The Role of Episodic Memory in Human Covariation Assessment

Behavioral and Electrophysiological Investigations on the Distinctiveness-Based Illusory Correlation

Dissertation
zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften
der Fakultät HW
Bereich Empirische Humanwissenschaften
der Universität des Saarlandes

vorgelegt von
Michael Weigl
aus Brackenheim

Saarbrücken 2019

Dekan:

Prof. Dr. Stefan Strohmeier, Universität des Saarlandes

Berichterstatter:

Prof. Dr. Axel Mecklinger, Universität des Saarlandes

Prof. Dr. Dirk Wentura, Universität des Saarlandes

Tag der Disputation: 15.07.2019

Abstract

Illusory correlations (IC) are subjectively assessed correlations which differ systematically from an actually observed correlation. In stereotyping people form associations between majorities and frequent, desirable behavior and minorities and infrequent, undesirable behavior, although group membership and behavior are actually uncorrelated. The Shared Distinctiveness Account (SDA) explains ICs by differential accessibility of infrequent and distinctive group-behavior combinations in episodic memory. However, this view has been challenged by proponents of an Information Loss Account (ILA), who claim that ICs result from regression to the mean due to noise in memory channels. This noise is supposed to especially affect infrequent group-behavior combinations. A third major account of the IC proposes that category accentuation causes ICs. According to the accentuation account, the focus of attention should lie on both the most frequent and the least frequent category combination in order to maximize the differentiation between the categories. The studies reported in this thesis aimed at determining the role of episodic memory in human covariation assessment and the relative merit of the different accounts of ICs by using behavioral and event-related potential (ERP) methods.

According to the ILA, ICs result from greater fading of infrequent group-behavior combinations in memory. Experiment 1 and 2 investigated whether ICs can be observed not only under standard conditions with skewed category frequencies (i.e. 2:1 ratio for positive and negative traits; Experiment 1), but also under conditions with equated category frequencies (i.e. 1:1 ratio for positive and negative traits; Experiment 2). Equated category frequencies preclude regression, but allow differential accessibility due to distinctiveness. We conducted a computer simulation study based on the ILA which showed that ICs are expected under standard conditions with skewed category, but not under conditions with equated category frequencies. Contrary to these simulations, our behavioral experiments revealed an IC under both, the skewed frequency condition and the equated frequency condition. Thus, information loss alone is not sufficient as an explanation for the formation of ICs. These

results imply that negative items contribute to ICs not only due to their infrequency, but also due to their emotional significance.

Since the results of Experiment 1 and 2 were inconclusive concerning the involvement of episodic memory, we conducted Experiment 3 in order to clarify the role of episodic memory in the generation of ICs. In this online questionnaire study, we improved the methods of prior studies by accounting for primacy and recency effects and by reducing the impact of response bias via the inclusion of novel distracters in the source memory task. Participants overestimated the frequency of negative behaviors in the minority and evaluated the minority less favorable than the majority indicating that the online questionnaire successfully induced an IC. Overall source memory accuracy was equal for the majority and the minority. However, memory for negative behavior was elevated for the minority even after controlling for response bias. This result is consistent with the prediction from the SDA. Experiment 3, thus, implies that heightened availability of distinctive group-behavior combinations underlies the IC.

In Experiment 4, we used ERP in order to compare the SDA with the accentuation account. According to the SDA, most attention should be paid to the least frequent category combination at learning, whereas the accentuation account predicts that most attention should be paid to the most frequent and the least frequent category combination. An active oddball task was used to elicit a P300, a marker for subjective probability and attention allocation. The SDA would be consistent with a linear increase of the P300 amplitude from the most frequent to the least frequent category combination, whereas the accentuation account would be consistent with larger P300 amplitudes for the most frequent and the least frequent category combinations than for the moderate frequent category combinations. Consistent with the SDA, we found a linear increase in the P300 as a function of infrequency. Furthermore, a frontal slow wave differentiated between the least frequent category combination and all other category combinations. Together, our results not only support the SDA, but also indicate that a fronto-parietal network is involved in the formation of mental contingency representations.

The von Restorff effect refers to the phenomenon that memory is enhanced for physically or semantically distinctive events. Distinctive events typically elicit a P300 and semantically distinctive events additionally elicit an N400. Memory studies with free recall generally report that P300 amplitude predicts subsequent memory performance. Some recent evidence indicates that N400 amplitude at study co-varies with familiarity-based recognition. In Experiment 5, the subsequent memory paradigm was combined with the von Restorff paradigm in order to test whether the N400 and the P300 at encoding predicts recognition based on familiarity and recollection, respectively. We recorded ERPs to physically and semantically distinctive items and control items at encoding and tested whether these items were later recognized on the basis of recollection or familiarity using the remember/know procedure. At encoding, physically and semantically distinctive items elicited a P300 and semantically distinctive items elicited an N400. Whereas the P300 amplitude to physically and semantically distinctive items at encoding was significantly larger for remembered than for known and forgotten items, the N400 did not differ between items subsequently remembered, known or forgotten. Unexpectedly, no von Restorff effect was found for physically and semantically distinctive items. However, high overall memory performance might explain the absence of a von Restorff effect. The ERP results support the view that P300 activity at encoding is linked to subsequent recollection-based recognition. Thus, recollection, but not familiarity, seems to depend crucially on the encoding context.

Experiment 6 was an ERP study designed to determine whether perceived item distinctiveness was related to subsequent memory and the IC. We exploited the fact that more distinctive items elicit larger P300 responses than less distinctive items, which also predicts subsequent memory performance differences for such items. Distinctiveness at encoding was created by presenting words that infrequently differed either in color or valence from frequently presented, positive words. Shared distinctive items deviated in both color and valence. We hypothesized that shared distinctiveness would lead to an enhanced P300 subsequent memory effect (SME), better source memory performance, and an overestimation of the frequency of shared distinctive items. Behavioral results indicated the presence of a shared distinctiveness effect in source memory and an overestimation of the frequency of shared distinctive items. In addition, memory was enhanced for positive items in the frequent color. This pattern was also reflected in the P300 for highly positive and negative items. However, shared distinctiveness did not modulate the P300 SME indicating that the processing of distinctive features might be only indirectly related to better encoding. This study shows that shared distinctiveness indeed leads to better source memory and ICs. In contrast to predictions based on the SDA, source memory did not predict the extent of IC. Since effects were observed for the most frequent and the least frequent category combination, our results imply that the processing of distinctiveness might lead to attention allocation to diametrical category combinations, thereby accentuating the differences between the categories.

Together, the results from the studies presented in this thesis provided little support for the ILA. In contrast, both, the SDA and the accentuation account, can account for parts of the results, but neither account provides a sufficient explanation for all results obtained in our studies. Furthermore, the studies in the present thesis were able to substantiate the effect of shared distinctiveness on memory in the IC paradigm. However, the results imply that episodic memory is predictive for the extent of IC, when participants deal with small samples as in Experiment 1, 2, and 3, but not when participants deal with large samples as in Experiment 6.

Zusammenfassung

Illusorische Korrelationen (IC) sind subjektiv bewertete Korrelationen, die sich systematisch von einer tatsächlich beobachteten Korrelation unterscheiden. Bei der Stereotypisierung bilden Menschen Assoziationen zwischen Mehrheiten und häufigen, erstrebenswerten Verhaltensweisen und Minderheiten und seltenen, unerwünschten Verhaltensweisen, obwohl Gruppenzugehörigkeit und Verhalten tatsächlich unkorreliert sind. Der Shared-Distinctiveness-Ansatz (SDA) erklärt ICs durch differenziellen Zugriff auf seltene und somit distinkte Gruppen-Verhaltenskombinationen im episodischen Gedächtnis. Diese Ansicht wurde jedoch von Befürwortern eines Information-Loss-Ansatzes (ILA) in Frage gestellt, die behaupten, dass ICs durch Regression zur Mitte aufgrund von Rauschen in Speicherkanälen entstehen. Dieses Rauschen soll vor allem selten auftretende Gruppen-Verhaltenskombinationen betreffen. Ein dritter Ansatz der IC schlägt vor, dass Kategorieakzentuierung ICs verursacht. Laut des Akzentuierungsansatzes sollte der Fokus sowohl auf der häufigsten als auch auf der seltensten Kategorienkombination liegen, um die Differenzierung zwischen den Kategorien zu maximieren. Die in dieser Arbeit berichteten Studien zielten darauf ab, die Rolle des episodischen Gedächtnisses bei der subjektiven Kovariationsbeurteilung zu identifizieren und den relativen Nutzen der verschiedenen Ansätze zur Erklärung der IC mithilfe von behavioralen Methoden und ereigniskorrelierter Potentiale (EKPs) zu ermitteln.

Laut dem ILA resultieren ICs aus einem stärkeren Informationsverlust von seltenen Gruppen-Verhaltenskombinationen im Gedächtnis. In Experiment 1 und 2 wurde untersucht, ob ICs nicht nur unter Standardbedingungen mit schiefen Kategorienhäufigkeiten (d.h. Verhältnis von 2:1 für positive und negative Merkmale; Experiment 1), sondern auch unter Bedingungen mit angeglichenen Kategorienhäufigkeiten (d.h. Verhältnis von 1:1 für positive und negative Eigenschaften; Experiment 2) beobachtet werden können. Angeglichene Kategorienhäufigkeiten schließen eine Regression zur Mitte aus, ermöglichen jedoch eine unterschiedliche Zugänglichkeit aufgrund von Distinktheit. Wir führten auf der Grundlage des ILA eine Computersimulationsstudie durch, die zeigte, dass ICs unter der Standardbedingung mit schiefen Kategorienhäufigkeiten erwartet werden, nicht

jedoch unter Bedingungen mit angeglichenen Kategorienhäufigkeiten. Im Gegensatz zu den Ergebnissen der Simulationen zeigten unsere Verhaltensexperimente eine IC sowohl unter der schiefen Häufigkeitsbedingung als auch unter der Bedingung mit angeglichenen Häufigkeiten. Informationsverlust allein reicht daher nicht aus, um die Bildung von ICs unter beiden Bedingungen zu erklären. Diese Ergebnisse implizieren, dass negative Stimuli nicht nur aufgrund ihrer Häufigkeit, sondern auch aufgrund ihrer emotionalen Bedeutsamkeit zu ICs beitragen.

Da die Ergebnisse von Experiment 1 und 2 hinsichtlich der Beteiligung des episodischen Gedächtnisses nicht eindeutig waren, führten wir Experiment 3 durch, um die Rolle des episodischen Gedächtnisses bei der Entstehung von ICs zu klären. In dieser Online-Fragebogenstudie haben wir die Methoden früherer Studien verbessert, indem Primacy- und Recency-Effekte berücksichtigt und die Auswirkung von Antwortverzerrungen durch die Einbeziehung neuartiger Distraktoren in die Quellengedächtnisaufgabe reduziert wurden. Die Teilnehmer überschätzten die Häufigkeit des negativen Verhaltens in der Minderheit und bewerteten die Minderheit weniger positiv als die Mehrheit, was darauf hinweist, dass der Online-Fragebogen erfolgreich eine IC induzierte. Die Quellengedächtnisleistung war für die Mehrheit und die Minderheit gleich. Das Gedächtnis für negative Verhaltensweisen war jedoch für die Minderheit erhöht, selbst wenn für Antwortverzerrungen kontrolliert wurde. Dieses Ergebnis stimmt mit der Vorhersage der SDA überein. Experiment 3 impliziert somit, dass der IC eine erhöhte Verfügbarkeit von unterschiedlichen Gruppen-Verhaltenskombinationen zugrunde liegt.

In Experiment 4 haben wir EKPs verwendet, um den SDA mit dem Akzentuierungsansatz zu vergleichen. Laut dem SDA sollte der seltensten Kategorienkombination beim Lernen größte Aufmerksamkeit geschenkt werden, wohingegen der Akzentuierungsansatz vorhersagt, dass der häufigsten und der seltensten Kategorienkombination die größte Aufmerksamkeit gewidmet werden sollte. Ein aktiver Oddball wurde verwendet, um eine P300, ein Marker für die subjektive Wahrscheinlichkeit und die Zuweisung von Aufmerksamkeit, auszulösen. Der SDA wäre konsistent mit einer linearen Erhöhung der P300-Amplitude von der häufigsten zur seltensten Kategorienkombination, wohingegen der Akzentuierungsansatz mit größeren

P300-Amplituden für die häufigsten und die seltensten Kategorienkombinationen im Vergleich zu moderat häufigen Kategorienkombinationen vereinbar wäre. In Übereinstimmung mit dem SDA fanden wir eine lineare Zunahme der P300 als Funktion der Seltenheit. Außerdem differenzierte eine Frontal-Slow-Wave zwischen der seltensten Kategorienkombination und allen anderen Kategorienkombinationen. Zusammengenommen unterstützen unsere Ergebnisse nicht nur den SDA, sondern zeigen auch, dass ein frontoparietales Netzwerk an der Bildung von mentalen Kontingenzrepräsentationen beteiligt ist.

Der Von-Restorff-Effekt bezieht sich auf das Phänomen, dass das Gedächtnis für physisch oder semantisch distinkte Ereignisse verbessert ist. Distinkte Ereignisse rufen normalerweise eine P300 hervor, und semantisch distinkte Ereignisse lösen zusätzlich eine N400 aus. Gedächtnisstudien mit freiem Abruf berichten im Allgemeinen, dass die P300-Amplitude die nachfolgende Gedächtnisleistung vorhersagt. Einige neuere Befunde weisen darauf hin, dass die Amplitude der N400 bei der Enkodierung mit vertrautheitsbasiertem Wiedererkennung zusammenhängen könnte. In Experiment 5 wurde das Subsequent-Memory-Paradigma mit dem Von-Restorff-Paradigma kombiniert, um zu testen, ob die N400 und die P300 beim Enkodieren die Wiedererkennung anhand von Vertrautheit bzw. Rekollektion vorhersagen. Wir haben EKPs zu physisch und semantisch distinkte Wörter und Kontrollwörter bei der Enkodierung aufgezeichnet und mit dem Remember/Know-Verfahren getestet, ob diese Wörter später aufgrund von Rekollektion oder Vertrautheit wiedererkannt wurden. Bei der Enkodierung lösten physisch und semantisch distinkte Wörter eine P300 aus und semantisch distinkte Wörter lösten eine N400 aus. Während die P300-Amplitude für physisch und semantisch distinkte Wörter beim Enkodieren für per Rekollektion erinnerte Wörter signifikant größer war als für vertraute und vergessene Wörter, unterschied die N400 nicht zwischen Wörtern, die später erinnert, gewusst oder vergessen wurden. Wider Erwarten wurde kein Von-Restorff-Effekt für physikalisch und semantisch distinkte Wörter gefunden. Eine hohe allgemeine Wiedererkennensleistung könnte jedoch das Fehlen eines Von-Restorff-Effekts erklären. Die EKP-Ergebnisse unterstützen die Ansicht, dass die P300-Aktivität bei der Enkodierung mit der nachfolgenden rekollektionsbasierten Wiedererkennung verknüpft ist. Daher scheint Rekollektion, aber nicht die Vertrautheit, entscheidend vom Enkodierungskontext abzuhängen.

Mit der EKP-Studie in Experiment 6 sollte festgestellt werden, ob die wahrgenommene Distinktheit mit der nachfolgenden Gedächtnisleistung und der IC zusammenhängt. Wir nutzten die Tatsache, dass distinkte Stimuli größere P300-Amplituden hervorrufen als weniger distinkte Stimuli und dass die P300 auch spätere Unterschiede in der Gedächtnisleistung für solche Stimuli vorhersagt. Distinktheit bei der Enkodierung wurde dadurch erzeugt, dass in manchen Durchgängen Wörter präsentiert wurden, die sich in Farbe oder Valenz von den häufig präsentierten, positiven Wörtern unterschieden. Stimuli mit geteilter Distinktheit (shared distinctiveness) unterschieden sich sowohl in Farbe als auch in Valenz von der Mehrheit. Wir testeten die Hypothese, dass die geteilte Distinktheit zu einem stärkeren P300-Subsequent-Memory-Effekt (SME), einer besseren Quellengedächtnisleistung und einer Überschätzung der Häufigkeit der Stimuli mit geteilter Distinktheit führen würde. Die behavioralen Ergebnisse zeigten einen Effekt für geteilte Distinktheit im Quellengedächtnis und eine Überschätzung der Häufigkeit von Stimuli mit geteilter Distinktheit. Darüber hinaus war das Gedächtnis für positive Stimuli in der häufigen Farbe verbessert. Dieses Muster spiegelt sich auch in der P300 für sehr positive und negative Stimuli wider. Die geteilte Distinktheit hatte keinen Einfluss auf den P300-SME, was darauf hindeutet, dass die Verarbeitung von distinkten Merkmalen möglicherweise nur indirekt mit einer besseren Enkodierung einhergeht. Diese Studie zeigt, dass die geteilte Distinktheit tatsächlich zu besserer Quellengedächtnisleistung und ICs führt. Im Gegensatz zu den auf SDA basierenden Vorhersagen sagte das Quellengedächtnis das Ausmaß an IC nicht vorher. Da Effekte für die häufigste und die seltenste Kategorienkombination beobachtet wurden, deuten unsere Ergebnisse darauf hin, dass die Verarbeitung der Distinktheit zu einer Aufmerksamkeitsallokation auf diametralen gegenüberliegenden Kategorienkombinationen führen kann, wodurch die Unterschiede zwischen den Kategorien verstärkt werden.

Insgesamt stützen die Ergebnisse der in dieser Arbeit vorgestellten Studien den ILA nicht. Im Gegensatz dazu können sowohl der SDA als auch der

Akzentuierungsansatz Teile der Ergebnisse erklären, aber keiner der beiden Ansätze liefert eine hinreichende Erklärung für alle Ergebnisse unserer Studien. Darüber hinaus konnten die Studien der vorliegenden Arbeit die Wirkung der geteilten Distinktheit auf das Gedächtnis im IC-Paradigma belegen. Die Ergebnisse implizieren jedoch, dass das episodische Gedächtnis das Ausmaß an IC vorhersagt, wenn Personen es lediglich mit kleinen Stichproben wie in Experiment 1, 2 und 3 zu tun haben, nicht jedoch, wenn Personen mit großen Stichproben wie in Experiment 6 umgehen müssen.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AIC	Akaike Information Criterion
ANOVA	Analysis of Variance
AT	Attention Theory
BIAS	Brunswikian Induction Algorithm for Social Cognition
BIC	Bayesian Information Criterion
Dm	Difference due to memory
ECG	Electrocardiography
EEG	Electroencephalography
EOG	Electrooculography
ERP	Event-Related Potentials
fMRI	Functional Magnetic Resonance Imaging
IC	Illusory Correlation
ICC	Intraclass Correlation
ICA	Independent Component Analysis
ILA	Information Loss Account
JOL	Judgment of Learning
K	Know
MANOVA	Multivariate Analysis of Variance
MCM	Multi-Component Model
MLM	Multilevel Linear Modeling
PDP	Process-Dissociation Procedure
R	Remember
ROC	Receiver-Operator Characteristic
SDA	Shared Distinctiveness Account
SME	Subsequent Memory Effect

List of Publications

Parts of this thesis are included in articles which have already been published in journals. The content has been adapted to fit into the argumentation of the present thesis. In order to facilitate the reading process, I will consistently use “we” instead of “I” throughout the thesis.

Part of the contents presented in Chapter 1, 2, 3, and 6 have already been published as:

Weigl, M., Mecklinger, A., & Rosburg, T. (2018). Illusory correlations despite equated category frequencies: A test of the information loss account. *Consciousness and Cognition, 63*, 11-28.

Acknowledgement

Many people supported me in some way in the past years and without their support I would not have been able to write the present thesis.

First and foremost, I would like thank Prof. Dr. Axel Mecklinger and PD Dr. Timm Rosburg for their excellent supervision and support.

Next, I would like to thank the many student assistants and interns, who helped me with the preparation of the experiments, data collection, and data preprocessing: Julia Maria Bock, Patrizia Schneider, Nicole Recklies, Nadja Fürst, Hong Hanh Pham, Mayte Jimenez Sastre, Bente Lubahn, and Ronja Thiel. I would also like to thank our lab technician Robert Schmidt for his assistance during data collection.

I am particular grateful for the constant support and advice I received from current and past members of the Experimental Neuropsychology Unit. Special thanks go to Dr. Emma Bridger, Dr. Anna-Lena Kursawe, Jun.-Prof. Dr. Siri-Maria Kamp, Dr. Regine Bader, Dr. Lisa Kuhn, Dr. Kathrin Eschmann, Ann-Kathrin Zaiser, and Gerrit Höltje.

In addition, I also would like to thank past and current members of the Cognitive Psychology Unit, namely Prof. Dr. Dirk Wentura Dr. Charlotte Schwedes, Dr. Timea Folyi, Dr. Andrea Paulus, and Dr. Michaela Rohr for fruitful discussions on statistics and methodology and advices for concrete challenges during data analysis.

I greatly profited from being an associative member of the IRTG. In particular, I would like to thank all principal investigators and students, with whom I discussed my research projects. The IRTG also made a research stay at the NACN lab of the Chinese Academy of Science in Beijing possible. There, I especially profited from discussions with Prof. Dr. Raymond C. K. Chan, Dr. Kui Wang, Xin-lu Cai, Han-yu Zhou, and Peter Bang. I am very grateful for all the hospitality and support I received during my time in China.

Last, but not least, I would like to thank Fatimah Köspardiana for proofreading my thesis and her moral support. Finally, my thanks go to all my friends and family members, who supported me during the past years.

1 Introduction

> The first thing the intellect does with an object is to class it along with something else. (William James, 1902/1928, The Varieties of Religious Experience: A Study in Human Nature, p. 9)

This famous quote by William James (1902/1928) highlights the human propensity for classification. However, humans do not restrict themselves with mere classifications, because “[t]he next thing the intellect does is to lay bare the causes in which the thing originates” (William James 1902/1928, p. 9). Indeed, the ability to extract patterns, regularities, and causal relationships from observation ranges among the most fundamental tools an organism needs for survival and adaptive behavior (Fiedler, 2000). This ability is evidenced in a large body of literature ranging from classical and operant conditioning (e.g. Kirsch, Lynn, Vigorito, & Miller, 2004) to the learning of complex artificial grammars (e.g. Reber, 1967). In fact, humans have such a strong propensity to detect patterns that they infer contingencies from the environment, even when there are no contingencies – a phenomenon called “illusory correlation“ (Chapman, 1967; Hamilton & Gifford, 1976).

More technically, an illusory correlation (IC) is a subjectively perceived correlation between two events, which varies systematically from the actual covariation between those events (e.g. Chapman, 1967; Fiedler, 2000). The two events might actually not correlate at all or correlate with each other in another direction than reported. ICs have been investigated in basic research (e.g. Chapman, 1967; Tversky & Kahneman, 1973) as well as in applied research, like psychodiagnostics (Chapman & Chapman, 1967; Starr & Katkin, 1969), clinical psychology (Alloy & Abramson, 1979), or organizational psychology (Feldman, Camburn, & Gatti, 1986). Very fruitful investigations on the IC were conducted in stereotyping research (e.g. see Hamilton, 1981 or Stoessner & Plaks, 2001 for reviews).

Our social world is composed of many groups of varying sizes and desirable behavior is, in general, more prevalent than negative behavior (Hamilton & Gifford, 1976). In this context, ICs refer to the phenomenon that people tend to associate majorities with desirable behavior and minorities with undesirable behavior, even though group membership and behavior might actually be uncorrelated. Since rare group-behavior combinations in this scenario are assumed to be more distinctive than other combinations, this type of ICs has been dubbed distinctiveness-based IC (Hamilton & Gifford, 1976). Even though the existence of the IC phenomenon can be considered as firmly established (see Fiedler, 2000; Mullen & Johnson, 1990, for integrative reviews), little agreement has been reached so far about the conditions which are necessary and sufficient for ICs to arise (Sherman et al., 2009). As a consequence, ICs were investigated from very different theoretical angles and various mechanisms have been proposed to underlie the development of ICs (see Sherman et al., 2009; Van Rooy, Vanhoomissen, & Van Overwalle, 2013, for an overview and discussion). One of the mechanisms which has attracted a lot of attention is episodic memory. However, the empirical evidence on the contribution of episodic memory to the development of ICs is equivocal. While some authors reported a link between episodic memory performance and ICs (e.g. Hamilton, Dugan, & Trolier, 1985; Risen, Gilovich, & Dunning, 2007), others could not corroborate a direct link between memory and covariation assessment (e.g. Van Rooy et al., 2013).

Neurophysiological methods like EEG or fMRI receive increasing attention in the investigation of social cognition (e.g. Fiske & Taylor, 2013). Most investigations focus on pre-existing stereotypes or on person perception (e.g. Bartholow, Fabiani, Gratton, & Bettencourt, 2001). Only few neurophysiological studies investigated how new intergroup attitudes are formed (Spiers, Love, Le Pelley, Gibb, & Murphy, 2016). The present thesis, therefore, aims at determining the role of episodic memory in the distinctiveness-based illusory correlation not only by using behavioral methods, but also by applying the event-related potential (ERP) technique. Three behavioral studies tested whether distinctiveness contributes to both, superior memory and ICs. Two ERP studies investigated the influence of distinctiveness on the P300, an ERP component linked to the processing of distinctiveness. A third ERP study tested whether perceived item distinctiveness as indexed by the P300 can predict subsequent memory and the amount of perceived covariation between two features.

2 Theoretical background

2.1 Social cognitive foundation of the illusory correlation

2.1.1 The phenomenology of illusory correlations

Loren J. Chapman (1967) was the first to define and demonstrate the phenomenon of illusory correlation. He defined an illusory correlation as

> the report by observers of a correlation between two classes of events which, in reality, (a) are not correlated, or (b) are correlated to a lesser extent than reported, or (c) are correlated in the opposite direction from that which is reported. (Chapman, 1967, p., 151)

In Chapman's (1967) experiment, participants learned a list of word pairs. In some cases, the words were substantially longer and thereby distinctive compared to other words (e.g. envelope-sidewalk). In other cases, some of the words were semantically associated (e.g. bread-butter). After the learning phase participants were asked to estimate the frequency of co-occurrence of the words. They systematically overestimated the frequency of distinctive and semantically associated words. Thus, Chapman established associations and distinctiveness as factors contributing to the formation of ICs.

Most studies approach the IC with a 2 x 2 contingency table in mind (McGarty & de la Haye, 1997). The normative, statistically correct approach to assess whether certain categories covary with other categories in a 2 x 2 contingency table as shown in Table 2.1a) would consist in the calculation of the phi coefficient:

$$\phi = \frac{ad - bc}{\sqrt{(a+c)(b+d)(a+b)(c+d)}}$$

Table 2.1b) presents individual cell frequencies which are commonly used in IC research (Mullen & Johnson, 1990). Such cell frequencies will result in φ = 0, i.e. a zero correlation. In IC research, the reported subjective frequency estimates are typically transformed into a phi coefficient, which is then compared with the normatively correct phi coefficient. The intriguing fact about ICs is that the intuitive covariation assessment by lay people systematically and predictably deviates from the normative result (Fiedler, 2000).

Table 2.1 A 2 x 2 contingency table. Table a) is used for in illusory correlation research to denote the cell frequencies. Table b) presents an example for a null correlation (φ = 0)

a)	Y1	Y2	b)	Y1	Y2
X1	a	b	X1	16	8
X2	c	d	X2	8	4

However, the IC is not a uniform phenomenon. Three different types of ICs have been identified in the literature – each associated with a specific pattern of results and a distinct research tradition (Fiedler, 2000; Hamilton, 1981; Tversky & Kahneman, 1973): expectancy-based ICs, ICs based on a positive-negative asymmetry, and the distinctiveness-based ICs.

In expectancy-based ICs, participants already have an expectation about the relationship between two variables, based on their experiences and personal beliefs. When study participants have to judge the covariation between well-known groups (e.g. accountants and salesmen) and certain traits (e.g. timid and talkative) in a new set of stimuli, their judgment on the new set is usually consistent with their pre-experimental expectations (Hamilton & Rose, 1980; see Fiedler, 2000, for an extensive discussion of the expectancy-based IC). In social psychology expectancy-based ICs were investigated in order to assess the impact of pre-existing stereotypes on contingency judgments (e.g. Hamilton & Rose, 1980; Slusher & Anderson, 1987). Moreover, clinical psychologists have applied the expectancy-based ICs to the study of phobias and related disorders (e.g. de Jong & Merckelbach, 2000; Tomarken, Mineka, & Cook, 1989; Tomarken, Sutton, & Mineka, 1995). In these studies, the expectancy-based IC (also called covariation bias in clinical studies) manifests itself as an overestimation of fear-consistent pairings (e.g. spiders and electric shock; de Jong & Merkelbach, 2000).

The IC based on a positive-negative asymmetry arises from the asymmetric psychological impact of positive and negative information (Fiedler, 2000). In this case, positive information refers to the presence of a feature or an effect (e.g. a symptom or a disease), whereas negative information refers to its absence (e.g. absence of a symptom or a disease). In the seminal study by Jenkins and Ward (1965), participants had to judge whether they can control an outcome (a light) by pressing one of two buttons. When button press and outcome were uncorrelated, participants reported more control under conditions with high success rate than under conditions with low success rates. Thus, the participants' feeling of control was mainly a function of positive feedback rather than negative feedback (see Fiedler, 2000, for an extensive discussion of the IC based on a positive-negative asymmetry). The IC based on a positive-negative asymmetry can also be seen as an illusion of control (Langer, 1975) under conditions, in which participants can make choices, but have no control over the outcome. However, unlike the illusion of control, ICs based on a positive-negative asymmetry can also arise in situations in which participants merely observe pairs of features (e.g. symptom and disease; Smedslund, 1963).

Participants in experiments on the distinctiveness-based IC have to infer a correlation about material for which they do not possess preexisting expectations about the relationship. In the seminal study of Hamilton and Gifford (1976), participants read short descriptions about members of two fictional groups – group A and group B, with group A having twice as many members as group B. For both groups, two-thirds of the description referred to desirable behavior and one-third to undesirable behavior. In other words, group membership and behavior were uncorrelated. Despite the absence of a correlation, participants evaluated the majority more favorable than the minority on three different dependent measures: group assignment, frequency estimation, and evaluative trait ratings. In other words, the participants showed a tendency to associate the majority with the frequent, desirable behavior and the minority with the infrequent, undesirable behavior. Thus, an illusory correlation was formed between group membership and desirability of the behavior. In a second experiment, Hamilton and Gifford reversed the frequency ratio of desirable to undesirable behavior and found that the minority was now evaluated more favorable than the majority. Again, this pattern was observed consistently across a range of dependent measures (Hamilton & Gifford, 1976, Experiment 2; see Mullen & Johnson, 1990 for a review). The popularity of the concept of ICs stems from the fact that it offers a cognitive explanation for the formation of stereotypes. Moreover, the experimental set-up resembles the situation we encounter in our everyday life: there are majorities and minorities in every society and minorities are by definition smaller than majorities. Furthermore, most people behave in a norm-consistent, desirable way (e.g. Alves, Koch, & Unkelbach, 2017; Fiske, 1980; Kanouse, 1984).

In the remainder of the thesis, we will primarily refer to the distinctiveness-based IC as this line of research inspired the most systematic empirical investigations and theorizing. Furthermore, this line of research most extensively investigated the relationship between episodic memory and illusory correlations.

2.1.2 Explanatory frameworks for the illusory correlation

Skewed frequency distributions are assumed to be essential for the distinctiveness-based IC to arise (e.g. Fiedler, 1991, 1996; Hamilton & Gifford, 1976). However, there is a still ongoing debate about the mechanisms by which skewed frequency distributions influence our judgment and a variety of models have been put forward to explain ICs (Sherman et al., 2009). As a consequence, ICs have been investigated from various theoretical perspectives (e.g. availability account: Rothbart (1981); memory trace model: Smith, 1991; pseudocontingencies: Fiedler, Freytag, & Meiser, 2009; recurrent connectionist model: Van Rooy, Van Overwalle, Vanhoomissen, Labiouse, & French, 2003; Rescorla-Wagner model: Murphy, Schmeer, Vallée-Tourangeau, Mondragón, & Hilton, 2011). The discussion in this section will focus on the four accounts which are most relevant for the experiments reported in this thesis: the Shared Distinctiveness Account (SDA), the Information Loss Account (ILA), the accentuation account, and Attention Theory (AT; Sherman et al., 2009).

2.1.2.1 The Shared Distinctiveness Account

The first and most often cited explanation is the Shared Distinctiveness Approach (Hamilton & Gifford, 1976). The SDA states that infrequent combinations are more distinctive and, therefore, better encoded than more common ones (Chapman, 1967; Hamilton & Gifford, 1976; Tversky & Kahneman, 1973). Infrequent combinations are therefore more easily available in memory than others. In the IC paradigm, negative behavior of the minority is the most infrequent category combination and, therefore, assumed to be most distinct and salient. This leads to deeper encoding and higher availability at retrieval. As individuals estimate the frequency of the combinations on the basis of their availability, infrequent combinations exert a stronger influence on our judgment than other combinations and are consequently overestimated (Tversky & Kahneman, 1973).

Evidence for better memory for shared distinctive items stems from studies using free recall (Hamilton et al., 1985) and one-shot ICs (Risen et al., 2007). Further support can be found in the memory literature: Distinctive items are in general better remembered than non-distinctive items (e.g. Alves et al., 2015; Fabiani & Donchin, 1995; Hunt, 1995, 2009; von Restorff, 1933; see also Schmidt, 1991, 2012, for an integrative account). Furthermore, memory is even better for items that are distinctive on several stimulus dimensions (e.g. Hunt & Mitchell, 1982; Kuhbandner & Pekrun, 2013). Additional evidence for the SDA stems from studies using reading and reaction times: reading times were longer for shared distinctive items than for other items indicating that participants spent more time on encoding shared distinctive items (Stroessner, Hamilton, & Mackie, 1992). However, reactions to shared distinctive items were facilitated at retrieval indicating that these items were more available in memory (Johnson & Mullen, 1994; McConnell, Sherman, & Hamilton, 1994).

This evidence has been criticized for not distinguishing between causes and effects of stereotyping (Meiser, 2008). Free recall performance and reaction times might have been facilitated by the newly acquired stereotype and, thus, might reflect the consequence rather than the cause of the illusory correlation. The evaluation of the evidence for the SDA is further hampered by the fact that researchers conceptualized distinctiveness quite differently. For example, Hamilton and Gifford (1976) defined distinctiveness as infrequency, whereas

Feldman et al. (1986) also considered negativity as distinctive. Furthermore, the memory advantage of the shared distinctive category combination could not be replicated with measures based signal-detection theory (Fiedler, Russer, & Gramm, 1993) or multinomial processing tree models (e.g. Klauer & Meiser, 2000). Moreover, ICs have been found in studies, in which the frequency of shared distinctive items was zero (i.e. no items in cell d of the 2 x 2 contingency table; e.g. Fiedler, 1991; Van Rooy et al., 2013).

2.1.2.2 The Information Loss Account

The Information Loss Account (ILA) offers an alternative explanation for ICs without assuming any differential processing of information (Fiedler, 1991, 1996, 2000; Smith, 1991). According to the ILA, we might observe and encode the different group members and their behavior correctly. Since perception and memory are far from perfect, noise can distort parts of information during encoding, storage, and retrieval. For example, due to such noise, we might misremember a rude person from the majority as a member of the minority. If participants are asked to make a judgment about their attitude towards the groups, unbiased aggregation of these distorted data alone is sufficient to lead to the erroneous conclusion that a correlation between groups and behavior is present. In the typical IC experiments the distribution of the valence of the behavior is skewed, i.e. positive stimuli are objectively more frequent than negative stimuli (or vice versa; Figure 2.1). For the majority, the preponderance of positive (or negative) behaviors becomes evident to the participant during encoding, because they can aggregate over a large number of instances. Therefore, the subjective frequency estimate for the majority are not so much affected by noise and should roughly correspond to the actual frequencies. The minority, however, is more strongly affected by this noise because single outliers have more influence on the estimates of smaller samples and the estimates regress to the mean (see 33% and 67% condition in Figure 2.1). Thus, the main appeal of the ILA is its parsimony. ICs can be explained via learning mechanisms that rely on the differences in sample size and unsystematic noise without assuming biased processing on the side of the participants (e.g. different processing of distinctive and non-distinctive information).

Evidence for the ILA stems from computer simulations that reproduce the IC effect without the assumption of biased processing (Fiedler, 1996, 2000; Smith, 1991). But there is also experimental evidence that the overestimation of frequencies increases when categories are split into sub-categories (Fiedler, 1991; Fiedler & Armbruster, 1994) or that ICs can be observed even in the absence of distinctive or infrequent information (Fiedler, 1991; Shavitt, Sanbonmatsu, Smittipatana, & Posavac, 1999; Van Rooy et al., 2013).

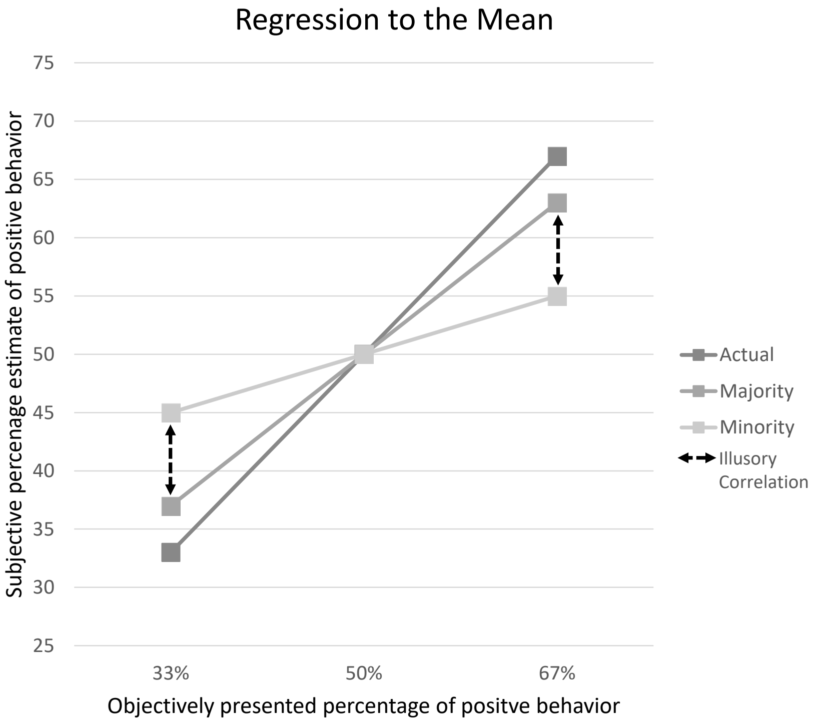

Figure 2.1 Regression to the mean as a potential cause for illusory correlation. In the 67% condition positive behavior is presented twice as often as negative behavior. In the 33% condition, in contrast, negative behavior is presented twice as often as positive behavior. In both cases, the subjective estimates of the majority (dark grey line) are closer to the actually presented frequency (black line) than the estimates of the minority (light grey line). The IC manifests itself as the difference between the subjective estimate of positive behavior for the majority and the minority (dotted double arrow). No IC would be expected if positive and negative behavior were equally frequent (50% condition). Adapted from Weigl, Mecklinger, and Rosburg (2018, Figure 1).

However, enhanced memory or prolonged reading times for distinctive category combinations as reported in some studies point to genuine differences in encoding and these differences are correlated with the extent of IC (Hamilton et al., 1985; Risen et al., 2007; Stroessner et al., 1992). Moreover, there is some evidence that IC disappear after extended learning indicating incomplete learning rather than information loss (Murphy et al., 2011; but see Kutzner, Vogel, Freytag, & Fiedler, 2011 for conflicting evidence).

2.1.2.3 Approaches Relying on Accentuation

A third mechanism proposed to underlie ICs is category accentuation (McGarty, Haslam, Turner, & Oakes, 1993; Sherman et al., 2009). Accentuation manifests itself as the exaggeration of both, between-category differences and within-category similarities (Tajfel, 1959; Tajfel & Wilkes, 1963). However, according to Tajfel (1959; see also Tajfel & Wilkes, 1963) accentuation only takes place if the category is predictive for the judgmental dimension. The accentuation effect has been shown in areas ranging from physical perception (Tajfel & Wilkes, 1963) to social categorization (Doise, Deschamps, & Meyer, 1978) or the processing of painful stimuli (van der Meulen, Anton, & Petersen, 2017). In the seminal experiment by Tajfel and Wilkes (1963) the length of lines were exaggerated when they were assigned to the arbitrary categories A and B. Krueger and Rothbart (1990, see also Krueger, Rothbart, & Sriram, 1989) have proposed that category members that heighten between-category differences and reduce within-category variance receive more attention at encoding and greater weight at judgment.

2.1.2.3.1 The Accentuation Account

In the Hamilton and Gifford (1976) IC paradigm, for example, two groups and positive and negative behavior are presented. Positive behavior of majority members and negative behavior of minority members would lead to a positive impression of the majority, whereas the reverse is true for negative behavior of majority members and positive behavior of minority members (McGarty et al., 1993). McGarty et al. (1993) proposed a pure accentuation account of the IC and argued that the greater absolute difference between the number of positive and negative items for the majority than the minority can be interpreted as a real group difference. Therefore, subjects want to enhance this difference by accentuation and, consequently, judge the majority to be more favorably than the minority.

More precisely, McGarty and colleagues (1993) assume that participants try to find meaning in the presented material and, therefore, actively search for differences between the majority and minority. The participants use the desirability of the behavior descriptions, because this is the only variable that would allow differentiation between the groups. The active search for meaning was most convincingly demonstrated in the second experiment by McGarty et al. (1993). In this experiment, participants only learned that there were a majority and a minority. Next, participants learned person descriptions. But in contrast to the typical IC experiment, McGarty and colleagues (1993) omitted the group labels from these descriptions. Despite the omission of the group labels, the participants formed a strong IC as predicted by the accentuation account. According to McGarty et al. (1993), it is not necessary to learn category combinations. Learning of the categories alone is sufficient to prompt a search for meaning in the material, which ultimately leads to the formation of an IC. Further support for the accentuation account stems from studies showing that ICs typically do not occur for dimensions which are meaningless in the experimental situation like gender or handedness (Haslam, McGarty, & Brown, 1996; Klauer & Meiser, 2000) and that ICs typically become stronger towards the end of an experiment (Berndsen, Spears, Van der Pligt, & McGarty, 1999).

On a conceptual level, Meiser (2008) criticized that the accentuation account offers just a pure description of the IC phenomenon and claims that the search for basic learning and memory processes are more fruitful. Indeed, a study by Murphy et al. (2011) challenged the view that ICs increase towards the end of the experiment. They found that ICs are strongest when learning is still incomplete. Once participants learned sufficient trials, the IC disappears. Moreover, the accentuation account faces some difficulties in explaining why people do not acquire an IC, when they have to make judgments during encoding (e.g. Pryor, 1986).

2.1.2.3.2 Attention Theory

Attention Theory (AT; Kruschke, 2003) was developed to account for the inverse base-rate effect (Medin & Edelson, 1988), but has been extended to other frequency-related phenomena like base-rate neglect and ICs (Sherman et al., 2009). In situations with skewed frequency distributions for categories (e.g. majority and minority) and their attributes (e.g. positive or negative behavior), the speed of acquisition is higher for the more frequent attribute as compared to the less frequent attribute. The attribute which is learned first not only defines the category that is acquired first, but also what other attributed are used for differentiation and accentuation.

In contrast to the accentuation account by McGarty et al. (1993), real category differences are not necessary for accentuation in AT, because any cause that leads to differences in the speed of acquisition of categories will result in accentuation (Sherman et al., 2009). According to AT, ICs arise, because the majority and positive behavior are more frequent than the minority and negative behavior. Therefore, the positivity of the majority is learned first. In AT, distinctive features of the less frequent category distinguish it from the more frequent category and allow further differentiation between the categories. When subjects form an impression of the minority, attention will shift to the negative behavior, because it is the only remaining attribute that allows differentiation from the majority. These attention shifts result in a less favorable impression of the minority relative to the majority. AT, therefore, combines the mechanisms of distinctiveness and accentuation. In contrast to the SDA, contextual rather than absolute distinctiveness is important for AT.

In essence, AT claims that the two diametrically opposed category combinations, the most frequent and the least frequent in the IC paradigm, receive more attention than the remaining two category combinations for the purpose of category accentuation and differentiation (Sherman et al., 2009). As Sherman et al. (2009) demonstrated in five experiments, AT is suitable to explain both, accentuation effects (e.g. Tajfel & Wilkes, 1963) and ICs (e.g. Hamilton & Gifford, 1976). Of note, ICs arose under conditions in which two orthogonal traits were considered rather than a single trait dimension. This led to the formation of different stereotypes for each group. Even though the AT by Sherman et al. (2009) does not make specific predictions regarding the

contribution of episodic memory to ICs, it seems plausible that the attention shifts should also contribute to better encoding.
Sherman et al. (2009) were able to confirm their AT approach to IC in five experiments. The strongest evidence for attention shifts were found in Experiment 5 (Sherman et al., 2009). Consistent with the predictions of AT, participants reacted faster to probes at the position of the common trait behavior of the majority or at the rare trait behavior of the minority than to probes at the position of the rare trait behavior of the majority or the common trait behavior of the minority. Attention shifts, thus, facilitated differentiation between the categories. Furthermore, attention shifts were also demonstrated in an eye-tracking study (Kruschke, Kappenman, & Hetrick, 2005).

2.2 Distinctiveness and Memory

Ever since Hedwig von Restorff (1933) demonstrated in her seminal experiment that memory for items which were distinctive or isolated during study was enhanced (the so-called Von Restorff effect or isolation effect), countless studies have investigated the link between distinctiveness and memory performance (for reviews see Hunt & Worthen, 2006; Schmidt, 1991; Wallace, 1965). These studies revealed that the effect of distinctiveness on memory crucially depends on the type of distinctive event. Schmidt (1991, 2012) identified four different types of distinctiveness: 1) primary distinctiveness, 2) secondary distinctiveness, 3) emotional significance, and 4) high priority stimuli. Since the concept of distinctiveness is central for the investigations on the IC in the present thesis, Schmidt's (1991, 2012) classification will be outlined in more detail in this section. For the purpose of the present thesis, we define a stimulus as being distinctive, if it fulfills one of the four criteria in Schmidt’s (1991, 2012) classification.

2.2.1 Primary Distinctiveness

Schmidt (1991) defined primary distinctiveness as the absence of feature overlap with content in working (or primary memory). Primary distinctive events or stimuli pop out from their immediate context, but might not be distinctive or unusual in a different context (Schmidt, 1991, 2012). These events or stimuli can even be quite mundane. For example, a word in red font in a list of words with black font is isolated[1]. If all words were written in a red font, however, the very same word would not be perceived as distinctive. Other examples of primary distinctiveness include differences in physical features like size or font, or semantic features like a category mismatch. In most studies, primary distinctiveness is operationalized as the infrequency of a stimulus relative to other stimuli. Most studies on the distinctiveness-based IC also rely on frequency manipulations to render certain groups or behaviors distinctive (Fiedler, 2000; see Mullen & Johnson, 1990, for a review). According to Schmidt (2012), the von Restorff or isolation effect is mainly an effect of primary distinctiveness. The isolation effect is easier to obtain in recall than in recognition (see Schmidt, 2012, for a discussion). Indeed, von Restorff herself faced some problems, when she tried to obtain an isolation effect in recognition (Hunt, 1995).

While early encoding views attributed the memory advantage for primary distinctiveness to increased attention devoted to the distinctive item during encoding (Green, 1958; Jenkins & Postman, 1948), later expectancy violation views proposed that the distinctive stimulus violates expectations about the composition of the stimulus series and prompts more extensive processing which ultimately benefits subsequent retrieval (e.g. Fabiani & Donchin, 1995). In Schmidt's (1991) incongruity theory, the distinctive stimulus is incongruent with the currently activated conceptual frame work which leads to an automatic increase in attention allocated to the distinctive stimulus followed by more controlled encoding processes like elaboration or rehearsal.

Studies using self-paced presentation rates provided some evidence in support for the claim that additional study time is devoted to distinctive items (e.g. Stroessner et al., 1992). Of note, distinctive items are remembered better than common items even with presentation rates which prevented additional study time (Waddill & McDaniel, 1998) or in the absence of a difference in study

[1] In research on the Von Restorff effect, the terms isolation and primary distinctiveness are often used synonymously (e.g. Schmidt, 2012). We will follow this convention in the present thesis.

time between common and distinct words (Hunt & Elliot, 1980). Furthermore, primary distinctiveness is not strongly affected by manipulations of attention (e.g. Bruce & Gaines, 1976; see also McDaniel & Geraci, 2006, for a review). In addition, primary distinctiveness typically does not lead to worse memory for background items (i.e. a suppression effect) in recognition, even though suppression effects are sometimes observed in recall (see Schmidt, 2006, for a review on the suppression effect). In general, suppression effects are more often observed for significant or emotional items than for primary distinctiveness (Schmidt, 2006).

Even though most encoding views imply that items can be distinctive only after some context has been established (McDaniel & Geraci, 2006), several studies, including von Restorff's original study, have convincingly shown that memory is increased even for distinctive items at the very first position of a study list (e.g. Hunt, 1995; McConnell et al., 1994; von Restorff, 1933). Furthermore, Dunlosky, Hunt, and Clark (2000) used judgments of learning (JOLs) as a measure of subjective salience and found that isolated items in the middle, but not at the beginning of a list received higher JOLs. However, an isolation effect was obtained for both positions. Hunt (2009) investigated whether salience effects for distinctive items would emerge at long retention intervals (48 H after encoding), but found that the isolation effect after 48 h was similar to the isolation effect immediately after the study phase. Geraci and Manzano (2010) replicated the study by Dunlosky et al. (2000) using delayed JOLs and found that items at the beginning of a list were perceived as distinctive, when participants became aware of the list context. This is in line with McConnell et al. (1994) who reported that ultimate rather than relative distinctiveness was critical for the distinctiveness effect. Furthermore, ICs seem to depend on ultimate distinctiveness (McConnell et al., 1994). Together, these studies imply that the processing of contextual deviance at encoding does play a role in primary distinctiveness irrespective of salience.

Retrieval can also play a role in primary distinctiveness (McDaniel & Geraci, 2006). According to Hunt and McDaniel (1993), distinctive stimuli possess features which can be used as highly diagnostic retrieval cues. Retrieval can benefit from these cues, because they enhance discriminability in recall and recognition or because they guide retrieval to the distinctive stimulus

(McDaniel & Geraci, 2006). Hunt and Smith (1996) reported that the isolation can be enhanced with appropriate cues at retrieval suggesting that access to the study episode is facilitated by retrieval cues. This might be another reason why memory for distinctive items at the beginning of a list is enhanced even though they are not salient at encoding (Dunlosky et al., 2000). Moreover, people generate more, but less useful cues for familiar categories and more diagnostic cues for the distinctive category (Peynircioğlu & Mungan, 1993). Bruce and Gaines (1976) proposed that distinctive items form a separate category with their own set of specific cues. This might explain why the recall of primary distinctive items is typically clustered (e.g. Fabiani & Donchin, 1995).

To sum up, the effect of primary distinctiveness on memory depends on both, encoding and retrieval (Schmidt, 2012; McDaniel & Geraci, 2006). Moreover, primary distinctiveness can be dissociated from effects of salience.

2.2.2 Secondary Distinctiveness

Secondary distinctiveness was defined by Schmidt (1991) as the absence of feature overlap with content in long-term or secondary memory. In other words, secondary distinctive events are those that are distinctive with respect to the life-time experience or are perceived as bizarre. Most people, for example, have never seen a dog riding a bike and might even consider the thought as bizarre. In the case of bizarre imagery, the effect of secondary distinctiveness is known as bizarreness effect in the literature (Schmidt, 2012). Uncommon orthography, unusual faces, humorous material, or low word frequency are other examples for secondary distinctiveness.

Study times seem to affect the effect of secondary distinctiveness as Kline and Groninger (1991) found that the bizarreness effect for sentences was present at a slow presentation rate (15 s), but absent at faster presentation rate (11 s). Interestingly, the bizarreness effect for simple sentences arises even at a presentation rate of 5 s (Waddill & McDaniel, 1998). Moreover, the effect of secondary distinctiveness is diminished under divided attention (McDaniel & Geraci, 2006). Similar results were obtained for face recognition (Shepherd, Gibling, & Ellis, 1991). These data imply that more study time is needed to comprehend and encode secondary distinctive stimuli (McDaniel & Geraci, 2006).

Another interesting finding is that recall benefits from secondary distinctiveness in mixed, but not in unmixed lists (McDaniel & Geraci, 2006). Since these results are surprising in light of the definition of secondary distinctiveness (i.e. deviance from general knowledge or lifetime experience), Schmidt (1991) suggested that incongruency with the activated framework prompts additional processing which produces a memory advantage for secondary distinctive items. In unmixed lists, there is no incongruency with the current conceptual framework and, consequently, no memory advantage is observed (McDaniel & Geraci, 2006). Moreover, Worthen, Marshall, and Cox (1998) reported that the bizarreness effect increased with increasing number of common items preceding the distinctive item.

In light of this evidence, McDaniel and Geraci (2006) argue that secondary distinctiveness might reflect an encoding phenomenon. However, there is also evidence that retrieval processes also affect the effects of secondary distinctiveness. McDaniel, DeLosh, and Merritt (2000) found that instructing participants to encode the serial position or introducing categorical relations between the target words in the sentences eliminated the bizarreness effect. Since both types of manipulations are known to affect retrieval, the results by McDaniel et al. (2000) imply that retrieval processes in addition to encoding processes play a role in the bizarreness effect and by extension secondary distinctiveness in general (McDaniel & Geraci, 2006).

2.2.3 Emotional Significance and High Priority Stimuli

Schmidt (1991, 2006, 2012) defined primary and secondary distinctiveness as the absence of feature overlap with content in working memory and long-term memory, respectively. Significance, in contrast, requires not only overlap with content in memory, but also relevance for the individual (Schmidt, 2006). Moreover, significance contains a value judgment acquired directly via experience or indirectly via culture (Schmidt, 2006). Schmidt (2006, 2012), therefore, argued to distinguish between distinctiveness in the strict sense (i.e. primary or secondary distinctiveness) and significance. As a result, any discussion of significance needs to distinguish between so-called high priority stimuli and emotional significance.

High priority stimuli are goal-relevant, but neither emotionally engaging nor arousing stimuli (Schmidt, 2012). However, high priority stimuli may elicit an orienting response (Sokolov, 1963). An example for an experimental high priority manipulation would be the instruction to particularly remember famous names. Such manipulations produce a memory benefit for the high-priority stimuli (Tulving, 1969). In contrast to primary or secondary distinctiveness, studies on high priority stimuli typically do not involve a match/mismatch manipulation (Schmidt, 1991, 2012).

Emotional significance refers to emotional engaging stimuli which are characterized by valence (i.e. whether it is a positive of negative stimulus) and arousal (i.e. the degree of activation due to the stimulus; Schmidt, 2012). Some stimuli are emotionally significant due to normative evaluations or prior (un-) pleasant experiences, whereas other stimuli are emotionally significant due to biological preparedness (Seligman, 1971). For example, Öhman, Flykt, and Esteves (2001) found that evolutionary significant stimuli like spiders and snakes tend to immediately “pop-out” from their surroundings. Since most studies on the distinctiveness-based illusory correlation use positive and negative behaviors or traits, we will focus exclusively on emotional significance for the remainder of this section.

There is ample evidence that memory for an emotional event increases with increasing levels of arousal (e.g. Bradley, Greenwald, Petry, & Lang, 1992; Cahill, Prins, Weber, & McGaugh, 1994; Kleinsmith, Kaplan, & Trate, 1963; Mather, 2007). Furthermore, the β-adrenergic receptor antagonist propranol which decreases arousal can reduce or eliminate the mnemonic effect of emotional material (e.g. Cahill et al., 1994; Hurlemann et al., 2005), whereas reboxetine which increases arousal leads to an enhancement of the emotion effect (Hurleman et al., 2005). Arousing stimuli typically enhance memory for the gist of an event, but may have a detrimental effect on peripheral aspects (see Schmidt, 2006, 2012 for reviews). This is especially true for threatening stimuli (Schmidt & Saari, 2007).

The valence of a stimulus also has an impact on memory (Schmidt, 2012). In most cases, negative events have a stronger impact than positive events in several domains of information processing ranging from attention to decision-making (see Alves, Koch, & Unkelbach, 2017; Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001, for reviews). This robust effects are typically explained with the higher relevance of threatening stimuli for survival and well-being which leads to deeper and more elaborate processing (Baumeister et al., 2001) or the fact that two pieces of negative information are less similar to each other and less redundant than two pieces of positive information (e.g. Alves et al., 2017). In recognition memory, for example, negative stimuli are recognized more accurately than positive stimuli (e.g. Alves et al., 2015; Ortony, Turner, & Antos, 1983).

In addition, certain stimuli might be emotionally significant only to certain groups of people. Kramer and Schmidt (2007) embedded a picture of a bottle of Pepsi or Jack Daniels in a series of common pictures and presented the material to high and low drinking participants. High drinking participants not only remembered the Jack Daniels bottle better than the Pepsi bottle, but also had better memory for the Jack Daniels bottle than low drinking participants. Furthermore, high drinking participants who saw the Jack Daniels bottle showed a suppression effect for the subsequent three pictures. In a similar study, Griffin (2005; cf. Schmidt, 2012) presented pictures of spiders to spider phobics and nonphobic participants and found that the picture of spiders produced a suppression effect only in spider phobics.

According to Schmidt (2012), five central conclusions can be drawn from studies on emotional significance: 1) emotional material like words, pictures, or movies are well remembered due to increased attention and analysis devoted to the stimulus and the effect can be enhanced by distinctiveness, 2) memory for the gist of emotional events comes at the expense of peripheral aspects due to the limited capacity of working memory and other information processing systems, 3) the activation of the neuroendocrine system can mediate or moderate the impact of emotional events on memory, 4) the effect of emotional material on memory depends on whether it is positive, negative, or threatening, 5) the impact of a specific stimulus on memory depends on the outcome of appraisal or evaluation processes.

2.3 Event-Related Potential Correlates of Distinctiveness

Event-related potentials (ERPs) are averaged EEG responses time-locked to a stimulus that show a characteristic time-course, polarity, and peak and that are related to sensory, motor, or cognitive processing (Fabiani, 2006). In comparison to other neurophysiological methods like fMRI, ERPs have a high temporal resolution and several ERP components like the N100, the Mismatch Negativity, the P300, or the N400 are sensitive to manipulations of distinctiveness (e.g. Fabiani, 2006; Michelon & Snyder, 2006). Therefore, ERPs are well suited to shed light on the processes during the encoding of distinctive events (Fabiani, 2006; Michelon & Snyder, 2006).

2.3.1 N100 and Mismatch Negativity

One of the earliest ERP components which is sensitive to distinctiveness is the auditory N100, an ERP component with a negative peak around 100 ms. The auditory N100 is elicited by any auditory stimuli above the sensory threshold and decreases with repetition (e.g. Budd, Barry, Gordon, Rennie, & Michie, 1998; Rosburg, Weigl, & Sörös, 2014), but recovers after stimulus changes suggesting that non-distinctive, repeated items receive less processing (e.g. Sable, Low, Maclin, Fabiani, & Gratton, 2004). Similar effects also have been observed for the visual N100 (cf. Fabiani, 2006).

The auditory Mismatch Negativity (MMN), a negative ERP component with a fronto-cental peak around 200 ms, is elicited by any deviant tone in an otherwise homogeneous auditory tone series independent of the focus of attention. An MMN can be elicited by deviance in frequency, duration, intensity (e.g. Näätänen, Pakarinen, Rinne, & Takegata, 2004; Weigl, Mecklinger, & Rosburg, 2016), but also by deviance in complex sounds (e.g. (Näätänen, Schröger, Karakas, Tervaniemi, & Paavilainen, 1993) or stimuli in a different modality (de Gelder, Böcker, Tuomainen, Hensen, & Vroomen, 1999). Even though the N100 and the MMN are sensitive to (primary) distinctiveness, these components are typically not associated with the formation of episodic memories.

2.3.2 The P300 and the Subsequent Memory Effect

The P300 (also labeled P3 or P3b), first described by Sutton, Braren, Zubin, and John (1965), is an ERP component with a parietal peak around 300 ms (see Figure 2.2). In an oddball task, the P300 can be elicited by a rare or distinctive, task-relevant stimulus in a series of otherwise homogeneous stimuli (e.g. an “X” in a series of “O”s; e.g. Fabiani, 2006; Polich, 2007). For words and complex stimuli, the peak is often around 600 ms (e.g. Fabiani & Donchin, 1995; Kutas, McCarthy, & Donchin, 1977). The P300 is inversely proportional to the probability of the rare stimulus (Duncan-Johnson & Donchin, 1977) and requires attention (in contrast to the MMN). Therefore, the P300 is often treated as a neural correlate of primary distinctiveness (Fabiani, 2006; Michelon & Snyder, 2006) and has also been linked to the orienting response (Sokolov, 1963). Furthermore, emotional stimuli elicit larger P300 relative to neutral stimuli (see Hajcak, MacNamara, & Olvet, 2010, for reviews). In addition to the P300, emotional stimuli also elicit a late positive potential (LPP; e.g. Schupp et al., 2000), a sustained positivity which partially overlaps with the P300. Even stereotype violations can elicit a P300 (Osterhout, Bersick, & Mclaughlin, 1997). According to Donchin (1981), the P300 reflects context updating, i.e. an updating of the content of working memory due to a change in the environment.

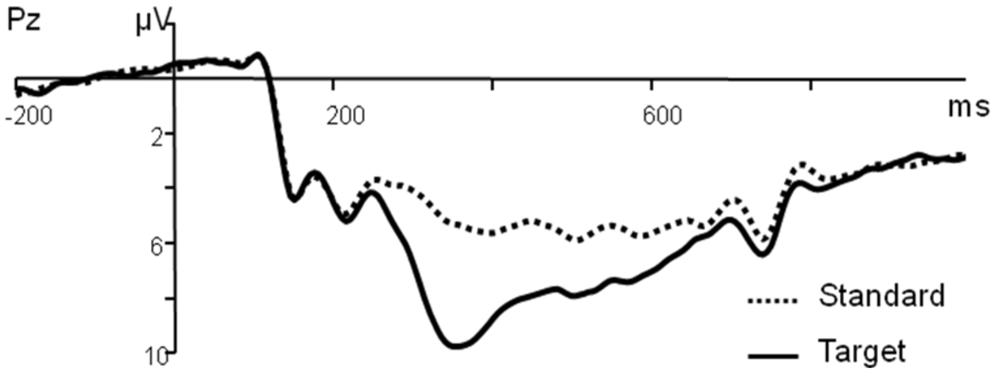

Figure 2.2 The P300 at the electrode Pz. The dashed line represents the frequent, task-irrelevant standard stimuli and the solid line represents the rare, task-relevant target stimuli. The figure is based on own unpublished data.

In addition to the classical P300, there is the novelty-P3 (or P3a) which has a frontal peak around 300 ms and is elicited by a task-irrelevant, novel stimulus and reflects an orienting response (cf. Polich, 2007). The novelty-P3 is typically tested in a novelty oddball, which contains rare, novel stimuli in addition to the rare target stimuli. The novelty-P3 is often accompanied by a P3b, but the neural generators differ between P3a and P3b as the P3a seems to involve both, the frontal lobe and the hippocampus (e.g. Knight, 1984, 1996).

Even though a lot of research has been conducted to identify the neural origins of the P300, a definite answer has not yet been found (Polich, 2007). However, there are indications that the posterior hippocampus and the dorsolateral prefrontal cortex contribute to novelty processing and the P3a (Knight, 1984, 1996). In contrast, lesions to the temporal-parietal junction diminish the P3b (Verleger, Heide, Butt, & Kömpf, 1994). In an fMRI study on perceptual, semantic, and emotional oddball stimuli, Strange, Henson, Friston, and Dolan (2000) found that there were two neural networks involved in deviance processing, a stimulus-independent network which includes the right prefrontal and the fusiform cortices and stimulus-dependent regions like the posterior fusiform regions for perceptual stimuli or the amygdala for emotional stimuli.

The effect of distinctiveness on encoding processes was often investigated by the analysis of so-called subsequent memory effects (SME; see Cohen et al., 2015; Fabiani, 2006; Friedman & Johnson, 2000, for reviews). For such an analysis, ERP data of items at encoding are sorted depending on whether items were later remembered or forgotten. The analysis of SMEs can provide insight in the neural foundation of cognitive processes that are responsible for successful storage (Cohen et al., 2015; Friedman & Johnson, 2000; Paller & Wagner, 2002). The P300 is considered to be an ideal candidate for the investigation of the link between primary distinctiveness and subsequent memory, because it does not only distinguish between isolated and nonisolated stimuli, but also between subsequently remembered and forgotten stimuli (e.g. Fabiani & Donchin, 1995; Fabiani, Karis, & Donchin, 1990; Kamp & Donchin, 2015; Kamp, Potts, & Donchin, 2015; Karis, Fabiani, & Donchin, 1984; Neville, Kutas, Chesney, & Schmidt, 1986; see Fabiani, 2006, for a review). The P300 reflects the encoding of item-specific information, whereas the encoding of inter-item associations is associated with a frontal slow wave, which appears between 900 and 2000 ms post-stimulus (e.g. Fabiani et al., 1990; Kamp et al., 2017).

For example, Fabiani and Donchin (1995) investigated the effect of physical and semantic distinctiveness on the P300 subsequent memory effect. For this purpose, participants encoded physically or semantically isolated words under a semantic or physical orienting task. Again, Fabiani and Donchin (1995) found that both, semantic and physical isolation led to a von Restorff effect and that the P300 was predictive for subsequent free recall performance. They also found that isolated words were predominantly reported at the beginning or at the end of recall suggesting organizational processes at retrieval (Bruce & Gaines, 1976). Fabiani & Donchin (1995) concluded that semantic isolation can indeed boost recall under conditions that stress item-specific processing even though semantic isolation might be deleterious for subsequent recall under relational processing. Further studies corroborated the link between positive ERPs and subsequent memory performance (e.g. Kamp et al., 2017; Otten & Rugg, 2001; Sanquist, Rohrbaugh, Syndulko, & Lindsley, 1980). In addition, some studies (e.g. Dolcos & Cabeza, 2002; Kamp, Potts, & Donchin, 2015) found a P300 SME for emotional stimuli indicating that the emotional stimuli were distinctive and that distinctiveness contributes to superior memory for emotional stimuli.

However, the P300 SME occurs only under certain circumstances. The P300 subsequent memory effect seems to be strongest in incidental or shallow encoding conditions (Michelon & Snyder, 2006). In the study by Karis et al. (1984), for example, a von Restorff effect was found for participants using rote memorization strategies, but not for participants using elaborate memorization strategies. Furthermore, the P300 predicted subsequent memory in participants using rote strategies. For participants using elaborate strategies, a frontal slow wave, but not the P300 predicted subsequent memory performance (Karis et al., 1984). This basic pattern of results was corroborated in two follow-up studies (Fabiani, Karis, & Donchin, 1986; Fabiani et al., 1990) which more systematically investigated the relationship between the P300 and the frontal slow wave, rote and elaborate encoding strategies, and subsequent memory performance. A similar pattern was obtained by Otten and Donchin (2000), who investigated whether central and peripheral feature of a distinctive event can lead to a von Restorff effect and a P300 SME. Centrality of the feature was manipulated by either varying the font size of the word (central feature) or placing a close or far frame around the word (peripheral feature). A von Restorff effect was obtained for all distinctive words irrespective of the centrality manipulation, but the P300 predicted subsequent memory only in the central condition. Furthermore, the P300 elicited by the frames was smaller than the P300 in the font size condition. However, the frontal slow wave predicted subsequent memory for the frame condition indicating that relational processing was necessary for binding the frame and the word in memory.

Together, these studies indicate that the P300 indexes primary distinctiveness. Furthermore, the P300 reflects context updating for distinctive items which is predictive for subsequent memory as long as only item-specific features are encoded (Fabiani, 2006; Fabiani & Donchin, 1995). Moreover, P300 SMEs can be obtained in recall and recognition memory tests. Subsequent rehearsal or elaboration as indexed by the frontal slow wave, however, can lead to the formation of inter-item associations which can overrule the mnemonic effects of the processing indexed by the P300 (e.g. Fabiani et al., 1990; Fabiani, 2006; Kamp et al., 2017).

2.3.3 The N400 and Subsequent Memory Performance

The N400, first described by Kutas and Hillyard (1980), is a negative ERP component with a centro-parietal peak around 400 ms (see Figure 2.3). It is elicited by semantic incongruity in sentences (e.g. “He spread the warm bread with socks”; Kutas & Hillyard, 1980), within word lists (e.g. Neville et al., 1986) or by the absence of semantic relations (e.g. “lipstick” and “waffle”; Voss & Federmeier, 2011). The N400 amplitude is diminished in an established semantic context (e.g. Kutas & Hillyard, 1980) or in word repetition (Olichney et al., 2000), but reinstated in terminal words with a low cloze probability (i.e. the subjective expectation). For words and sentences, the N400 often peaks prior to the P300, and reflects semantic integration or access to semantic memory (see Kutas & Federmeier, 2011, for a review). Furthermore, the N400 is often treated as the neural correlates of secondary distinctiveness (Michelon & Snyder, 2006), but it can be also elicited by primary distinctiveness (Fabiani & Donchin, 1995).

The literature is equivocal on whether there is an N400 SME (e.g. Mangels, Picton, & Craik, 2001; Meyer, Mecklinger, & Friederici, 2007; Neville et al., 1986). Several studies failed to establish a link between the N400 and subsequent memory performance (e.g. Fabiani & Donchin, 1995; Neville et al., 1986). The P300, rather than the N400 amplitude, seems to be predictive for recall of semantically distinctive items (e.g. Fabiani & Donchin, 1995). Furthermore, Bartholow et al. (2001) found that cued-recall performance was next to zero for incongruous words indicating that words which elicit a strong N400 are often hard to remember. At first sight, these findings seem to be at odds with studies reporting superior memory for bizarre events (bizarreness effect; see Schmidt, 1991, for a review). However, behavioral research showed that bizarreness influences memory in both, a facilitative and disruptive way (see Schmidt, 2006, 2012, for a review). Fabiani (2006) speculates that the N400 might have an effect on confusion or response bias rather than on recall or recognition accuracy.

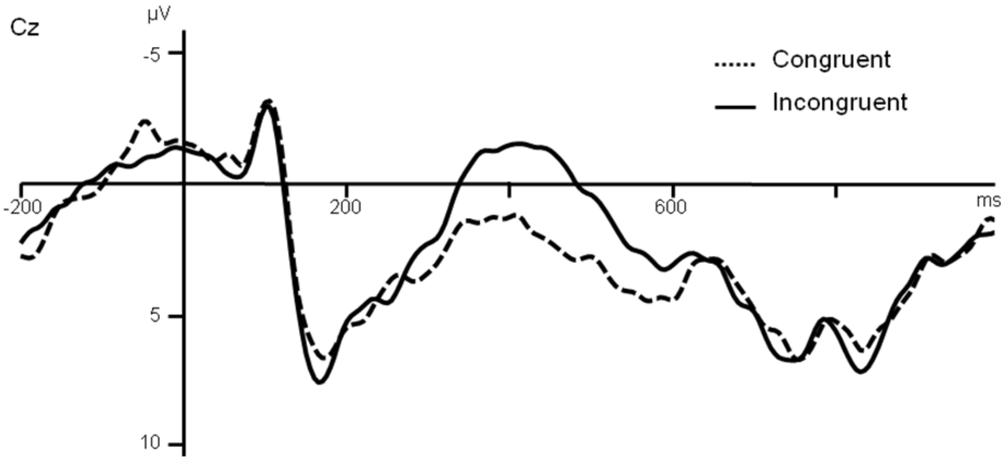

Figure 2.3 The N400 at the electrode Cz. The dashed line represents the semantically congruent words and the solid line represents the semantically incongruent words. The figure is based on own unpublished data.

However, more recent studies provide some evidence that the N400 might be a likely candidate for supporting subsequent recognition rather than subsequent recall (e.g. Kamp et al., 2017; Mangels et al., 2001; Meyer et al., 2007). Recognition memory is supported by two independent processes: familiarity and recollection (Mandler, 1980; Yonelinas, 2002; Yonelinas, Aly, Wang, & Koen, 2010). Whereas familiarity is considered to be an automatic, fast, and context-free recognition process based on the degree of a feeling of oldness, recollection is considered to be a slower, more deliberate all-or-none recognition process in which recognition judgment is based on the retrieval of context information. Meyer et al. (2007) found that the N400 at encoding correlated with the FN400, an ERP index of familiarity. Friedman and Trott (2000) applied the remember/know (R/K) procedure in their subsequent memory experiment and found SMEs for subsequently remembered and subsequently known items in the ERPs starting around 300 ms in their sample of older participants. Mangels, Picton, and Craik (2001) used the R/K procedure and not only found an SME which differentiates between remember and know, but also a negative going ERP component in the time range of the N400 (dubbed "N340" by the authors) which predicted subsequent recognition on the basis of familiarity and recollection. These studies indicate that the N400 might predict familiarity-based recognition.

Studies using intracranial recordings also revealed an N400-like potential in the parahippocampal cortex and a later positive going potential in the hippocampus (Elger et al., 1997; Fernández et al., 1999; see Friedman & Johnson, 2000, for a review). These findings are consistent with data indicating that the hippocampus is related to recollection and the perirhinal cortex to familiarity (Aggleton & Brown, 1999; Yonelinas et al., 2010).

In more recent studies by Kamp and colleagues (e.g. Kamp et al., 2015, 2017), further evidence for an N400 SME was obtained. Kamp et al. (2017) presented word pairs in two encoding conditions. In the definition condition, the word pairs were presented together with a definition of the concept formed by the word pair. In the sentence condition, participants had to mentally insert the words in a sentence with blanks. They found an N400-like SME only for the definition condition. Since the definition condition should promote unitization and allows familiarity-based recognition of associative information (e.g. Bader, Mecklinger, Hoppstädter, & Meyer, 2010), the findings by Kamp et al. (2017) are consistent with the interpretation that the N400 might predict familiarity-based recognition. An N400 SME was also found in the study by Kamp et al. (2015), in which positive, negative, or neutral words were used as isolates or as nonisolates in a von Restorff paradigm. Kamp et al. (2015) found an N400 SME for emotional words. For negative stimuli, smaller N400 amplitudes were associated with higher likelihood of recall, whereas the reverse was true for positive stimuli. This effect, however, is difficult to interpret, because the effect of emotional significance on the N400 is less clear than the effect of emotional significance on the P300 (e.g. Hajcak et al., 2010).

To sum up, the studies reviewed in this section highlight that the N400 SME is of a more elusive nature than the P300 SME. However, studies like Kamp et al. (2017) or Mangels et al. (2001) indicate that an N400 SME can be obtained under certain conditions. Given its relation to semantic processing, the N400 might be a putative predictor for familiarity-based recognition rather than a general indicator of subsequent memory.

2.4 Shared Distinctiveness, Memory, and Illusory Correlation

In most studies reviewed so far, only one feature was manipulated to render an item distinctive. These studies advanced our understanding of the relationship between distinctiveness and memory, but bear only limited value for a discussion of shared distinctiveness. Since shared distinctiveness plays an important role in the distinctiveness-based IC (Hamilton & Gifford, 1976), this section reviews studies, in which items contained two or more distinctive features, and relates them to the distinctiveness-based IC.

The additive effects of distinctiveness in the von Restorff paradigm were investigated in several studies. Bruce and Gaines (1976) combined high priority with primary distinctiveness. Participants were instructed to especially attend uppercase items, which were presented in a list together with lowercase words. An uppercase word was additionally isolated by brackets. Bruce and Gaines (1976) found that the isolated uppercase word was remembered better than the other attended uppercase words in the list. In a similar study by Hunt and Mitchell (1982), participants were presented a list with an orthographically distinctive item, a categorically distinctive item, or an orthographically and categorically distinctive item. Hunt and Mitchel (1982) found that memory was superior for distinctive items as compared to control items. Critically, memory for orthographically and categorically distinctive items was better than memory for items with only one distinctive feature.

Emotional significance can also change the effect of isolation in a von Restorff paradigm. Kamiya (1997), for example, reported that the von Restorff effect for items isolated by color was enhanced, when the isolated items were emotional. Similar results were obtained by Kuhbandner and Pekrun (2013), who combined primary distinctiveness with emotional significance and high priority stimuli. A word in the middle of the list was isolated by the colors red, green, or blue. Furthermore, the isolated words had a positive, a negative, or a neutral meaning. Kuhbander and Pekrun (2013) found a von Restorff effect for all colors. However, the von Restorff effect was enhanced for negative isolates presented in red and for positive isolates presented in green. Since the color red signals negativity (e.g. wrong) and the color green signals positivity (e.g. correct), the enhancement arises from an interaction between emotional significance, primary distinctiveness, and high priority.

There are also ERP studies on emotional significance in a von Restorff paradigm. In the study by Kamp et al. (2015), participants learned lists with positive or negative isolates surrounded by neutral words and lists with neutral isolates surrounded by either positive or negative words. A von Restorff effect was observed for negative, but not for positive or neutral isolates. Furthermore, recall for neutral isolates was even worse than recall for positive standards. Negative standards and isolates elicited a P300 which was predictive for subsequent recall indicating that the distinctiveness of negative words contributed to their superior memorability. Bartholow et al. (2001) presented their participants fictional person descriptions with a given trait followed by sentences describing the person’s behavior which was either consistent or inconsistent with the trait, neutral, or semantically incongruous. Behaviors violating the participants’ expectation not only elicited a larger P300, but were also recalled more often than expectancy-consistent behavior indicating that the inconsistent behavior was more distinctive than the consistent behavior. This pattern was stronger for negative behaviors than for positive behaviors. In contrast, recall for semantic incongruity was poor. In a subsequent study, Bartholow, Pearson, Gratton, and Fabiani (2003) investigated the effect of alcohol intoxication on person perception and memory. Again, recall for negative inconsistent behavior was better than for positive inconsistent behavior under placebo conditions. In the alcohol condition, however, positive inconsistent behaviors were better recalled than negative inconsistent behaviors.

But the combination of emotional significance and primary/secondary distinctiveness might not only lead to a memory advantage for emotional stimuli, but also affect the processing of surrounding items (Schmidt, 2006, 2012). In an emotional oddball paradigm by Hurlemann et al. (2005), a color photograph was embedded in a series of black-and-white line drawings. The photograph could be either positive, negative, or neutral. At the bottom of each picture was a corresponding label. A total of 36 lists à 8 items were presented. In comparison to neutral stimuli, negative and positive stimuli led to a suppression effect for subsequent stimuli. However, positive stimuli increased memory for items preceding the photograph, whereas negative stimuli decreased memory for the preceding item. Similar results were obtained by Schmidt (2002) who presented photographs of nude and clothed persons as isolates or nonisolates. The isolation effect for the nude exceeded the isolation effect for the clothed person. At the same time, memory for background details and subsequent pictures was worse for isolated nudes than for isolated clothed persons indicating that participants needed some time to recover from the presentation of the nude. Significant and distinctive stimuli, thus, attract attention leaving less capacity for processing background details and the following stimuli (Schmidt, 2006, 2012).

Hence, there is ample evidence in the memory literature that shared distinctiveness can further enhance the memory benefit for distinctiveness. The SDA places shared distinctiveness in a central position in its theoretical framework of the distinctiveness-based IC (Hamilton et al., 1985; Hamilton & Gifford, 1976). The shared distinctiveness of category combinations with two infrequent attributes, usually the negative behavior of minority members, should lead to better encoding and memory as compared to more common category combinations. In contrast to the clear-cut results from the memory studies reviewed above, the shared distinctiveness effect has been shown to be much more elusive in studies on the distinctiveness-based IC.

There are IC studies using cued recall (see Mullen & Johnson, 1990, for a review), free recall (e.g. Hamilton et al., 1985), or one-shot ICs (Risen et al., 2007), which support the heightened memory hypothesis of the SDA. More indirect evidence for shared distinctiveness in the IC paradigm comes from studies using reading and reaction times. Reading times were longer for shared distinctive items than for other items indicating that participants spent more time on encoding shared distinctive items (Stroessner et al., 1992). Furthermore, reactions to shared distinctive items were facilitated at retrieval indicating that these items were more available in memory (Johnson & Mullen, 1994; McConnell et al., 1994).

This evidence has been challenged by other memory studies (e.g. Fiedler et al., 1993; Klauer & Meiser, 2000). Klauer and Meiser (2000) included novel distractors in their group assignment task and found that participants were better in discriminating between old and new, when the behavior was negative than when it was positive. Furthermore, participants were biased to assign positive items to the majority. Critically, source monitoring performance was equal for all category combinations. These results were replicated in other studies (Bulli & Primi, 2006; Meiser & Hewstone, 2001). In addition, Bulli und Primi (2006) reported that old/new discrimination was worse for positive items of the majority and negative items of the minority than for negative majority items and positive minority items.

Typically, participants in the Hamilton and Gifford (1976) IC paradigm are instructed to memorize the items for a subsequent memory test. However, the type of instruction affects both, the extent of IC and the memory performance. Pryor (1986) compared a memorization condition as used by Hamilton and Gifford (1976) with an online impression condition, in which participants were instructed to form an impression of each group. Pryor (1986) reported that memory was comparable in an impression formation condition and a memory condition, even though an IC was found only in the memory condition. Consistent with Klauer and Meiser (2000), Pryor found a stronger bias in favor for the majority. However, memory for negative minority items was elevated in both conditions. In a similar vein, Meiser (2003) found that ICs were diminished in an impression formation condition as compared to a memorization condition. However, memory was better in the impression formation condition than in the memorization condition.

Furthermore, ICs can be observed even in the absence of distinctive or infrequent information, i.e. the frequency of shared distinctive items was zero (Fiedler, 1991; Shavitt et al., 1999; Van Rooy et al., 2013). The study by Van Rooy et al. (2013) is particularly informative, because they presented four groups with decreasing group size and found that memory was enhanced for negative behavior of the smallest group. However, an IC was also observed in a condition without distinctive items in the smallest group.

To sum up, the shared distinctiveness effect, which was often observed in the memory literature, has been difficult to replicate in the social cognitive literature of ICs. The discussion of shared distinctiveness effects in the IC paradigm is further complicated by the fact that researchers rarely provide a precise definition of distinctiveness based on the memory literature or use existing definitions as those provided by Schmidt (1991, 2012). At a closer look, several discrepancies between the designs in IC studies and memory research become apparent. First of all, most studies on distinctiveness rely on free recall, whereas the cued recall task in the IC paradigm is more similar to a source monitoring task. Therefore, the responses in the memory task from the IC paradigm are more susceptible to response bias. Second, studies on shared distinctiveness in memory research typically use fewer distinctive items than studies in IC research. Third, most memory studies control for primacy and recency effects. In contrast, this problem is rarely discussed in IC studies. Finally, a control condition without distinctive items is usually included in memory studies of shared distinctiveness and the effect of distinctiveness is evaluated with respect to the control condition. In IC studies, however, no comparable control condition exists. This further complicates the assessment of the relationship between episodic memory and ICs. The present thesis will address some of these points in order to determine the role of episodic memory in covariation assessment.

2.5 Research Questions

From the evidence reviewed in this section, it becomes clear that the IC is a robust and systematic bias in subjective covariation assessment, but the mechanisms underlying ICs are still a matter of debate. Episodic memory is supposed to play a pivotal role in the SDA, which explains ICs by differential accessibility of infrequent and distinctive group-behavior combinations after encoding (Hamilton et al., 1985; Hamilton & Gifford, 1976). The ILA, in contrast, explains ICs as a result of regression to the mean due to noise in memory channels that especially affects the infrequent group (Fiedler, 1991, 2000). A third mechanism proposed to underlie ICs is category accentuation as proposed by the accentuation account (McGarty et al., 1993) and AT (Sherman et al., 2009). Attention should lie on both the most frequent and the least frequent category combination in order to maximize the differentiation between the categories.

The main objectives of the present thesis are to determine a) the role of episodic memory in human covariation assessment and b) the relative merit of the three mechanisms, namely shared distinctiveness, information loss, and accentuation, by the use of both, behavioral and ERP methods. For this purpose, we conducted six experiments.

In the Experiments 1 and 2, infrequency was contrasted with distinctiveness. More precisely, we tested whether ICs can be observed not only under conditions with skewed frequency ratios (1:2 ratio for desirable and undesirable behavior), but also under conditions with equated category frequency for valence (1:1 ratio for desirable and undesirable traits). Equated category frequencies preclude regression to the mean, but allow differential accessibility due to emotional significance, a form of distinctiveness according to Schmidt (2012). An IC under such conditions would be clearly incompatible with the ILA, but compatible with the SDA. Experiment 3 used a methodologically refined source memory task in order to test whether memory for distinctive category combinations was enhanced as predicted by the SDA.

Experiment 4 was designed to compare the SDA with the accentuation approach to ICs by using an oddball paradigm (Polich, 2007). Since the P300 is an ERP component closely linked to the processing of distinctiveness (e.g.

Fabiani, 2006), the SDA would predict a linear increase of the P300 amplitude from the least frequent to the most frequent category combination. In contrast, the accentuation hypothesis would be compatible with larger P300 amplitudes for the most frequent and the least frequent category combinations as compared to the moderate frequent category combinations (Sherman et al., 2009).

Based on the SME literature reviewed in the sections 2.3.2 and 2.3.3, we applied the subsequent memory paradigm in Experiment 5 in order to test whether and how the N400 and the P300 were related to subsequent memory performance. ERPs were recorded to physical and semantic isolates as well as control items at encoding. We then tested whether these items were later recognized on the basis of recollection or familiarity using the remember/know procedure. Thus, we hypothesized that N400 is related to familiarity-based recognition and that P300 is linked to recollection-based recognition, which shares some similarities with free recall.

We conducted Experiment 6 in order to further illuminate the relationship between the P300, (shared) distinctiveness, memory, and IC. In this final study, we combine the insights gained from the previous five experiments and the literature reviewed in this section in order to test whether perceived distinctiveness as indexed by the P300 can be related to subsequent memory and the amount of perceived covariation between two features.

3 Behavioral Investigations on the Distinctiveness-Based Illusory Correlation

3.1 Introduction

The SDA explains ICs by differential accessibility of distinctive group-behavior combinations after encoding, whereas the ILA states that the primary source of ICs is regression to the mean due to noise in memory channels which especially affects the infrequent group (see sections 2.1.2.1 and 2.1.2.2, respectively). Even though the SDA and the ILA are not mutually exclusive, it is necessary to test boundary conditions to judge the relative merit of both accounts. One highly important boundary condition arises naturally from a closer look at the regression to the mean argument of the ILA. An illusory correlation would be expected only if the frequency distribution for behavior is skewed (33% and 67% condition in Figure 2.1), because only these circumstances allow differences in regression between the majority and the minority. Thus, no IC would be expected when positive and negative behaviors are equally distributed (50% condition in Figure 2.1).

A reassessment of the classical IC paradigm reveals that several kinds of distinctiveness are involved: Minority members are distinctive in the context of presentation due to infrequency (primary distinctiveness). Negative behavior is not only distinctive due to rarity in the stimulus set (primary distinctiveness), but also per se due to its rarity in real life (secondary distinctiveness) and unpleasantness (emotional significance) as documented by the literature on the asymmetry between positive and negative information (Fiske, 1980; Kanouse, 1984). Furthermore, positive information is more similar than negative information which additionally renders negative information distinctive (Alves, Koch, & Unkelbach, 2016; Alves et al., 2017; Koch, Alves, Krüger, & Unkelbach, 2016). Thus, equating the category frequencies for negative and positive stimuli allows dissociating primary distinctiveness (resulting from infrequency) from other forms of distinctiveness.

In such a setting with equated frequencies, the SDA would still predict the presence of an IC, whereas the ILA would predict its absence, because regression to the mean would equally affect positive and negative items as illustrated (Figure 2.1).

The first and most important research question of the present study was, thus, whether an IC could be observed in a condition with equated frequencies (with positive and negative items being equally frequent). For this purpose, findings in the equated frequency condition were compared to a condition with skewed frequencies (i.e. the standard paradigm with positive items twice as frequent as negative items). To the best of our knowledge, no previous study has directly addressed such a comparison.

A second research question concerned the relationship between memory and IC. Both, the SDA and the ILA, assume a causal contribution of memory in the formation of ICs, even though they assign different roles to memory. Based on the SDA, memory should be better for the minority than for the majority and best for negative items of the minority. Moreover, the better memory for such distinct group-behavior combinations should be predictive for the extent of ICs. In contrast, the ILA would predict similar memory performance for all group-behavior combinations, because the random noise is supposed to affect all of them in the same manner.

The third research question of the present study was whether ICs would remain stable over time – a question still understudied. Feldman et al. (1986) investigated ICs using delays of 12 or 24 hours. Unfortunately, the results of this study are inconclusive, because their paradigm failed to induce a reliable IC in the first place. Distinctiveness effects on episodic memory are relatively stable over time (e.g. Hunt, 2009). Thus, on the basis of the SDA, we would expect the IC to be unaffected by a time delay. However, it also seems plausible to assume that the distortion of the encoded material due to information loss increases over time. Memory steeply declines over a short range of time after initial encoding (Ebbinghaus, 1885/1966; see Rubin & Wenzel, 1996 for an extensive review). Thus, introducing a short delay after initial encoding should already lead to some information loss and regression to the mean. The extent of IC as a function of noise presumably follows an inverted u-shaped pattern. As long as estimates for both, the majority and the minority, can still regress, increases in noise will increase the extent of IC, because more information will be lost for the minority than the majority. If information about the minority is almost completely lost and only the estimates for the majority can show further regression, then increases in noise will reduce the difference between both groups and the extend of IC (see also Fiedler, Russer, & Gramm, 1993). Thus, while the SDA predicts that a time delay does not affect ICs, the ILA predicts that a short time delay after establishing an IC might lead to an increase of ICs.

As a first step, we tested our theoretical assumptions on the impact of equated category frequencies and temporal delay in a computer simulation based on the Brunswikian Induction Algorithm for Social cognition (BIAS; Fiedler, 1996, 2000; Fiedler & Walther, 2004). BIAS implements the assumptions of the ILA that the mere unbiased processing of information in a noisy environment is sufficient for the formation of ICs. Subsequently, we tested the predictions derived from the simulation in two behavioral experiments – one with the standard skewed distribution (Experiment 1) and the second with an equal distribution (Experiment 2).

3.2 Simulation study

3.2.1 Method

BIAS applies the principles of information loss and aggregation on stimulus matrices to simulate cognitive biases. Computationally, the BIAS model assumes that information is represented in a stimulus matrix in which the columns represent individual stimuli or events and rows represent cues (e.g. features of a stimulus), i.e. a stimulus is not a scalar value, but a vector (Figure 3.1; Fiedler, 1996, 2000). Each stimulus vector is derived from the distal entity. The distal entity is a vector that defines the true values, i.e. how the values of the stimulus would look like in an ideal, noise-free environment. Due to noise (e.g. misperceptions, forgetting, natural variations) the stimulus vector will deviate from the distal entity, i.e. some elements of the stimulus vector randomly differ from the distal entity. For example, red and white roses are both instances of the distal entity roses and, therefore, resemble each other even though genetic and ontogenetic variations alter their appearance to some degree. The error variance due to noise is removed by applying an aggregation rule (e.g. averaging or summation) on each row. If there is a prevailing tendency in the data, it will become apparent after aggregation. If, for example, a person is asked about his/her attitude towards a product, this person will aggregate over arguments for and against the product. The attitude would be positive, if there are more pros than cons (Fiedler, 1996).

In our simulation, a stimulus was represented by a vector with ten elements (Figure 3.1; Fiedler, 1996). Each element represented a feature or cue. The presence of a feature was coded as 1, and the absence of a feature was coded as -1. The ten elements represented the valence of the trait. Positive and negative traits were coded as the exact opposite of each other. In order to explore the impact of the noise level on the extent of IC, either one, two, three, four, or five elements (corresponding to total information loss) were randomly chosen and reversed.

For the simulation of the skewed frequency condition, the simulation included 16 positive and 8 negative behaviors for the majority, as well as 8 positive and 4 negative behaviors for the minority. For the simulation of the equal frequency condition, the simulation included 12 positive and 12 negative behaviors for the majority, as well as 6 positive and 6 negative behaviors for the minority. In order to assess the prevailing tendency for each run, the stimuli were summed up row-wise for each group separately and the resulting aggregated vector was correlated with the predefined ideal vector for positive stimuli (Figure 3.1). A high correlation means that the aggregate corresponds closely to the ideal vector. The average correlations scores of the two groups across 1000 simulations were compared by paired t-tests. In these simulations, an IC was judged to be present, if the aggregate of the majority correlated significantly higher with the ideal vector of positive stimuli than the aggregate of the minority. All simulations were run in R 3.3.1 using custom-written R code.

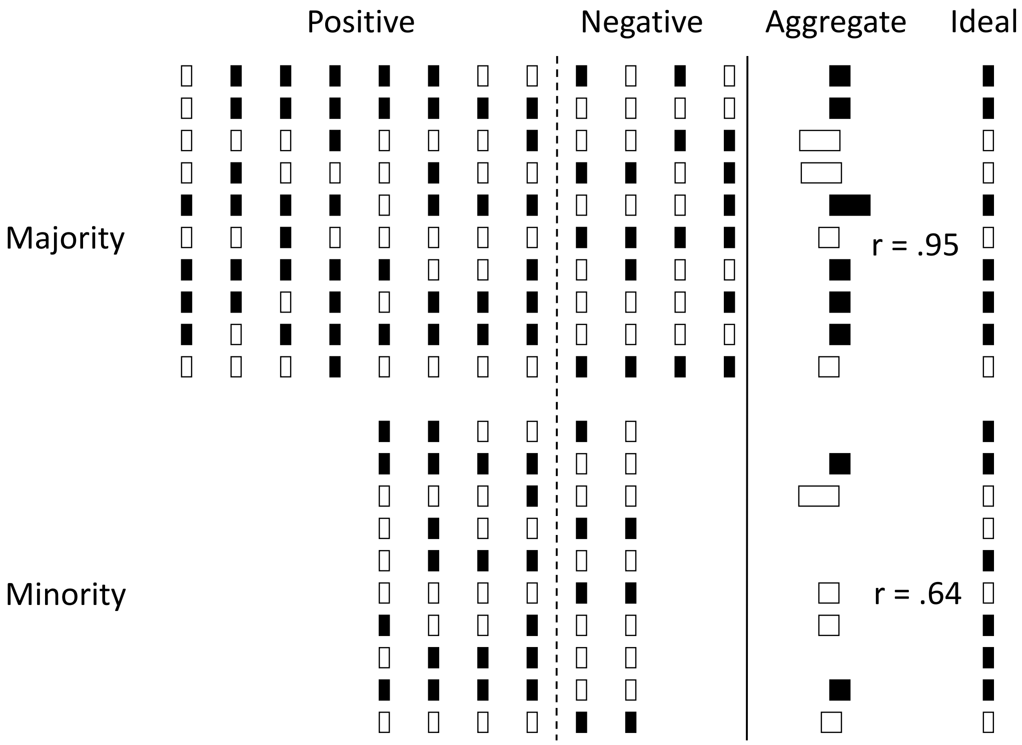

Figure 3.1 Illustration of the aggregation process in BIAS. The majority is represented by 12 stimuli – 8 positive and 4 negative. The minority is represented by 6 stimuli – 4 positive and 2 negative. During the aggregation process, each row is summed up. Black rectangles are counted as +1 and white rectangles as -1. The aggregate vector is then correlated with the ideal type of positive behavior. The aggregate of the majority correlates more closely with the ideal type than the aggregate of the minority. Adapted from Fiedler (1996, Figure 2).

3.2.2 Results

The results of the computer simulation are depicted in Figure 3.2. As hypothesized, there were ICs in the skewed frequency condition for the noise levels one to four, i.e. the correlation between the ideal vector for positive behavior and the aggregate vector was higher for the majority than for the minority (all *t*-values > 21.03, *p* < .001). Thus, the overall positivity became more apparent for the majority than the minority at these noise levels. As expected, no IC was observed at the highest noise level when all five elements were changed ($t(999) = 0.10$, $p = .917$), because all information was erased by noise. Also as predicted, the extent of ICs initially rose with increasing noise levels and then fell for more extreme levels of noise: The ICs of all five noise levels significantly differed from each other (all p-values < .008). In contrast, for the equated frequency condition, there were no significant ICs at any noise level ($|t(999)| < 1.06$, $p > .289$) and no differences between the five noise levels either (all p-values > .330).

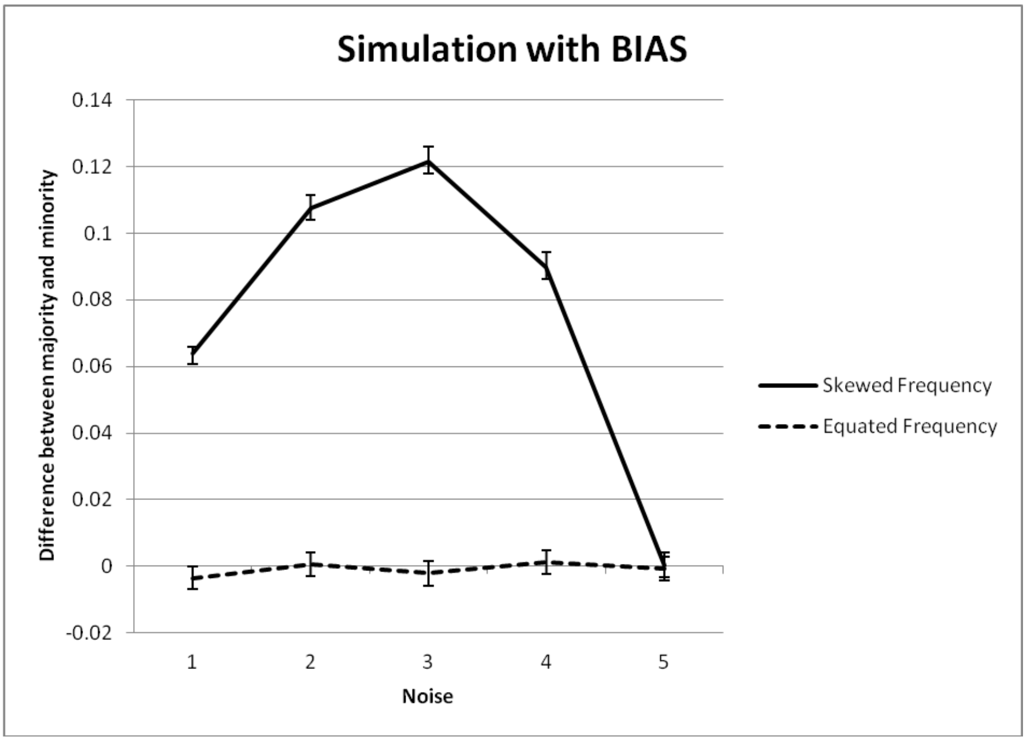

Figure 3.2 Results of the computer simulation with BIAS. The illusory correlation is depicted as difference between the majority's correlation with the ideal vector and the minority's correlation with the ideal vector for the skewed frequency condition (solid line) and the equated frequency condition (dashed line). In the skewed frequency condition an illusory correlation was present at all noise levels but the 5th. In the equated frequency condition, the illusory correlation was absent at all noise levels.

To sum up, the simulations showed that ICs were only present in the skewed frequency condition and the extent of ICs initially increased with increasing noise and then declined with extensive noise levels. These results have two implications for our experiments: 1) If nothing but information loss exerts an effect on judgment, an IC should be observed in Experiment 1 (skewed frequency condition), but not in Experiment 2 (equated frequency condition). 2) If information loss is moderate, a short delay should increase the extent of IC. Thus, if the assumptions of the ILA hold, the pattern of the behavioral data should resemble the pattern found in the simulations.

3.3 Experiment 1 and 2: Skewed versus Equated Frequency Condition

3.3.1 Experiment 1: Skewed Frequency Condition

3.3.1.1 Method

3.3.1.1.1 Participants

Thirty-four students (26 female) of the Saarland University participated in Experiment 1. Participants received course credit or comparable compensations for their participation. Data of six participants were not included in the analysis. One participant did not use the correct response keys. The five other excluded participants were either not German native speakers (n = 2), who were granted participation in the study in order to obtain course credits, or were participants who inferred the hypotheses of the study as reported in the post-experimental questionnaire (n = 3). The latter three participants reported quite specific hypotheses about the purpose of the experiment (e.g. that the experiment was about distinctive features of minorities or about the development of stereotypes about minorities, or even explicitly mentioned ICs) and were excluded on the grounds of previous findings: the awareness of influence can hamper the investigation of judgmental biases (see Bless, Fiedler, & Strack, 2004, pp. 122-124, for a discussion). Specifically to the IC, knowledge about the task has been shown to reduce the extent of IC (e.g.Chapman, 1967; Lilli & Rehm, 1983, 1984). Thus, from our point of view, the inclusion of these participants would diminish the validity of the data. The exclusion of the three participants did, however, not affect the major findings. The final sample comprised 28 participants (23 female, median age 23 years, range 18-31 years).

3.3.1.1.2 Materials

For Experiment 1, 24 positive and 12 negative adjectives were drawn from a pool of 48 trait adjectives and matched for arousal, imageability, and word length (see Appendix B. for details). Descriptions of traits were chosen instead of descriptions of behavior, because by this approach sentence lengths and, thus, encoding times could be equated. Thirty-six German male first names that were popular in the time period from 1986 to 1993 were selected from a website (http://www.beliebte-vornamen.de/). We selected this time period to ensure that the participants were familiar with the first names. Consistent with other studies on the IC (e.g. Hamilton & Gifford, 1976; Stroessner et al., 1992), the names were restricted to male names in order to control for possible effects of the stimulus persons' gender.

For each participant, a list of 36 person descriptions was created. The 36 first names and trait adjectives were randomly combined to a description and assigned either to the majority, group A, or to the minority, group B. As shown in Table 1, sixteen descriptions with positive traits and eight descriptions with negative traits were assigned to group A, eight descriptions with positive traits and four descriptions with negative traits were assigned to group B.

Table 3.1 Distribution of positive and negative traits across groups for Experiment 1 and 2.

	Experiment 1		Experiment 2	
	Positive traits	Negative traits	Positive traits	Negative traits
Majority	16	8	12	12
Minority	8	4	6	6

3.3.1.1.3 Procedure

The experiment was programmed and run using E-Prime 2.0. Participants sat in front of a 17 inch monitor and were individually tested. All displays were centered and had white background. Words and sentences were presented in black 18 pt Courier New font. The instruction followed those of Hamilton and Gifford (1976) and Pryor (1986). The experiment was described as being concerned with how people perceive and retain information about others. Participants were told that they would read descriptions of students made by persons close to them and that each person belonged to one of two groups, which actually existed and were arbitrarily named group A and group B for the purpose of the experiment (see Table 3.2 for an overview over the experiment).

Table 3.2 Overview over the procedure in Experiment 1 and 2. The tasks were presented in the order in which they are listed (except the frequency estimation and the evaluative trait rating for which the order was counterbalanced across subjects). The list below each task summarizes the measures we derived from this task.

Encoding Task
Filler Counting Task (1 min.)
Immediate Testing • Group Assignment (assessing bias against the minority, illusory correlation, and source memory performance) • Frequency Estimation (assessing bias against the minority and illusory correlation) • Evaluative Trait Rating (as in Frequency Estimation)
Filler Task (40 min.)
Delayed Testing • Group Assignment (assessing bias against the minority, illusory correlation, and source memory performance) • Frequency Estimation (assessing bias against the minority and illusory correlation) • Evaluative Trait Rating(as in Frequency Estimation)
Sentence Valence Rating (calculating the encoded correlation)
Control Questionnaire

During encoding participants saw 36 descriptions. They were instructed to read the descriptions and memorize all the information for a subsequent memory test. Each trial started with a fixation cross presented for 500 ms. Then a blank appeared for 200 ms. Next, the descriptions appeared on the screen for 4000 ms. Each description contained a male first name, the group membership, and a positive or negative trait (e.g. “Oliver from group A is nice.” or “Andreas from group B is stubborn.”). The descriptions were presented in a random order. At the end of the trial another blank was presented for 200 ms.

After the encoding task participants had to count backward from 100 in steps of three for one minute and to enter the number they reached. Then they started the group assignment task. This task was intended to assess the bias against the minority and probe the source memory for group membership. Participants were told to assign the descriptions to one of the groups as fast and accurately as possible. Each trial started with a blank presented for 200 ms. Next, the sentences from the encoding task were presented again in random order without group information (e.g. “Oliver is nice.”). Participants could respond for 2500 ms by pressing the F or the J key of the keyboard. Response keys to group assignments were counterbalanced across participants. Only when participants did not respond within the given time window, a feedback screen appeared for 500 ms, reminding the participant to respond faster on the next trials. Each trial ended with another blank shown for 200 ms.

Next, participants filled out computerized versions of the frequency estimation task, in which participants estimated the percentage of negative traits for each group separately, and the evaluative trait rating, in which participants rated each group separately on ten traits (*helpful*, *tolerant*, *sociable*, *affectionate*, *honest*, *ingenious*, *friendly*, *unreliable*, *industrious*, and *irresponsible* taken from Fiedler et al., 1993) using a 10-point rating scale. The frequency estimation task and the evaluative trait rating as well as the group labels were counterbalanced across participants.

For about 40 min participants conducted an unrelated filler task. After this delay participants performed a second time the group assignment task, the frequency estimation task, and the evaluative trait rating in the same sequence as in the first part of the experiment. After these tasks, the sentences of the encoding task were presented again. Participants had to rate each sentence by using the keys 1, 2, and 3 whether the sentence was negative, neutral, or positive. This sentence valence rating served as a manipulation check to ascertain that the sentences were accurately encoded according to our intended design. When participants pressed one of the keys, a blank was presented for 400 ms followed by the next sentence.

At the end of the experiment the participants were given a questionnaire which asked for participants' sensitivity for the hypotheses and related control questions.

3.3.1.1.4 Data Analysis

At both, immediate and delayed testing, group assignments, frequency estimation, and evaluative trait rating were used to assess the amount of bias against the minority. In order to control for Type I errors while at the same time preserving statistical power, these measures were subjected to a multivariate analysis of variance (MANOVA) with the factors Group and Time point. We used the difference between the number of positive and negative descriptions assigned to each group in the group assignment task, the estimated relative frequency of negative behavior for each group, and the mean evaluative trait rating scores for each group as dependent measures. Significant main effects or interactions were followed up with univariate analyses of variance (ANOVA).

Source memory performance was used to test for better memory of shared-distinctive items. Memory was assessed by calculating the unbiased hit rates (Wagner, 1993) for the valence condition (positive vs. negative) and the two time points (immediate vs. delayed) separately (see Appendix A. for details). Unbiased hit rates are the conditional probability that the stimulus is correctly classified given the stimulus is shown (e.g. correctly assigning "A" to a group A item) multiplied with the conditional probability that the correct response category is chosen given this response category (e.g. correctly responding "A" when responding "A").

The unbiased hit rates were transformed with an arcsine transformation (Wagner, 1993) to ensure normal distribution and entered in a repeated measure ANOVA with Group (majority vs. minority), Valence (positive vs. negative), and Time point (immediate vs. delayed) as independent variables.

In order to provide a comprehensive picture of the memory performance, we also analyzed the memory data using Pr, a measure for discrimination, and Br, a measure for response bias (Snodgrass & Corwin, 1988). Pr and Br were calculated for negative and positive valence and for the two time points (immediate and delayed) separately and entered in a repeated measure ANOVA with Valence (positive vs. negative) and Time point (immediate vs. delayed) as independent variables.

In order to quantitatively assess the extent of ICs, correlation coefficients were calculated individually for each participant, dependent variable, and time point separately: Phi coefficients for group and valence were calculated from the group assignment and the frequency estimation data. Point-biserial correlations for group and evaluation were calculated from the evaluative trait rating data from the learning phase and the sentence rating data at the end of the experiment. The latter correlation coefficient represents the encoded correlation. With it, we can not only ensure that participants accurately encoded the sentences according to our experimental design (i.e. that participants interpreted a negative description as negative). We can also preclude that the IC was already formed during encoding (i.e. that erroneous encoding acts as a source of the noise according to the ILA), which might hamper the interpretation of the retrieval data. Please note that the phi coefficient and the point-biserial correlation are both simplified versions of the Pearson correlation coefficient and are therefore comparable with each other. Positive correlations indicate that the majority is associated more with positive descriptions and the minority more with negative descriptions. Fisher z-transformation was applied to all correlation coefficients for statistical tests. All correlation coefficients were subjected to a MANOVA to test for difference from a null vector. If a significant effect was observed, follow-up one-tailed t-tests were conducted.

All statistical analyses were conducted using IBM SPSS 24 (IBM Corp., Armonk, NY, USA). Effect sizes for t-tests were calculated with G*Power 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009). The alpha criterion was set to p =

.050 for all analyses. The Benjamini-Hochberg procedure for multiple comparisons was used to adjust the p-values in follow-up tests.

3.3.1.2 Results

3.3.1.2.1 Group Assignments, Frequency Estimation, and Evaluative Trait Ratings

The MANOVA revealed a main effect for Group (Wilk's Λ =.64, *F*(3, 25) = 4.64, *p* = .010). The univariate follow-up analyses indicated that participants assigned more negative descriptions to the minority (*F*(1, 27) = 14.93, *p* = .003, η_p^2 = .36; see Figure 3.3 Top), estimated the frequency of negative traits to be higher in the minority than in the majority (*F*(1, 27) = 4.59, *p* = .041, η_p^2 = .15; see Figure 3.3 Middle), and evaluated the minority less favorable than the majority (*F*(1, 27) = 10.30, *p* = .045, η_p^2 = .28; see Figure 3.3 Bottom), even though the ratio of positive and negative descriptions at encoding were equal in both groups. Thus, an IC was present in all three measures. The main effect for Time point was not significant (Wilk's Λ = .74, *F*(3, 25) = 2.94, *p* = .052). Critically, the interaction between Group and Time point was not significant (Wilk's Λ = .98, *F*(3, 25) = 0.15, *p* = .930), indicating that the IC was stable over time.

3.3.1.2.2 Memory

The results for the unbiased hit rates can be seen in Figure 3.4. There was a significant main effect for Group (*F*(1, 27) = 147.45, *p* <.001 , η_p^2 = .85). Source memory was better for descriptions of the majority than for descriptions of the minority. This effect was modulated by Valence (*F*(1, 27) = 7.25, *p* = .012, η_p^2 = .22). No other effects were observed (all *F*s < 1.84, *p* > .187). The Group x Valence interaction was followed-up by t-tests for descriptions of each group: positive descriptions of the majority were better recalled than negative descriptions of the majority (*t*(27) = 3.28, *p* = .006, two-tailed, Cohen's *d* = 0.64). Contrary to the predictions of the SDA, no difference was observed between the recall of positive and negative descriptions of the minority (*t*(27) = -0.67, *p* = .506, two-tailed, Cohen's *d* = -0.13).

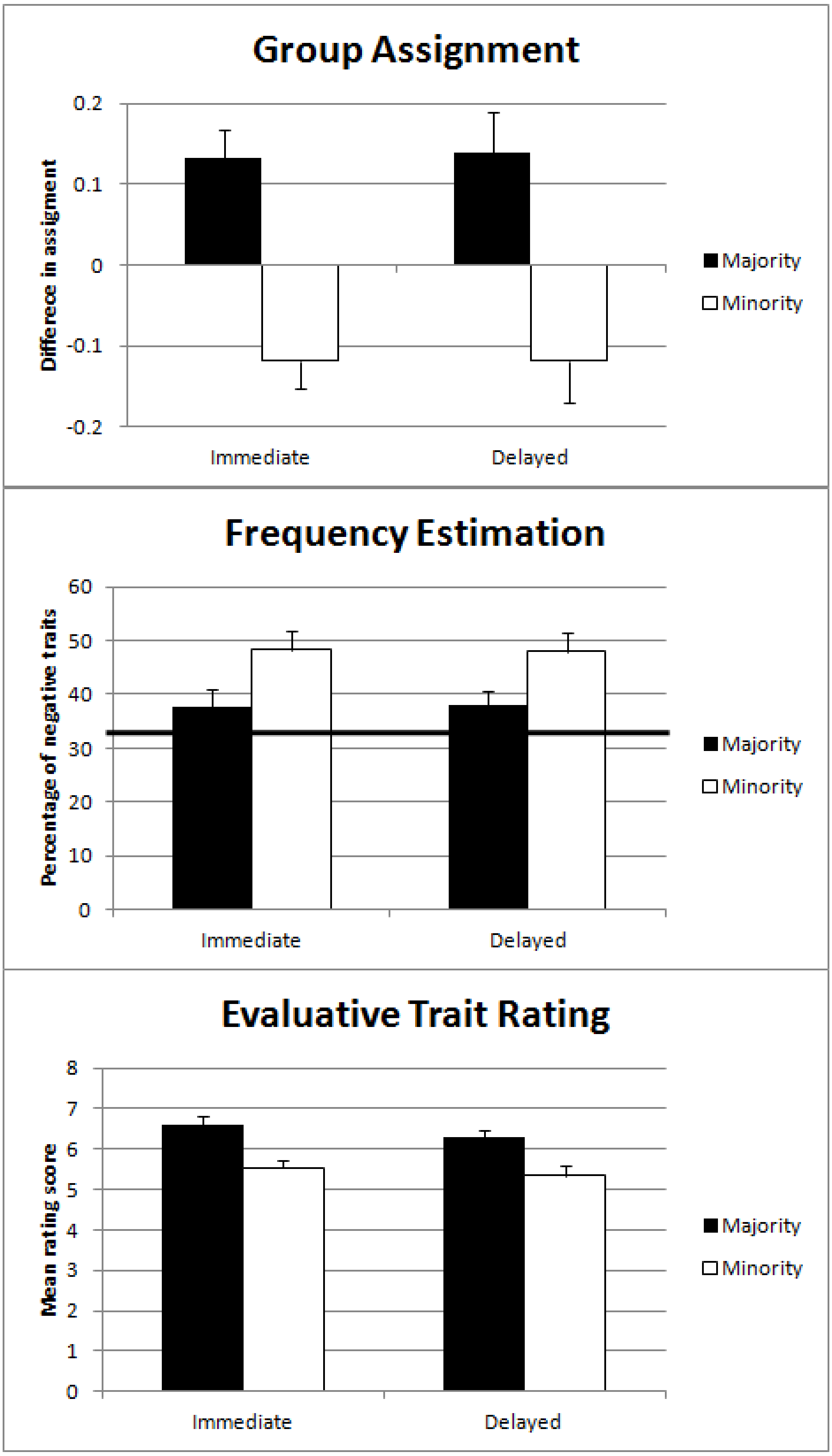

Figure 3.3 Overview over the results of Experiment 1. Top: Difference between positive and negative traits assigned to each group. Middle: Estimated ratio of negative trait in percent. The bold line represents the true ratio. Bottom: Mean evaluative trait ratings. Error bars represent between-subject standard errors.

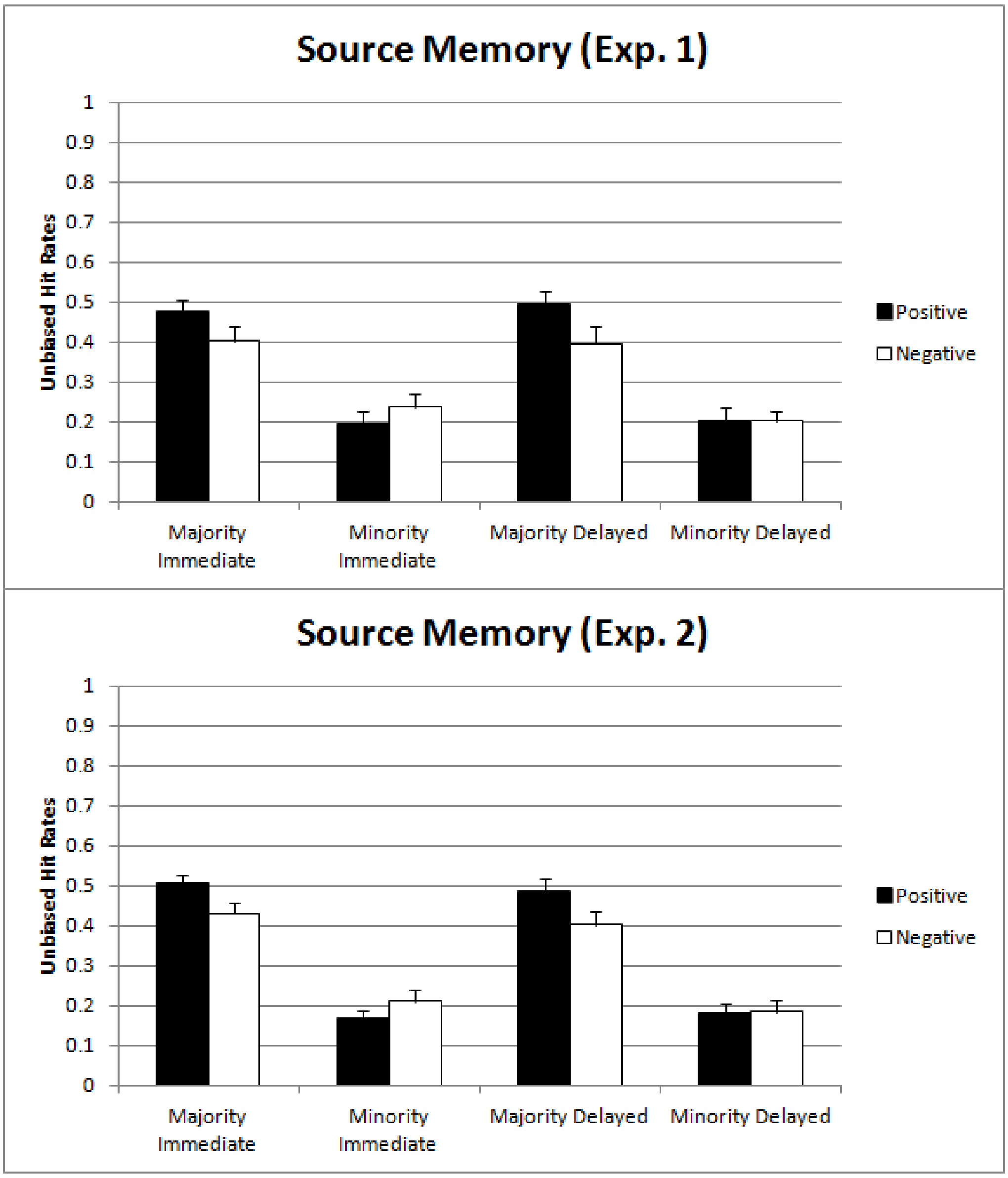

Figure 3.4 Unbiased hit rates for source memory in Experiment 1 (Top) and Experiment 2 (Bottom). Error bars represent between-subject standard errors.

The results for Pr and Br can be seen in Figure 3.5. The analysis of item discrimination Pr did not reveal any significant effects (all *F*s < 1, *p* > .332). In contrast, the analysis of the response bias Br revealed a significant effect for Valence ($F(1, 27) = 8.57$, $p = .007$, $\eta_p^2 = .24$) indicating that participants were more likely to assign positive descriptions to the majority than negative descriptions. No other effects were significant (all *F*s < 1, *p* > .579).

3.3.1.2.3 Measures of Illusory Correlation

The overall MANOVA was significant (Wilk's Λ = .49, *F*(7, 21) = 3.09, *p* = .021). As can be seen in Table 3.3, an IC was observed for all correlation coefficients except the coefficient for the frequency estimation task at immediate testing (*p* = .053) and the encoded correlation (*p* = .215). The latter finding indicates that the IC was formed after encoding. Furthermore, the encoded correlation did neither correlate with any single IC coefficient (all p-values >.282) nor with the average of all single IC coefficients (*r* = -.07, *p* = .708).

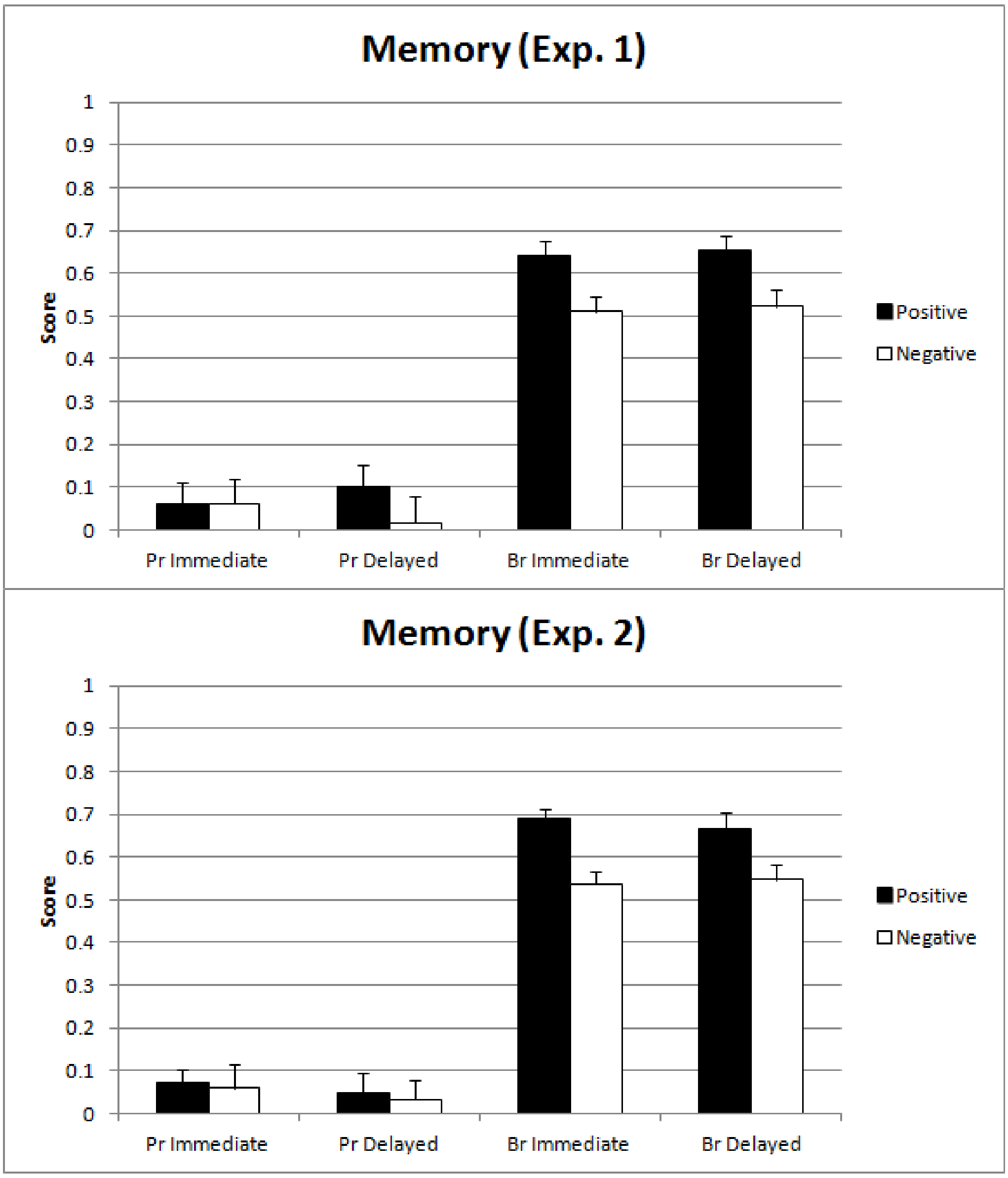

Figure 3.5 Pr and Br scores for memory in Experiment 1 (Top) and Experiment 2 (Bottom). Error bars represent between-subject standard errors.

Table 3.3 Size of the illusory correlation in Experiment 1 and 2 across measures and time points. Please note that the presented measures are untransformed correlation coefficients. Statistical tests, however, were conducted with the Fisher z-transformed coefficients. The Benjamini-Hochberg procedure was used to adjust the p-values.

		Group assignment		Frequency estimation		Evaluative trait rating		Encoded correlation	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Exp.1	Immediate	.12**	.19	.11†	.32	.26**	.41	-	-
	Delayed	.13*	.27	.10*	.27	.23**	.35	.02	.11
Exp.2	Immediate	.13**	.17	.10**	.19	.25**	.39	-	-
	Delayed	.10*	.30	.08*	.22	.22*	.43	.01	.08

Significantly different from zero: † *p* < .10 one-tailed, * *p* <.05 one-tailed, ** *p* <.01 one-tailed

3.3.1.3 Discussion

The aim of Experiment 1 was to reproduce the effect of an illusory correlation for the standard skewed frequency distributions (Hamilton and Gifford, 1976) and to validate our experimental design and materials. Our experiment indeed successfully evoked ICs, which were observed consistently across all task for assessing ICs (group assignment, frequency estimation, and evaluative trait rating). By obtaining these assessments in Experiment 1, we established a baseline for Experiment 2, in which we sought to test our prediction that, contrary to the predictions of the ILA, equated frequencies for positive and negative traits still result in an IC.

Other noteworthy findings of Experiment 1 were that ICs did not change over time and were not attributable to erroneous encoding. Moreover, the analysis of source memory accuracy revealed that descriptions of the majority were better remembered than descriptions of the minority, and positive descriptions of the majority were better remembered than its negative descriptions. Contrary to the SDA, negative descriptions of group B members were not better remembered than positive descriptions. In addition to this, response bias, but not item discrimination, differed between positive and negative descriptions. We review these findings in detail in the general discussion.

3.3.2 Experiment 2: Equated Frequency Condition

3.3.2.1 Method

3.3.2.1.1 Participants

41 students (35 female) of the Saarland University participated in Experiment 2. Participants received partial course credit or comparable compensations for their participation. Those who did not finish the experiment properly (*n* = 2 did not use the correct response keys, *n* = 1 could not finish task due to a technical error), who reported severe difficulties to focus on the task (*n* = 3), or who reported quite specific hypotheses about the purpose of the study as reported in the post-experimental questionnaire (*n* = 5) were excluded from further analysis. The inclusion of the participants who inferred the hypotheses of the study would have rendered some effects marginally significant or non-significant. However, similar as for Experiment 1, we opted to exclude the latter participants because otherwise the validity of the conclusions drawn from the data could be considered as compromised (see section 3.3.1.1.1 for a discussion of the relevant research). The final sample comprised 30 participants (26 female, median age 20 years, range 18-30 years).

3.3.2.1.2 Materials

For Experiment 2, 18 positive and 18 negative traits were drawn from the pool of 48 adjectives and again matched for arousal, imageability, and word length (see Appendix A.). For each participant, the 36 first names and trait adjectives were randomly combined to a description and assigned either to group A (majority) or to group B (minority). Twelve positive and twelve negative person descriptions were assigned to group A, six positive and six negative descriptions were assigned to group B (see Table 3.1).

3.3.2.1.3 Procedure and Data Analysis

For Experiment 2 the same experimental and analytical procedures were used as for Experiment 1.

3.3.2.2 Results

3.3.2.2.1 Group Assignment, Frequency Estimation, and Evaluative Trait Rating

The overall MANOVA revealed a main effect for Group (Wilk's Λ = .73, *F*(3, 27) = 3.37, *p* = .033). Similar to Experiment 1, the univariate follow-up analyses indicated that participants assigned more negative descriptions to the minority (*F*(1, 29) = 9.53, *p* = .012, η_p^2 = .25; see Figure 3.6 Top), estimated the frequency of negative traits to be higher in the minority than in the majority (*F*(1, 29) = 6.26, *p* = .018, η_p^2 = .18; see Figure 3.6 Middle), and evaluated the minority less favorable than the majority (*F*(1, 29) = 6.71, *p* = .018, η_p^2 = .19; see Figure 3.6 Bottom). Thus, an IC was present in all three measures. No main effect for Time point was observed (Wilk's Λ = .92, *F*(3, 27) = 0.82, *p* = .495). As in Experiment 1, the interaction between Group and Time point was again not significant (Wilk's Λ = .97, *F*(3, 27) = 0.28, *p* = .840), indicating that the IC was stable over time.

3.3.2.2.2 Memory

As in Experiment 1, arcsine-transformed unbiased hit rates were entered in a repeated measure ANOVA with the factors Group (majority vs. minority), Valence (positive vs. negative), and Time point (immediate vs. delayed). There was a significant main effect for Group (*F*(1, 29) = 332.46, *p* <.001 , η_p^2 = .92). Source memory performance was better for the majority than for the minority (Figure 3.4). This effect was modulated by Valence (*F*(1, 29) = 8.74, *p* = .007 , η_p^2 = .23). No other effects were observed (all *F*s < 1, *p* > .413). The Group x Valence interaction was followed up by t-tests. The follow-up t-test for the majority revealed a significant difference (*t*(29) = 2.40, *p* = .046, two-tailed, Cohen's *d* = -0.44) between positive and negative descriptions, with positive descriptions being better recalled than negative descriptions. As in Experiment 1, no such difference was observed for the minority (*t*(29) = -1.11, *p* = .275, two-tailed, Cohen's *d* = -0.20).

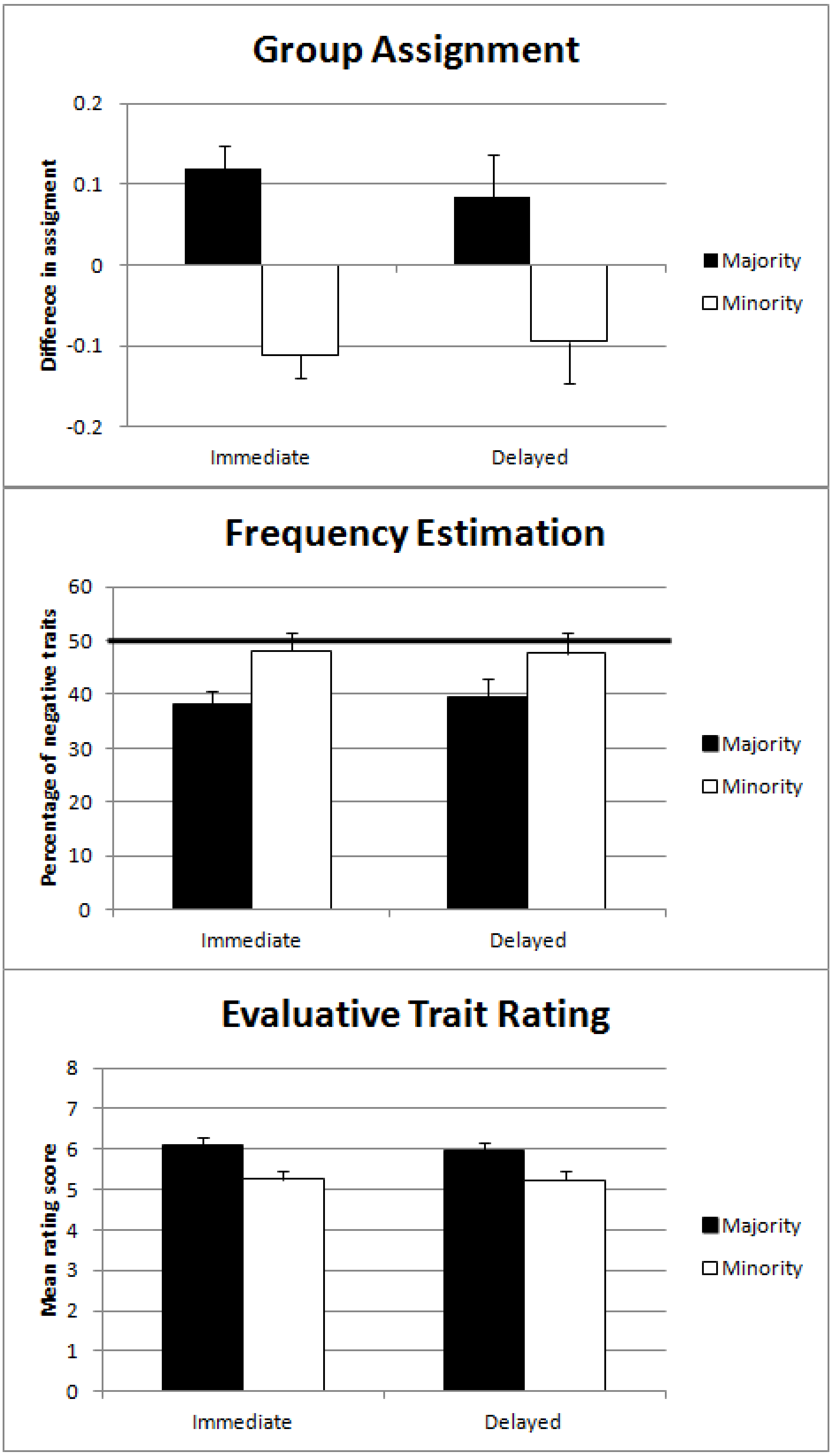

Figure 3.6 Overview over the results of Experiment 2. Top: Difference between positive and negative descriptions assigned to each group. Please note that the difference score os smaller than in Experiment 1 due to the equated frequencies of positive and negative descriptions. Middle: Estimated ratio of

negative traits in percent. The bold line represents the true ratio. Bottom: Mean evaluative trait ratings. Error bars represent between-subject standard errors (continued from p. 53).

The results for Pr and Br can be seen in Figure 3.5. As in Experiment 1, no effects were observed for Pr (all *F*s < 1, *p* > .649). In contrast, the analysis of Br revealed a significant effect for Valence ($F(1, 27) = 12.65$, $p = .001$, $\eta_p^2 = .30$), indicating that participants were more likely to assign positive descriptions to the majority than negative descriptions. No other effects were significant (all *F*s < 1, *p* > .622).

3.3.2.2.3 Measures of Illusory Correlation

Using the same procedure as in Experiment 1, correlation coefficients were calculated for each dependent variable and each time point separately. The overall MANOVA was significant (Wilk's $\Lambda = .57$, $F(7,23) = 2.47$, $p = .048$). As in Experiment 1, an IC was observed for all tasks but the sentence valence rating (encoded correlation: $p = .348$; Table 3.3). Furthermore, the encoded correlation did neither correlate with any single IC coefficient (all *p*-values > .174) nor with the average of all single IC coefficients ($r = -.14$, $p = .447$).

3.3.2.2.4 Comparison between Experiment 1 and 2

In order to put the results into perspective, we also compared the results from Experiment 1 and 2 statistically. We found no differences between Experiment 1 and Experiment 2 in the bias against the minority (all *F*s < 1.23, *p* > .310), memory task (all *F*s < 2.18, *p* >.145), and IC coefficients (all *F*s < 1, *p* > .635). Next, we collapsed the samples of both experiments in order to calculate two linear regressions that assess the relative contribution of source memory performance on the ICs, separately for the frequency estimation task and the evaluative trait rating. We used the arcsine-transformed unbiased hit rates as predictors and the IC as criterion. All measures were collapsed across time points. The memory performance for each of the four category combinations significantly predicted the IC in both tasks (see Table 3.4). As indicated by the sign of the regression coefficients, memory for positive descriptions of the majority and negative descriptions of the minority increased the IC, whereas memory for negative descriptions of the majority and positive descriptions of the minority decreased it.

Furthermore, the model was not improved when experiment was included as categorical predictor (frequency estimation: $\Delta R^2 < .01$, $F(1, 52) = 0.06$, $p = .809$; evaluative trait rating: $\Delta R^2 < .01$, $F(1, 52) = 0.04$, $p = .845$) or as moderator (frequency estimation: $\Delta R^2 = .07$, $F(4, 48) = 1.55$, $p = .203$; evaluative trait rating: $\Delta R^2 = .02$, $F(4, 48) = 0.37$, $p = .828$).

However, as indicated by the bivariate correlations (Table 3.5) between the variables, memory performance for positive descriptions of the majority and for negative descriptions of the minority act as suppressor variables, i.e. they do not predict the extent of illusory correlation themselves, but contribute to the prediction by removing criterion-irrelevant variance from the other predictors (Pandey & Elliott, 2010). This might be the case due to partial structural dependence between the variables. In other words, the variance that is not shared between the variables predicts IC. Nevertheless, the results indicate that memory for both, positive and negative descriptions, plays a key role in IC.

Table 3.4 Standardized regression coefficients(β) of the arcsine-transformed unbiased hit rates in the prediction of the illusory correlation in the frequency estimation task and the evaluative trait rating.

	Majority		Minority	
	Positive	Negative	Positive	Negative
Frequency Estimation	.36**	-.56***	-.39**	.28*
Evaluative Trait Rating	.53***	-.53***	-.34**	.37**

Frequency estimation: $R^2 = .43$, evaluative trait rating: $R^2 = .41$
Significantly different from zero: * $p < .05$, ** $p <.01$, *** $p <.001$

Table 3.5 Correlations between unbiased hit rates and IC. Data were collapsed across the experiments.

	Frequency Estimation	Evaluative Trait Rating	Majority Positive	Majority Negative	Minority Positive
Majority Positive	-.00	.21	-		
Majority Negative	-.47***	-.32*	.35**	-	
Minority Positive	-.43***	-.27*	.48***	.44***	-
Minority Negative	-.01	.11	.08	.48***	.13

Significantly different from zero: * *p* < .05, ** *p* <.01, *** *p* <.001

3.3.2.3 Discussion

Experiment 2 tested whether an IC could be observed even in the equated frequency condition. In essence, all the results from Experiment 1 could be replicated in Experiment 2.

3.3.3 Discussion of Experiment 1 and 2

Experiment 1 and 2 revealed an IC irrespective of the experimental condition (skewed vs. equal distribution). Thus, participants associated the majority with positive traits and the minority with negative traits. The IC was not affected by the delay. The results from Experiment 1 and 2 provided evidence which were more in line with the SDA than the ILA. However, the results of the group assignment task remained inconclusive about whether source memory was actually heightened for the least frequent category combination.

3.3.3.1 Implications of Experiment 1 and 2 for the Information Loss Account

The computer simulations showed that the ILA would predict the presence of an IC in the skewed frequency condition, but not in the equal frequency condition. The main result of Experiment 1 and 2 is that the same extent of ICs can be observed irrespective of the frequency ratio in all three common measures of ICs (frequency estimation, group assignment, and evaluative trait rating). The latter results are clearly incompatible with the notion of the ILA that ICs can be explained by the nature of skewed frequency distributions and pure information loss. The ILA can only account for the current findings when additional assumptions are made, such as that the processing varies depending on the content or context. Fiedler (2000) offered a computational implementation of selective forgetting and weighting of information in the BIAS model. However, it would necessitate the introduction of psychological constructs like distinctiveness into the ILA, which the ILA originally sought to replace. The data from the group assignment task provide some support for the ILA, insofar as source memory is better for the majority than for the minority and best for positive descriptions of the majority. Due to the preponderance of the majority and of positive descriptions, the formation of inter-item associations, which facilitate memory retrieval, are more likely for the majority than the minority (Fiedler et al., 1993). Nevertheless, the general pattern of results does not provide support for the ILA.

3.3.3.2 Temporal Characteristics of the Illusory Correlation

No decline of ICs across time was found in Experiment 1 and 2. In previous studies (Chapman, 1967; Lilli & Rehm, 1983) ICs wore off when participants were tested on another sequence of stimuli immediately after the IC was assessed (e.g. learning 48 stimuli twice), most likely as a result of increased transparency of the judgment situation (Lilli & Rehm, 1983, 1984). Consistent with this idea, the IC was not reduced when participants were engaged in extended learning (e.g. learning 96 stimuli in a row; Lilli & Rehm, 1983) and no IC was found when participants were given only a summary table about the groups (Hamilton et al., 1985). In our case, participants did not run through the same experimental procedure again and could, therefore, not profit from their knowledge of the task, but had to rely on the initially encoded information.

The absence of a decline in the extent of the IC might imply that the IC helps forming an expectation, which is later used to guide behavior and attention giving rise to the expectancy-based IC (Hamilton, 1981; see also Garcia-Marques & Hamilton, 1996). Even though one might have expected to see more decay in the memory judgments and an increase in the IC itself on the basis of the ILA, all measures proved to be quite stable over the 40 min delay. This is more consistent with the SDA. A study by Hunt (2009) indicated that the effects of distinctiveness on memory persist even after a retention interval as long as 48 h.

However, there are several limitations regarding the interpretation of the delay manipulation that should be discussed. First of all, our delay was relative short. Even though some memory decay might be expected after a short time period (e.g. Rubin & Wenzel, 1996), a longer retention interval might allow a more conclusive assessment of the stability of ICs. Second, the sample size was rather low. Thus, the absence of an effect might simply imply that the statistical power was insufficient. Third, and more critically, we measured the impact of a delay in a within subject design, i.e. subjects might have been influenced by their initial choices and ratings (e.g. via remembering or response priming). Therefore, the second assessment might not be independent from the first assessment. Furthermore, the first test offered an opportunity to further consolidate memory. Indeed, studies on the testing effect imply that repeated tests on the same material might even lead to an improved performance (e.g. Karpicke & Roediger, 2008; Rosburg, Johansson, Weigl, & Mecklinger, 2015). Future studies should also investigate the impact of delays in a between-subject design and use longer retention intervals.

3.3.3.3 Implications of Experiment 1 and 2 for the Shared Distinctiveness Account

The results of the frequency estimates, group assignments, and evaluative trait rating are consistent with the SDA, but in the group assignment task no heightened retrieval accuracy for negative descriptions of the minority was found in Experiment 1 and 2. Based on the SDA, we expected memory to be better for the minority than the majority and to be best for negative descriptions of the minority. Instead we found the reverse pattern. Memory was better for the majority than the minority and positive descriptions of the majority were remembered best. Our findings do, therefore, not support the SDA.

The latter finding is partially consistent with Fiedler et al. (1993), who reported that memory is better for positive items compared to negative items. However, the results from Experiment 1 and 2 are inconsistent with the findings from memory studies on ICs that used multinomial processing tree models, which report that item memory is better for negative than for positive items, whereas source memory is equal for all sources (Bulli & Primi, 2006; Klauer & Meiser, 2000; Meiser, 2003).

Most IC studies only control for the desirability of the behavior descriptions (e.g. Van Rooy et al., 2013). We used trait descriptions instead of behavior descriptions. This allowed us to exert more control for factors known to affect memorability like arousal, imageability, or word length. However, the word frequency was higher for positive traits than for negative traits in our study, reflecting an actual linguistic difference between positive and negative words (e.g. Zajonc, 1968). Nevertheless, the difference in word frequency between positive and negative trait adjectives cannot account for the present results, because across both experiments source memory was better for the majority than the minority for both, positive (t(57) = 17.04, p < .001, d = 2.24) and negative items (t(57) = 10.81, p < .001, d = 1.42). And last, but not least, negative items tended to be remembered better than positive items for the minority at the immediate test (t(57) = -1.70, p = .095, d = -0.22; see also Figure 5).

The latter finding seems to provide some tentative support for the SDA hypothesis. However, given the small effect size, a sample of roughly 130 participants would be necessary to test this effect with sufficient power. Furthermore, as indicated by the regression analysis, only memory performance for minority members with positive descriptions and majority members with negative descriptions is a reliable predictor of IC. Positive descriptions of the majority or negative descriptions of the minority contribute to the prediction of IC only by removing criterion-irrelevant variance. Thus, even though a sufficiently powered design might provide evidence for heightened memory for shared distinctive items, this memory advantage alone might not be a strong determinant of the IC. Thus, this additional analysis also did not provide support for the SDA.

We also analyzed our data by calculating the traditional measures of items discrimination Pr and response bias Br (Snodgrass & Corwin, 1988). As major finding, the analysis indicated that positive items were more likely attributed to the majority than to the minority. This effect cannot be accounted for by differences in frequency between positive and negative items, because the response bias effect was similar in the skewed and equated frequency condition. However, this finding is in line with a study by Alves et al. (2015) on the effect of valence on recognition memory. Similar to our study, Alves and colleagues reported that there were differences in response bias, but not in item discrimination for positive and negative items. They attributed this effect to the fact that positive items are more similar to each other than negative items (Alves et al., 2017, 2015). Thus, future studies that test the IC with trait descriptions should also control for the influence of similarity.

At first sight, the results from the traditional memory measures Pr and Br seem to be at odds with the results from the unbiased hit rates. The traditional measures revealed differences only in response bias, whereas the unbiased hit rates reveal superior memory for the majority as compared to the minority. However, these two types of measures quantify different aspects of memory. Although Pr (or d’) scores are corrected for response bias by subtracting false alarm rate (Snodgrass & Corwin, 1988), these measures assume that detection is equal for both the majority and the minority. The majority is treated as target and the minority as distractor. Genuine memory differences (i.e. differences in discrimination ability) between majority and minority would be ascribed to response bias. In the current study, we used the unbiased hit rates, because we were interested in mnemonic differences between the majority and the minority. Such differences would be masked by the traditional approaches. Unbiased hit rates allow the assessment of such memory differences and have been successfully used in previous source memory studies (Bell et al., 2012; Suzuki & Suga, 2010).

If the observed memory advantage for the majority in the unbiased hit rates was solely due to a response bias in favor for the majority, then all Br scores should be significantly larger than .50. Br > .50 indicates a response bias that favors the majority over the minority. A one-sample t-test of Br revealed that across both experiments only the Br scores for positive items were significantly different from .50 (immediate: $t(57) = 8.08$, $p < .001$, Cohen's $d = 1.06$; delayed: $t(57) = 6.23$, $p < .001$, Cohen's $d = 0.82$). The Br scores for negative items did not differ from .50 (immediate: $t(57) = 1.03$, $p = .306$, Cohen's $d = 0.14$; delayed: $t(57) = 1.30$, $p = .198$, Cohen's $d = 0.17$). These results indicate that the differences in the unbiased hit rates cannot be attributed solely to response bias in favor for the majority, but might instead primarily reflect genuine memory differences.

To sum up, the experimental design used in Experiment 1 and 2 was not optimal for assessing the validity of the SDA. The inclusion of foil items and testing the same number of items from the majority, minority and foils should make the IC paradigm less susceptible to guessing strategies and offers more possibilities for assessing memory performance. Therefore, we decided to conduct a third behavioral experiment with these methodological refinements in order to clarify the role of memory in ICs.

3.4 Experiment 3: Source Memory in an Optimized Illusory Correlation Paradigm

In Experiment 3, we used the online questionnaire method, which has been shown to induce reliable ICs (Weigl, Mecklinger, & Rosburg, 2015), in order to test whether memory for distinctive category combinations is enhanced as predicted by the SDA. For this purpose, we introduced several improvements in the methodology of Experiment 1 and 2 and previous studies (e.g. Fiedler et al., 1993; Hamilton et al., 1985; Klauer & Meiser, 2000). More precisely, we introduced new distracter items in the test phase, took into account primacy and recency effects by removing items at the beginning and the end of the study phase from the test phase and equated the frequency of majority, minority and distracter items in the source monitoring task in order to reduce the impact of response bias. Furthermore, we improved on the frequency estimation task by allowing participants to independently estimate the frequency of positive and negative behavior for both groups. This allows us to assess whether participants are actually more accurate in estimating frequencies for the majority as predicted by the ILA.

3.4.1 Methods

3.4.1.1 Participants

Master students from a statistics course received a link to the online questionnaire and were encouraged to fill out the questionnaire at home. 50 out 89 students who started the online questionnaire completed it. Additionally, one participant, who indicated to have given untrue answers (as checked via a control question at the end of the questionnaire) was excluded, two participants were excluded because of unreasonable frequency estimates (i.e. their estimates differed more than three standard deviations from the group mean) and five participants were excluded because they took more than one hour to complete the questionnaire. Thus the final sample comprises 42 participants (34 female, age: mean: 24.18, range: 21-30). The students received sweets for their participation. Furthermore, parts of the data set were used in the statistics course to demonstrate the application of multivariate statistics.

3.4.1.2 Materials

A list of 36 behavior descriptions was selected from the norms provided by Ehrenberg, Cataldegirmen, and Klauer (2001). The majority in this experiment was composed of employees from a large company and the minority was composed of employees from a small company. Sixteen desirable and eight undesirable sentences described employees of the larger company and eight desirable and four undesirable sentences described employees of the smaller company (see Table 3.6). For one half of the participants, the larger company was labeled Company Alpha and the smaller Company Delta. For the other half the labeling of the companies was reversed. Eight desirable and four undesirable sentences from the larger company and all sentences from the smaller company served as critical items in a source monitoring task. An additional set of eight desirable and four undesirable behavior descriptions served as distracters in the source monitoring task.

Table 3.6 Distribution of positive and negative words in the study phase and the test phase.

	Study phase		Test phase	
	Positive	Negative	Positive	Negative
Majority	16	8	8	4
Minority	8	4	8	4
New	-	-	8	4

The items were matched for valence, gender typicality, number of words per sentence, and number of characters per sentence (all *F*s < 1.30, *p* > .287; see Table 3.7). Following Hamilton and Gifford (1976), each stimulus item consisted of a German male first name, a company designation (Alpha or Delta) and a behavior description (e.g. "Jan from Company Delta is very patient with elderly people.").

Table 3.7 Mean (± SD) for the positive and negative words.

	Majority (N = 12)	Minority (N = 12)	New (N = 12)
Valence	4.63 (2.08)	4.64 (2.07)	4.65 (2.05)
Gender Typicality	4.12 (0.66)	4.00 (0.63)	4.04 (0.55)
Number of Words	11.50 (2.24)	12.08 (1.83)	11.17 (2.79)
Number of Characters	70.75 (13.90)	80.00 (10.65)	72.83 (18.64)

3.4.1.3 Procedure

The online questionnaire was implemented using SoSci Survey (https://www.soscisurvey.de). The instructions followed those of the Hamilton and Gifford (1976) experiment. The experiment was described to be concerned with how people process and retain information about others and that participants would read behavior descriptions of employees working in one of two companies, a large company and a small company.[2] Next, participants were instructed to go through the list of 36 behavior descriptions and rate on a 5-point scale whether the description induces a positive or a negative impression of the described person. After completing the rating task four dependent measures were obtained: 1) evaluative trait rating, 2) frequency estimates, and 3) source monitoring task. In the evaluative trait rating participants rated each group separately with respect to ten traits on a 10-point scale. Traits were taken from Fiedler et al. (1993; see also Experiment 1 and 2). Afterwards, they had to decide for each group separately whom they would rather choose as friend or business partner: a member from company Alpha/Delta or a stranger. Then, participants had to estimate the absolute frequency of desirable and undesirable behavior in each group. Finally participants were given a list containing 24 descriptions of the encoding task with first names and group designation removed together with the 12 distracter items (e.g. “P always pays back borrowed money as soon as possible.”). Only items from the middle of the encoding task were chosen to avoid primacy and recency effects. For each description participants were instructed to decide whether the description belonged to a member of company Alpha or company Delta or whether it was a new description. Demographic variables, personality measures, and manipulation check questions were assessed at the end of the questionnaire.

[2] Additionally, half of the participants were told that the employees work in their company only for monetary reasons (low entitativity condition). The other half was told that the employees work in their company, because they strongly identify themselves with their company (high entitativity condition). The effect of entitativity was assessed with an additional social decision task. Since entitativity did not influence any dependent variables in our preliminary analyses (all *p*s > .250), we collapsed the data across the low and high entitativity groups. The entitativity manipulation will not be discussed any further.

3.4.1.4 Data Analysis

The analysis of the data was similar to our previous studies (Experiment 1 and 2; Weigl et al., 2015). For the encoding rating and the evaluative trait rating, we used one-sided one-sample t-tests in order to compare the ratings for the majority with the ratings for the minority. For the frequency estimation task and the group assignment task, we used one-sampled t-tests in order to assess whether the frequency estimates or the assignment frequency differed from the actual frequencies and 2 x 2 repeated-measure analyses of variance (ANOVAs) with the factors Group (majority vs. minority) and Valence (positive vs. negative) on the relative deviations from the true frequencies in order to assess differences between the conditions in the degree of over- or underestimation relative to the true frequencies.

Source memory performance was measured by calculating the unbiased hit rates (Wagner, 1993) for each type of item (Table 3.6 right-hand side). Unbiased hit rates are calculated by multiplying the conditional probability of correctly classifying an object given that it is present with the conditional probability of correctly applying a response category given that it is applied (see Appendix A. for more details). The resulting values were arcsine transformed for statistical analysis. We conducted a 3 x 2 repeated measure ANOVA with the factors Type (majority vs. minority vs. new) and Valence (positive vs. negative) on the arcsine-transformed unbiased hit rates.

In order to quantitatively assess the extent of IC, we converted the encoding rating, the evaluative trait rating, the frequency estimates and the group assignments into measures of IC. The data from the encoding rating and the evaluative trait ratings were converted into correlation coefficients by calculating the point-biserial correlation. Phi coefficients were calculated from the data of the frequency estimation task and the group assignment task. The resulting correlations were transformed by using the Fisher's Z transformation and subjected to a multivariate analysis of variance (MANOVA) to test for difference from a null vector. If a significant effect was observed, follow-up one-sided t-tests were conducted.

All data were analyzed using IBM SPSS version 22.0 (IBM Corp., Armonk, NY, USA). In repeated-measure ANOVAs, Greenhouse-Geisser corrected p-values are reported, if the assumption of sphericity was violated. Effect sizes for t-tests were calculated with G*Power 3 (Faul, Erdfelder, Lang & Buchner, 2007). Significance level was set to *p* = .050 for all analysis. Two-sided tests were used, unless stated otherwise.

3.4.2 Results

3.4.2.1 Encoding Rating, Evaluative Trait Rating, Frequency Estimation, and Group Assignment

Encoding rating. In order to assess whether participants already started associating the majority with positive behavior and the minority with negative behavior at encoding, the valence rating of the behaviors were averaged for each company separately. A paired sample t-test revealed that the larger was not rated more positively than the smaller company (*t*(41) = -3.73, *p* = .999, one-sided, Cohen's *d* = -0.58). On the contrary, participants rated the majority (*M* = 3.40, *SD* = .16) less favorable than the minority (*M* = 3.49, *SD* = .20).

Evaluative trait rating. As predicted, participants rated the majority (*M* = 6.50, *SD* = 1.04) more positive than the minority (*M* = 5.54, *SD* = 1.23) in the evaluative trait ratings (*t*(41) = 3.34, *p* = .001, one-sided, Cohen's *d* = 0.51).

Frequency estimates. The absolute estimated frequencies and the true frequencies are shown in Table 3.8. As predicted, participants overestimated the frequency of negative behavior in the minority, but were quite accurate in the frequency estimates for positive behavior of minority members as assessed with one-sample t-tests (see Table 3.8). Of note, they underestimated the frequency of both positive and negative behavior in the majority.

The Group x Valence repeated-measure ANOVA on relative deviations from the true frequency revealed a main effect for Group (*F*(1, 41) = 30.33, *p* < .001, η_p^2 =.43), a main effect for Valence (*F*(1, 41) = 18.66, *p* < .001, η_p^2 =.21), and an interaction between Group and Valence (*F*(1, 41) = 16.24, *p* < .001, η_p^2 =.28). Consistent with the impression from the one-sample t-tests, follow-up paired t-tests revealed a significant difference in the relative deviation from the true frequency between positive and negative behaviors in the minority (*t*(41) = -4.56, *p* < .001, Cohen's *d* = -0.70), but not in the majority (*t*(41) = -0.42, *p* = .680, Cohen's *d* = -0.06).

Group assignment. The number of behaviors attributed either to group Alpha or Delta and the true frequencies in the group assignment task are shown in Table 3.8. The group assignment task revealed that participants assigned significantly more positive behaviors to the majority and negative behaviors to the minority than would have been expected based on the true frequencies. Furthermore, participants assigned less positive behaviors to the minority as compared to the true frequency.

The Group x Valence repeated-measure ANOVA on the relative deviations from the true frequency revealed an interaction between Group and Valence ($F(1, 41) = 11.93$, $p = .001$, $\eta_p^2 = .23$). The main effects for Group and Valence were not significant ($F(1, 41) = 0.64$, $p = .430$, $\eta_p^2 = .02$ and $F(1, 41) = 0.01$, $p = .924$, $\eta_p^2 = .00$, respectively). Follow-up paired-sample t-tests revealed that more positive behaviors were assigned to the majority than to the minority ($t(41) = 4.28$, $p < .001$, Cohen's $d = 0.66$) and more negative behaviors were assigned to the minority than to the majority ($t(41) = -2.29$, $p = .028$, Cohen's $d = -0.35$).

Table 3.8 True and estimated frequencies in the frequency estimation task and in the group assignment task. The values in brackets represent standard deviations.

		Frequency Estimation		Group Assignment	
		True	Estimated	True	Assigned
Majority	Positive	16	12.31 (5.66)**	8	10.33 (3.19)**
	Negative	8	6.36 (2.90)**	4	3.28 (2.39)†
Minority	Positive	8	7.21 (3.79)	8	6.11 (3.32)**
	Negative	4	6.14 (3.53)**	4	4.92 (2.33)*

Significant deviation from true value: † $p < .10$, * $p < .05$, ** $p < .01$

3.4.2.2 Source Memory Performance

The descriptive statistics for the unbiased hit rates are shown in Table 3.9. The Type x Valence repeated measure ANOVA revealed a main effect for Type

(F(1.31, 53.65) = 220.66, p < .001, η_p^2 = .84), but not for Valence (F(1, 41) = 0.20, p = .654, η_p^2 = .01). Orthogonal contrasts revealed that there was no difference between the majority and minority (F(1, 41) = 0.48, p = .494, η_p^2 = .01), indicating that source memory performance was comparable for both groups, but that there was a difference between the new distractors and both groups combined, indicating that item memory was better than source memory (F(1, 41) = 255.58, p < .001, η_p^2 = .86). Moreover, there was an interaction between Type and Valence (F(2, 82) = 10.66, p < .001, η_p^2 =.21). Orthogonal contrasts revealed this interaction was reliable when the majority was compared with the minority (F(1, 41) = 14.11, p = .001, η_p^2 = .26), as well as when the new distractors were compared with both groups combined (F(1, 41) = 4.42, p = .042, η_p^2 = .10). Follow-up t-tests revealed that for the majority, positive behaviors were better remembered than negative behaviors (t(41) = 3.12, p = .003, Cohen's d = 0.48), whereas, consistent with our expectation, negative behaviors were better remembered than positive behaviors for the minority (t(41) = -2.56, p = .014, Cohen's d = -0.39). For new distracter items, performance was better for negative items than for positive items (t(41) = -2.32, p = .026, Cohen's d = -0.36).

Table 3.9 Mean (± SD) for the unbiased hit rates

	Majority	Minority	New
Positive	.28 (.15)	.17 (.14)	.84 (.22)
Negative	.17 (.20)	.29 (.21)	.88 (.22)

3.4.2.3 Measures of Illusory Correlation

The overall MANOVA was significant (Wilk's Λ = .55, F(4, 38) = 7.91, p < .001). As can be seen in Table 3.10, an IC was observed for all correlation coefficients except the encoded correlation. The latter finding indicates that the IC was formed after encoding. Furthermore, the encoded correlation did neither correlate with any single IC coefficient nor with the average of all single IC coefficients (r = .10, p = .52).

As in Experiment 1 and 2, we calculated two linear regressions in order to assess the relative contribution of source memory performance on the ICs, separately for the frequency estimation task and the evaluative trait rating. We used the arcsine-transformed unbiased hit rates as predictors and the IC as criterion. The memory performance for each of the four category combinations significantly predicted the IC in both tasks (see Table 3.11). As indicated by the sign of the regression coefficients, memory for positive descriptions of the majority and negative descriptions of the minority increased the IC, whereas memory for negative descriptions of the majority and positive descriptions of the minority decreased it. Furthermore, the model was not improved when the unbiased hit rates for positive and negative new items were included (frequency estimation: ΔR^2 = .03, F(2, 35) = 1.11, p = .342; evaluative trait rating: ΔR^2 < .01, F(2, 35) = 0.26, p = .771).

Table 3.10 Illusory correlation coefficients for each task.

			Deviance from zero correlation		Intercorrelations		
	Mean	SD	t(41)	p(one-sided)	r_EVA	r_EST	r_REC
r_ENC	-.03	.05	-3.63	.999	.07	.10	-.03
r_EVA	.23	.37	3.97	.000		.74*	.76*
r_EST	.11	.20	3.72	.001			.71*
r_REC	.23	.42	3.14	.002			

r_ENC = encoded correlation, r_EVA = evaluative correlation, r_EST = estimated correlation, r_REC = reconstructed correlation, * p < .001

In addition, as indicated by the bivariate correlations between the variables (Table 3.12), the unbiased hit rates for old items were all significantly correlated with the illusory correlation coefficients. Thus, the inclusion of foil items successfully removed the partial structural dependence in the variables which were observed in the regression analysis of Experiment 1 and 2.

Table 3.11 Standarized regression coefficients (β) of the arcsine-transformed unbiased hit rates in the prediction of the illusory correlation in the frequency estimation task and the evaluative trait rating.

	Majority		Minority	
	Positive	Negative	Positive	Negative
Frequency estimation	.32*	-.20	-.36*	.32*
Evaluative trait rating	.31*	-.34**	-.48**	.23*

Frequency estimation: $R^2 = .49$; evaluative trait rating: $R^2 = .64$
 * p < .05, ** p < .01

Table 3.12 Correlations between unbiased hit rates and illusory correlation coefficients.

	Frequency Estimation	Evaluative Trait Rating	Majority Positive	Majority Negative	Minority Positive
Majority Positive	.42**	.41**	-		
Majority Negative	-.43*	-.60***	-.39**	-	
Minority Positive	-.38**	-.54***	.22	.33*	-
Minority Negative	.42**	.33*	.31*	.07	-.06

Significantly different from zero: * *p* < .05, ** *p* <.01, *** *p* <.001

3.4.3 Discussion of Experiment 3

In this study, we introduced several methodological refinements in order to test whether memory for the most distinctive category combination is enhanced as predicted based on the SDA. Evaluative trait ratings, frequency estimates, and group assignments all indicate the successful induction of an IC. Overall source memory accuracy was equal for the majority and the minority. However, memory for negative behavior of the minority and for positive behavior of the majority was elevated even after controlling for response bias. Furthermore, memory predicted the extent of IC in the other measures. In addition, the frequency judgments indicate that – in contrast to predictions based on the ILA – people not only misrepresent frequency information about the minority, but also frequency information about the majority. One could argue that the underestimation of the frequencies concerns the absolute frequencies, whereas the ILA is primarily concerned with the frequency ratios (Fiedler, 2000). In contrast to most studies on the IC (Haslam & McGarty, 1994), participants in Experiment 3 were allowed to provide independent estimates of the frequencies for all four possible category combinations. Since the frequency ratio was the same for negative items (eight negative majority items and four negative minority items) and for the minority items (eight positive and four negative minority items; see also Table 3.6), the ILA would predict equal estimates for both frequency ratios. However, it is important to note that the frequency was underestimated for negative majority items, but not for positive minority items (see also Table 3.8). Furthermore, the estimated ratios were different on a trend level ($t(40) = 1.72$, $p = .092$). Thus, the overall pattern of results is consistent with the prediction from the SDA, but hard to accommodate with the ILA. This study, thus, implies that it is heightened availability of distinctive group-behavior combinations rather than random information loss that generates ICs.

3.5 General Discussion: Information Loss versus Shared Distinctiveness

The main goal of the computer simulation and the three behavioral experiments was to compare the relative merits of the two accounts for ICs, the Shared Distinctiveness Account (SDA) and Information Loss Account (ILA). For this purpose, we: 1) used computer simulations based on the BIAS model to derive exact predictions whether an IC would be present in an equal frequency condition under the assumption of pure information loss, 2) tested whether the predictions derived from the simulations correspond to the behavioral data, 3) explored whether observed ICs were stable over time shortly after initial encoding, and 4) assessed whether shared distinctiveness actually enhanced memory. The main result of Experiment 1 and 2 was that, in contrast to predictions from simulations with the BIAS model, skewed and the equated frequency distributions led to an IC. The main result of Experiment 3 was that shared distinctiveness indeed enhanced memory as predicted by the SDA.

The presence of an IC despite equated frequencies in Experiment 2, as well as the misrepresentation of the frequencies for the majority in Experiment 3, contradict the ILA. The ILA can only account for the results of Experiment 2 and 3 by introducing additional parameters, which allow the processing to vary as a function of the context. For example, assuming that negative items from the majority are especially susceptible to noise could account for the results in the frequency estimation task of Experiment 2 and 3. However, a compelling theoretical reason is needed to explain why only negative majority items are affected by noise, but not negative minority items. One construct which could account for such differential processing would be distinctiveness – the concept which the ILA originally sought to replace.

Whereas the group assignment task of Experiment 1 and 2 did not reveal any heightened retrieval accuracy for negative descriptions of the minority, Experiment 3 provided evidence for superior memory for negative minority items. Since the experimental procedure of Experiment 3 was specifically designed to assess source memory performance, the results of Experiment 3 provide strong support for the SDA. In the literature, there is some debate about whether shared distinctiveness actually enhances memory. In general, studies using cued recall (see Mullen & Johnson, 1990, for a review), free recall (e.g. Hamilton et al., 1985), or one-shot illusory correlations (Risen et al., 2007) support the heightened memory hypothesis of the SDA, whereas studies relying on signal detection theory (Fiedler et al., 1993) or a multinomial processing tree approach (e.g. Bulli & Primi, 2006; Klauer & Meiser, 2000) failed to accrue evidence for the SDA. The results from our behavioral experiments indicated that evidence for superior memory for negative items of the minority is most likely to be found, when the experimental design is optimized for the unambiguous assessment of source memory performance.

A limitation of all three behavioral experiments is that participants were informed beforehand that there will be a majority group and a minority group. Indeed, ICs based on expectancies or self-relevance have been reported even when frequencies for one dimension have been equated (Spears, Eiser, & van der Pligt, 1987; Spears, van der Pligt, & Eiser, 1986). The instructions were designed to closely follow those of Hamilton and Gifford (1976) which explicitly state the presence of a majority and a minority. This instruction might already have activated a pre-experimental association between belonging to a majority and positive traits and belonging to a minority and negative traits (McGarty et al., 1993). However, the very same instructions lead to a reversal of ICs, when negative behavior is more prevalent indicating that participants respond according to the displayed information (Hamilton & Gifford, 1976, Exp. 2). Therefore, it seems unlikely that pre-existing associations alone are responsible for the observed effect. Nevertheless, future studies should not inform the participants about the presence of a majority in order to check whether such information affects frequency judgments.

A more general limitation concerns the nature of our data. Even though the data obtained in the behavioral experiments are largely in line with the SDA, the data do not allow unambiguous conclusions about the underlying processes. According to the SDA, shared distinctiveness leads to better encoding and higher availability of the least frequent category combination. The higher availability of instances from the least frequent category combinations in turn affects the frequency estimation and the evaluative trait ratings. However, the current behavioral data do not allow us to decide between the causal chain proposed by the SDA and an alternative scenario: shared distinctiveness simultaneously affects all three dependent measures at the same time. One potential solution to this problem would be the use of time-sensitive methods like event-related potentials which are able to shed more light on the cognitive processes at encoding that contribute to the development of ICs. Of course, the present results also have implications for other accounts of the IC. However, these will be discussed in Chapter 6.

3.6 Conclusion from the Behavioral Investigations

The results from the three behavioral experiments indicate that the mere frequency ratio of positive to negative information is not the only psychologically active mechanism in the distinctiveness-based IC. Rather, the asymmetry in the processing of positive and negative information (Alves et al., 2017; Baumeister et al., 2001) might play an important role in the formation of ICs and stereotypes. However, a pure distinctiveness approach was supported only when optimal design choices were made as in Experiment 3. In order to further elucidate the role of memory and learning in the distinctiveness-based IC, we decided to investigate the cognitive processes at encoding that contribute to the development of ICs with ERPs. This approach will be described in more detail in the subsequent chapters.

4 Electrophysiological Studies on Distinctiveness

The purpose of the two experiments reported in chapter 4 was to test the basic assumptions underlying the main ERP experiment on the IC, Experiment 6, which will be reported in the next chapter. Experiment 4 was designed to test the assumption that the P300 component can be used as a neural marker for shared distinctiveness. Furthermore, Experiment 4 contrasted the shared distinctiveness assumption with the accentuation hypothesis. Experiment 5 tested whether the P300 component can predict subsequent recollection-based memory and, in addition, explored whether the N400 at encoding is related to subsequent familiarity-based recognition.

4.1 Experiment 4: An ERP Comparison between Distinctiveness and Accentuation in the Illusory Correlation Paradigm

4.1.1 Introduction

Despite decades of research on the IC (see Mullen & Johnson, 1990; Stroessner & Plaks, 2001, for reviews), there is still debate over the exact mechanisms leading to ICs. In this research, we aimed at evaluating the relative merit of two mechanisms, namely distinctiveness (Hamilton & Gifford, 1976) and accentuation (McGarty et al., 1993; Sherman et al., 2009). The SDA (Hamilton, Dugan, & Trolier, 1985; Hamilton & Gifford, 1976) proposes that some group-behavior combinations are more distinctive than others. These distinctive combinations receive more attention and are better encoded than less distinctive ones. However, the accentuation approach offers a different explanation for the IC (McGarty et al., 1993). In the IC paradigm, a contrast between the positive behavior of majority members and the negative behavior of minority members results in maximal differentiation and, as a by-product, in ICs.

The research literature is ambiguous whether shared distinctiveness (Hamiton & Gifford, 1976) or accentuation (McGarty et al., 1993) is the most plausible mechanism for explaining ICs. Attention Theory (AT; Kruschke, 2003; Sherman et al., 2009) is a newer account for the distinctiveness-based IC which combines the mechanisms of distinctiveness and accentuation in a single theoretical framework. In contrast to the accentuation account by McGarty et al. (1993), accentuation can arise even in the absence of real category differences in AT. The crucial component in AT is the differential speed of category acquisition. Participants in an IC experiment first learn to associate the positive behavior with the majority, because they see more majority items and positive items than minority or negative items. In order to differentiate the minority from the majority, attention shifts to distinctive attributes, in this case to the negative behavior for the minority. Thus, AT proposes that relative, rather than absolute, distinctiveness contributes to the IC effect.

In a series of experiments, Sherman et al. (2009) provided evidence for the AT, but Experiment 5 provides the strongest evidence for the attention shift mechanism. After the participants acquired an IC in an impression formation task, they performed an X probe task. A pair of behavior descriptions (i.e. a common and a rare trait behavior) was simultaneously presented for either a majority or a minority member. Then, an X would appear either on the side of the common trait or on the side of the rare trait. Consistent with the predictions of AT, participants reacted faster to probes at the position of the common trait behavior of the majority or at the rare trait behavior of the minority than to probes at the position of the rare trait behavior of the majority or the common trait behavior of the minority. Attention shifts, thus, facilitated differentiation between the categories.

However, these experiments do not allow to determine whether accentuation already happens during encoding of the stimuli. It is possible that only distinctiveness plays a role at encoding and accentuation takes place after encoding or at retrieval. Event-related potentials (ERPs) provide high resolution temporal information and are therefore suitable to determine the cognitive processes involved in the processing of the stimuli immediately at encoding. In this ERP study, we implemented the IC paradigm in an active oddball task. This task elicits a P300, which indexes attention allocation and the processing of distinctiveness (Polich, 2007). Participants attended a regular series of stimuli (e.g. sequences of the letter O) in which a rare stimulus – the target stimulus – appears (e.g. the letter X). Furthermore, there was a frequent color (e.g. purple) and a rare color (e.g. orange).

We hypothesized that the P300 for the most frequent and the most infrequent category combination would be larger than the P300 for moderately frequent category combinations, if the accentuation hypothesis was true. At first sight, a high P300 for the most frequent category combination seems implausible from the perspective of the research literature on the P300 (see Polich, 2007, for a review), but there is evidence that category members which heighten between-category differences and reduce within-category variance receive more attention at encoding and greater weight at judgment (Krueger, 1991; Krueger & Rothbart, 1990). In contrast, we expected that the P300 would linearly increase as a function of rareness of the category combination, if only distinctiveness plays a role at encoding. Furthermore, we explored whether distinctiveness and accentuation manifest themselves in the frontal slow-wave, a component involved in working memory (García-Larrea & Cézanne-Bert, 1998; Monfort & Pouthas, 2003; Schubotz, 1999). Since the frontal slow-wave is also often observed at the encoding of inter-item associations (Kamp et al., 2017), it might be the case that accentuation is primarily manifested in the frontal slow-wave.

4.1.2 Method

4.1.2.1 Participants

24 healthy, right-handed students (18 female; age: median = 23 yrs, range: 19-29 yrs) of the Saarland University were recruited. Four additional participants had to be excluded due to excessive artifacts. Participants received partial course credit or comparable monetary compensations for their participation. All participants gave written informed consent.

4.1.2.2 Procedure

The behavioral experiment was programmed using PsychPy 2 Version 1.84.1 (Peirce, 2007, 2009). Participants were tested individually and sat in front of a 17 inch monitor. All displays were centered and had grey background. All words, sentences, and symbols (except the targets) were presented in white

Arial font. The participants were instructed to attend a sequence of visual stimuli that contained the letters “X” and “O” presented either in purple or orange. The participants’ task was to categorize each stimulus depending on the letter and its color.

In total, there were 360 trials. The four possible letter-color combinations were presented with the following frequencies (see also Table 4.1): The frequent letter was presented in frequent color in 150 trials (Nontarget-Standard) and in the infrequent color in 90 trials (Nontarget-Deviant). The infrequent letter was presented in the frequent color in 90 trials (Target-Standard) and in the infrequent color in 30 trials (Target-Deviant). Thus, there was a negative correlation (φ = -.125) between the letters and the colors, i.e. the infrequent letter was less likely to co-occur with the infrequent color than with the frequent color. This particular frequency distribution was chosen, because it would equate the number of trials which should be inside and outside the focus of attention according to AT. Furthermore, the present ratios have been shown to elicit an illusory correlation (Weigl et al., 2015). The trials were presented in randomized order with the constraint that no feature was repeated more than three times.

Each trial started with the presentation of one of the two letters for 200 ms followed by a fixation star presented for 2300 ms. They had to classify each colored letter by pressing the keys D, F, J, or K as fast and accurate as possible. The response window was 2500 ms (corresponding to the length of a trial). The frequency of the letters and colors and the response keys were counterbalanced across participants.

After the participants completed the experimental task, they judged the relative frequency of the letters, the colors, and each color-letter combination on a questionnaire.

Table 4.1 Frequencies of the letter-color combinations in the experiment. The frequencies as they would be necessary for a zero-correlation are given in brackets.

		Color		
		Frequent color (Standard)	Infrequent color (Deviant)	Σ
Letter	Frequent letter (Nontarget)	150 (160)	90 (80)	240
	Infrequent letter (Target)	90 (80)	30 (40)	120
	Σ	240	120	360 (φ = -.125)

4.1.2.3 EEG Recordings and ERP Processing

An elastic cap (Easycap, Herrsching, Germany) with 28 embedded silver/silverchloride EEG electrodes was attached to the participant’s head (recording sites: Fp1, Fp2, F7, F3, Fz, F4, F8, FC5, FC3, FCz, FC4, FC6, T7, C3, Cz, C4, T8, CP3, CPz, CP4, P7, P3, Pz, P4, P8, O1, O2, as well as the right mastoid (M2)). EEG was continuously recorded with reference to the left mastoid (M1). The ground was placed in AFz. EOG activity was recorded with two electrodes placed on the outer canthi and by a pair of electrodes placed above and below the right eye. Electrode impedances were kept below 5 kΩ. Data were sampled at 500 Hz and filtered online from 0.016 Hz to 250 Hz.

Offline, EEG data were processed with the Brain Vision Analyzer 2.04 (Brain Products, Gilching, Germany). Data were downsampled to 200 Hz and then high-pass filtered at 0.1 Hz (48dB/oct). Ocular, ECG and muscle artifacts were removed using an independent component analysis (ICA) based algorithm implemented in the software. After the ICA correction, data were re-referenced to linked mastoids. For the ERP analysis, a 30 Hz low-pass filter was applied (48 dB/oct). Data were then segmented into epochs of 2200 ms (including 200 ms pre-stimulus baseline). Baseline-corrected data were screened for remaining artifacts and all segments that contained amplitudes outside the range of -70 to 70 µV or voltage steps exceeding 50 µV/ms were removed.

The mean amplitude of the P300 was calculated for the time window from 300 to 400 ms as the peak of the P300 was at around 360 ms in the grand average across all four conditions at Pz. The frontal slow wave was extracted as mean amplitude for the time window from 1200 to 2000 ms.

4.1.2.4 Data Analysis

The accuracies and reaction times from the oddball task were analyzed using Friedman's ANOVA due to violations of the assumption of normality. Since we were especially interested in the comparison between the most frequent and the least frequent category combination, we conducted three follow-up Wilcoxon tests. In the first test, we compared the least frequent category combinations with the mean of the remaining three combinations. In the second test, we compared the most frequent category combination with the mean of the two moderate frequent combinations. In the last test, we compared the two moderate combinations with each other. One participant had to be removed from the analyses due to missing values.

In order to assess the accuracy of the frequency estimates, several one-sample t-tests were calculated which compared the estimated relative frequency with the true relative frequency. The frequency estimates were also used to calculate a phi coefficient. Three participants failed to report correct conditional relative frequencies (i.e. the conditional frequencies did not add up to 1.0) and were therefore excluded from the analysis of the conditional probabilities.

In order to test for general differences in the ERPs, we calculated a one-way analysis of variance (ANOVA) with repeated measures for the P300 at Pz and the frontal slow wave at Fz separately. The one-way ANOVA included the four factor levels Nontarget-Standard, Nontarget-Deviant, Target-Standard, and Target-Deviant. We tested our specific hypotheses regarding the distinctiveness account and the accentuation account by using a priori polynomial contrasts. If the distinctiveness account were true, then the linear contrasts should become significant. In contrast, the quadratic contrasts should become significant, if the accentuation account were true. We refrained from reporting the cubic contrast, because we did not have any specific hypotheses regarding this contrast and cubic contrasts are often of little theoretical relevance (Field, Miles, & Field, 2012).

All statistical analyses were conducted using IBM SPSS 24 (IBM Corp., Armonk, NY, USA). For all repeated-measure ANOVAs, sphericity was checked with Mauchly's test and a Greenhouse-Geisser correction was applied when necessary. Effect sizes for t-tests were calculated with G*Power 3.1 (Faul et al., 2009). The alpha criterion was set to p = .050 for all analyses.

4.1.3 Results

4.1.3.1 Behavioral Results

The descriptive statistics for the accuracy and the reaction times in the oddball task can be found in Table 4.2. Overall, the accuracy of the participants was very high. The Friedman's ANOVA was significant ($\chi^2(3)$ = 25.13, p < .001) indicating differences between the conditions. The follow-up Wilcoxon tests revealed that participants were less accurate in the least frequent condition (i.e. Target-Deviant) as compared to the three other conditions (z = -3.65, p < .001). No differences in accuracy emerged between the other category combinations (z = -1.33, p = .183 and z = -0.08, p = .935). The Friedman's ANOVA for the reaction times was also significant ($\chi^2(3)$ = 28.46, p < .001) indicating differences between the conditions. The follow-up Wilcoxon tests revealed that the participants were significantly slower in the least frequent condition (i.e. Target-Deviant) as compared to the other three conditions (z = -3.53, p < .001). Furthermore, the participants were significantly faster in the most frequent condition (i.e. Nontarget-Standard) as compared to the two moderate frequent conditions (z = -2.25, p = .024). The two moderate frequent conditions did not differ from each other (z = -0.61, p = .543). In short, the behavioral results indicate that the participants were less accurate and reacted slower in the least frequent condition as compared to the other conditions.

The relative frequency of the frequent colors and frequent letters was .67. The subjective frequency estimates for the colors and the letters were highly accurate (colors: M = .67, SD = .08; letters: M = .69, SD = .10) and did not statistically differ from the true relative frequency (colors: *t*(20) = 0.10, *p* = .924, Cohen's *d* = 0.02; letters: *t*(20) = 0.76, *p* = .456, Cohen's *d* = 0.17). In the estimation of the frequent color given the frequent letter (the most frequent category combination, Nontarget-Standard), the true relative frequency was .63. However, the participants significantly overestimated this frequency (*M* = .70, *SD* = .11; *t*(20) = 2.85, *p* = .010, Cohen's *d* = 0.62). In the estimation of the infrequent color given the infrequent letter (the least frequent category combination, Target-Deviant), the true relative frequency was .25. In this case, the participants frequency estimation did not significantly differ from the true value (*M* = .29, *SD* = .15; *t*(20) = 1.15, *p* = .263, Cohen's *d* = 0.25). The analysis of the phi coefficient indicated that the subjective correlation significantly differed from the actual, negative correlation (*M* = -.02, *SD* = .12; *t*(20) = 4.20, *p* < .001, Cohen's *d* = 0.92), but not from a zero-correlation (*t*(20) = -0.71, *p* = .48. Cohen's *d* = 0.15).

Table 4.2 Mean, standard deviation (in brackets), median, and interquartile range (in brackets) for the accuracy and the reaction time in ms.

	Accuracy		Reaction Time	
	Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)
Nontarget Standard	.94 (.09)	.97 (.06)	607 (84)	589 (81)
Nontarget Deviant	.89 (.16)	.94 (.09)	625 (110)	600 (112)
Target Standard	.92 (.10)	.96 (.07)	627 (95)	609 (96)
Target Deviant	.82 (.15)	.87 (.17)	688 (141)	625 (230)

4.1.3.2 ERP Results

Figure 4.1 shows the grand averages for the P300 and the frontal slow wave at the electrodes Fz and Pz. As expected, the P300 is manifested at Pz and the frontal slow wave is manifested at Fz. The ERP results of the experiment indicate that the P300 is smallest for the most frequent category combination and largest for the least frequent category combination. The P300 for category combinations of medium frequency were of intermediate strength. The frontal slow-wave, in contrast, was similar for all category combinations except the least frequent one.

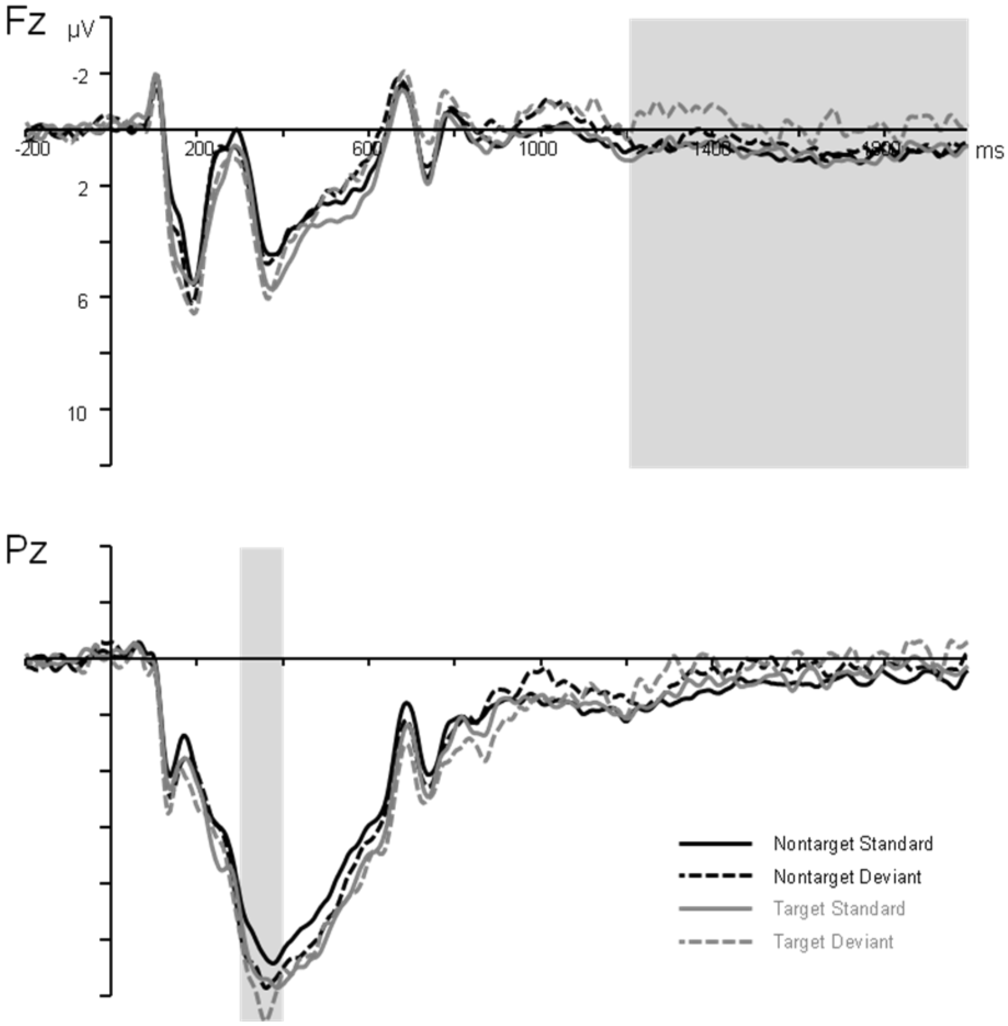

Figure 4.1 Grand average of the P300 and the frontal slow wave at the electrodes Fz and Pz. The grey bars denote the time window used for the statistical analysis.

This visual impression was corroborated by the ANOVAs. The overall ANOVA for the P300 revealed a significant effect (*F*(2.34, 53.71) = 3.31, *p* = .037, η_p^2 = .13). In a next step, we examined the a priori contrasts[3]. The linear contrast was significant (*F*(1, 23) = 6.35, *p* = .019, η_p^2 = .22), whereas the quadratic contrast was not significant (*F*(1, 23) = 0.55, *p* = .466, η_p^2 = .02). As depicted in Figure 4.1 and Figure 4.2, there is a linear increase of the P300 as a function of rarity. Thus, the results for the P300 support the distinctiveness account rather than the accentuation account.

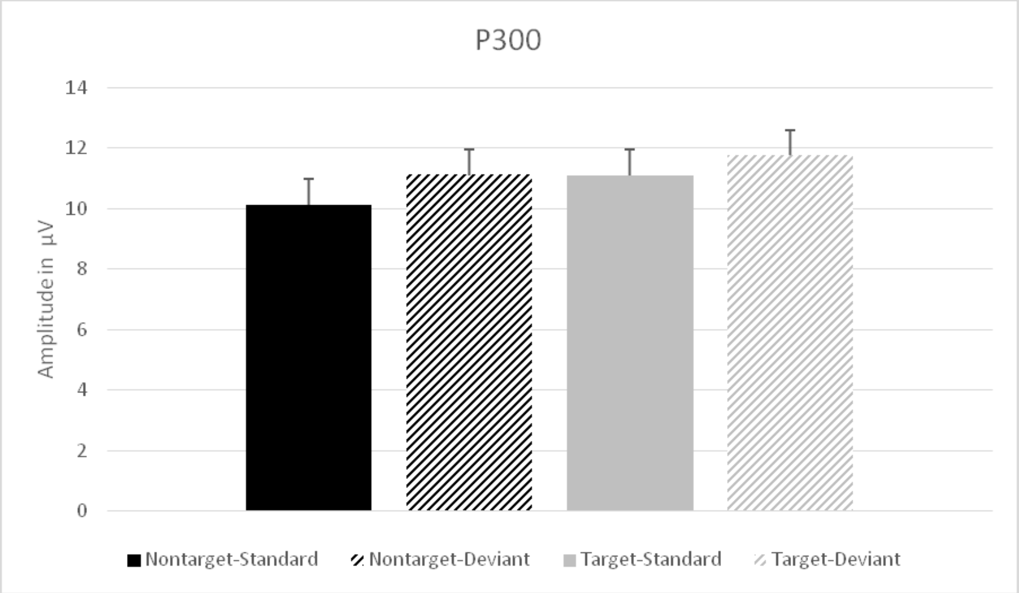

Figure 4.2 The mean amplitude for the P300 at the electrode Pz. The error bars indicate within-subject confidence intervals.

[3] Some might argue that a more specific coding scheme would provide more adequate tests for our hypotheses. Namely, a contrast comparing the least frequent category combination with the other three combinations would provide a more adequate test for distinctiveness and a comparison comparing the mean of the most and the least frequent combination with the mean of the two moderate combinations would provide a more adequate test for accentuation. For the P300, the contrast testing distinctiveness was significant (*F*(1, 23) = 4.69, *p* = .041, η_p^2 = .17), whereas the contrast testing accentuation was not (*F*(1, 23) = 0.55, *p* = .466, η_p^2 = .02). For the frontal slow-wave, the contrast testing distinctiveness was significant (*F*(1, 23) = 4.55, *p* = .044, η_p^2 = .17), whereas the contrast testing accentuation was not (*F*(1, 23) = 2.15, *p* = .157, η_p^2 = .09). Thus, the more specific contrast scheme leads to the same conclusion as the polynomial contrasts.

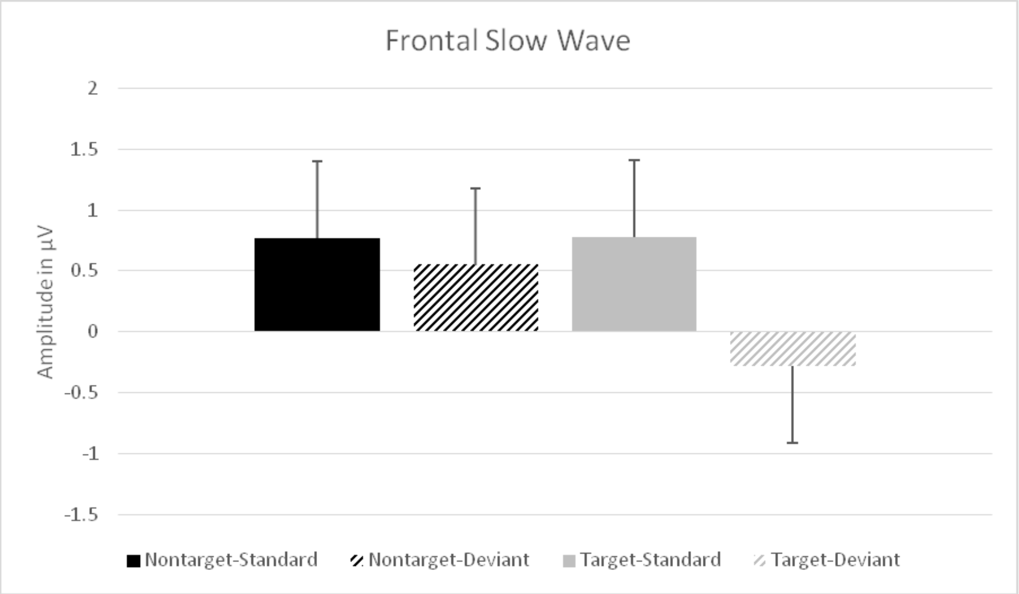

Figure 4.3 The mean amplitude of the frontal slow wave at the electrode Fz. and the frontal slow wave. The error bars indicate within-subject confidence intervals.

The overall ANOVA for the frontal slow wave revealed a significant effect ($F(3, 69) = 2.52$, $p = .065$, $\eta_p^2 = .10$). The linear contrast was marginally significant ($F(1, 23) = 4.10$, $p = .055$, $\eta_p^2 = .15$), whereas the quadratic contrast was not significant ($F(1, 23) = 2.15$, $p = .157$, $\eta_p^2 = .09$). As can be seen on Figure 4.1 and Figure 4.3, only the Target-Deviant differed from the other category combinations. The analysis, therefore, does not offer direct support for either account.

4.1.4 Discussion

The current study investigated whether distinctiveness and accentuation, two mechanisms proposed to underlie the distinctiveness-based IC (Sherman et al., 2009), can already be observed at encoding. We used behavioral and ERP measures to answer this research question.

The most important result of the current ERP study is that the P300 amplitude increased as a function of the rarity of the category combinations. This finding is not only in line with the literature on the P300 as an index of subjective probability (Duncan-Johnson & Donchin, 1977; Polich, 2007), but also with the distinctiveness approach of the IC (Hamilton et al., 1985; Hamilton & Gifford, 1976). At the same time, this result also contradicts the accentuation hypothesis (Sherman et al., 2009).

The behavioral data also indicate differential processing of the least frequent category combination as compared to the other combinations. Participants were both, less accurate and slower in classifying the least frequent category combination relative to the other combinations. This result is in line with a study by Stroessner et al. (1992), who also reported that the encoding latencies were longest for the least frequent category combination. Thus, the behavioral results also support the distinctiveness account. At first sight, these results seem hard to align with the findings by Sherman et al. (2009, Experiment 5). They reported a facilitation effect in the reaction times to the most and the least frequent category combination and interpreted this result as evidence for attention shifts that lead to accentuation. However their reaction time task followed an initial category acquisition task, whereas we assessed the reaction times directly during category acquisition. Thus, accentuation might still take place, but at later processing stages after category acquisition.

Therefore, it seems to be the case that the perception of distinctive items leads to context updating. Furthermore, the reaction time results indicate prolonged processing of the distinctive stimulus. One caveat of this study is that subjects had to respond to each stimulus category. Therefore, it might be possible that the observed P300 pattern was only elicited, because all stimuli were task-relevant to the task. The results for the frontal slow-wave seem consistent with the shared distinctiveness view, because the least frequent category combination is processed differently. However, this result might be also accommodated with the accentuation effect, because the least frequent category combination receives special processing in the formation of inter-item associations as indexed by the frontal slow-wave.

Future studies that investigate the electrophysiological basis of the accentuation effect should follow the original paradigm by Tajfel and Wilkes (1963) more closely. Using six different stimuli (e.g. six tones ascending in frequency) could be presented either as separate stimuli or as part of two categories. Finding ERP differences between the two conditions which are identical on the physical level, but differ on the psychological level, would provide strong evidence for accentuation (see Tajfel & Wilkes, 1963).

Consistent with the literature on frequency estimation (e.g. Fiedler, Kutzner, & Vogel, 2013), the participants were highly accurate in estimating the marginal frequencies in the frequency estimation task. In contrast to what would be expected on the basis of distinctiveness, the most frequent, but not the least frequent category combination was significantly overestimated. The extended learning situation might have contributed to the current result. So far, only few studies have investigated ICs using more than 48 trials and the results are equivocal. Some studies indicate that the IC disappears once learning reaches an asymptote (Murphy et al., 2011; Spiers et al., 2016), whereas other studies reported even after extended learning (Kutzner et al., 2011; Lilli & Rehm, 1983). The current study contributes to this literature. Since the same ratios with a lower absolute number of trials (i.e. 36 instead of 360) elicited a significant IC in a prior study (Weigl et al., 2015), the current results might reflect a pre-asymptotic phase in which the subjective covariation estimation declined from positive to zero. Future studies should extend on the current study by using even longer stimulus sequences. Such studies could determine whether the subjective covariation estimation converges to the true negative correlation as would be expected based on the asymptotic learning curve argument (Murphy et al., 2011).

Alternatively, if such studies find that participants still estimate the correlation to be zero, the current findings could be interpreted as a type of IC, because the subjective covariation deviates from the objectively presented covariation pattern (e.g. Chapman, 1967). Since negative contingencies as in our experiment are typically the hardest to detect (e.g. Nisbett & Ross, 1980), the participants might rather expect a zero-correlation than a negative correlation or rely on the marginal probabilities as a guide for the estimation of the conditional probabilities. In fact, their estimates are close to the true value of the marginal frequencies (*p*s > .220).

Even though the Experiments 3 and 4 provided some support for the SDA, it is so far not clear whether the enhanced memory for shared distinctive items is based on recollection or familiarity. In order to derive detailed predictions for an ERP experiment that tests the effect of shared distinctiveness on memory, we first conducted another ERP study to see how distinctiveness is reflected in the ERP components N400 and P300.

4.2 Experiment 5: Can ERPs at Encoding Predict Subsequent Familiarity-Based and Recollection-Based Recognition

4.2.1 Introduction

Recognition memory is supported by two independent, but not mutually exclusive processes: automatic, fast, and context-free recognition process called familiarity and a slower, more deliberate all-or-none recognition process called recollection (Mandler, 1980; Yonelinas, 2002; Yonelinas et al., 2010). Recollection and familiarity can be measured in several different ways (see Yonelinas, 2002, for an overview). Reaction time methods exploit that recollection-based recognition is – in general – slower than familiarity-based recognition. Three other widely used methods – the process-dissociation procedure (PDP; Jacoby, 1991), receiver-operator characteristics (ROCs; Yonelinas, 1994), and the remember/know procedure (R/K procedure; Tulving, 1985; Yonelinas & Jacoby, 1995) – try to computationally estimate the extent of recollect and familiarity on the basis of the amount of recognized material.

In addition, recollection and familiarity can be dissociated with ERPs. Familiarity is associated with a mid-frontal component, which roughly peaks around 400 ms (the early mid-frontal old/new effect or FN400), whereas recollection is associated with a late left parietal component in the time window between 400 and 700 ms (the late left-parietal old/new effect or LPC; Rugg & Curran, 2007). Several fMRI and patient studies additionally revealed that recollection and familiarity are processed in different brain regions (Yonelinas et al., 2010). Recollections seems to crucially depend on the hippocampus, whereas familiarity is associated with the surrounding medial temporal lobe structures (e.g. parahippocampal gyrus and the ento- and perirhinal cortex; Aggleton & Brown, 1999; Yonelinas, 2002; Yonelinas et al., 2010).

Experiment 5 is concerned with the question whether familiarity and recollection can already be dissociated in the ERPs at encoding. In other words, we were interested whether ERPs at encoding can already predict if an item will be later recognized on the basis of familiarity and recollection. The encoding processes that might contribute to familiarity-based or recollection-based recognition can be investigated with the subsequent memory paradigm (see section 2.3.2 and 2.3.3).

There is ample evidence that the frontal slow wave SME contributes to the formation of inter-item associations, whereas the P300 SME is related to item-specific processing at encoding (Fabiani et al., 1990; Kamp et al., 2017). Since recall is highly similar, though not identical, to recollection-based recognition (Yonelinas, 2002) and hippocampal activity around encoding is related to subsequent recollection (see Cohen et al., 2015, for a review), both, the P300 SME and the frontal slow wave SME, might reflect encoding processes that lead to successful subsequent recollection-based remembering. To date, however, there is still debate whether there is a specific ERP component that predicts familiarity-based recognition. Given its relation to semantic processing, the N400 might be a putative predictor for familiarity-based recognition. While early studies reported that the N400 does not predict subsequent memory at all (e.g. Fabiani & Donchin, 1995; Neville et al., 1986), more recent studies provide some evidence that the N400 might be a likely candidate for supporting subsequent familiarity-based recognition (e.g. Kamp et al., 2017; Mangels et al., 2001; Meyer et al., 2007).

In the present study, we wanted to test whether the N400 at encoding is related to familiarity-based recognition and whether the P300 at encoding is related to recollection-based recognition. For this purpose we used a modified Von Restorff paradigm (von Restorff, 1933) at encoding. In the Von Restorff paradigm, a stimulus which pops out from its immediate context is presented. Memory for this distinctive or isolated stimulus during study is enhanced (Von Restorff or isolation effect; von Restorff, 1933). The distinctive stimuli can be quite mundane like differences in physical features as size or color, or semantic features like a category mismatch.

The R/K procedure can be used with few trials and also taps into the phenomenology of recollection and familiarity (Yonelinas, 2002), whereas the PDP requires the learning of sources and two test phases and ROCs require more than 60 trials per condition to be accurately estimated. Therefore, we decided to use the R/K procedure at test.

Since physically or semantically isolated stimuli typically elicit a P300 and semantic isolates additionally elicit an N400 (Fabiani, 2006), we used both, physically and semantically isolated words in our study task. Based on the vast literature on the P300 SME in recall (see Fabiani, 2006, for a review) and the functional overlap between recall and recollection-based recognition (Yonelinas, 2002), the P300 SME might be confined to subsequent recollection-based recognition. We therefore predicted a P300 SME for the physically and semantically isolated words which were subsequently remembered. In contrast, no P300 SME should be observed for physically and semantically isolated words which were only known in the subsequent memory test. Based on the studies reviewed above which provided some evidence for a link between the N400 at encoding and subsequent familiarity-based recognition, we predicted an N400 SME for semantically isolated words which were remembered or known in the subsequent memory test.

4.2.2 Method

4.2.2.1 Participants

24 students of the Saarland University (14 female; median age: 24 yrs, range: 18 – 30 yrs) participated in this study for partial course credit. Two further subjects were excluded due to excessive artifacts in the study phase EEG. All participants gave written informed consent prior to participation.

4.2.2.2 Materials

For the present study 40 word categories – each containing 16 semantically related words – were used. These words were derived from the category norms (Battig & Montague, 1969; McEvoy & Nelson, 1982; Schröder, Gemballa, Ruppin, & Wartenburger, 2011; Van Overschelde, Rawson, & Dunlosky, 2004) and from self-generated categories. The items at the critical list positions were matched for word length ($F(7, 312) = 0.44$, $p = .879$, $\eta_p^2 = .01$) and word frequency ($F(7, 312) = 0.17$, $p = .990$, $\eta_p^2 = .01$) as assessed with dlexDB (Heister et al., 2011).

4.2.2.3 Procedure

The experiment consisted of two phases: a study and a test phase. There was a self-determined break between the study phase and the test phase. At the end of the experiment, participants filled out a questionnaire about the experiment.

Study phase. Forty study lists were presented without any break between the lists. Each list contained 12 words (Table 4.3). Eleven words were members of a category (e.g. mammals). The isolated items as well as the baseline item were presented at the list positions 5-8. The semantic isolate was a study word from a different semantic category (e.g. pretzel). The physical isolate was a study word presented in a different color, but belonged to the same semantic category as the other list words. The baseline item was a normal word from the list category which served as a reference in the subsequent test phase. The first item of a new list category was also treated as a semantic isolate. All remaining study words were filler items which were excluded from the subsequent test phase. The used colors were reversed for half of the participants.

Each trial began with the presentation of a fixation cross for 500 ms (see Figure 4.4). Next, a word in either blue or yellow font was presented for 500 ms followed by a question mark. For half of the participants blue was the color of the isolate and yellow was the color of the other items. The colors were reversed for the other half of the participants. The participants’ task was to judge whether the word has a positive, neutral, or negative meaning by pressing the numpad keys 1, 2, or 3, respectively. The question mark was presented until the participant responded or for 2500 ms, if no response was made. Afterwards, a blank screen was presented for 700 ms plus the difference between 2500 ms minus the response time.

Table 4.3 Organization of a study list. Only items at the positions 1 or 5-8 were used in the subsequent recognition test. All filler items were excluded from the recognition test.

Position	Item status	Example
1	Semantic isolate (first item from list category)	Dog
2-4	Filler items (item from list category not in the test phase)	Cow
5-8	Baseline item (item from list category)	Horse
	Physical isolate (item from list category in different color)	Rabbit
	Semantic isolate (item from different category)	Pretzel
	Filler item (item from list category not in the test phase)	Deer
8-12	Filler items (item from list category not in the test phase)	Donkey

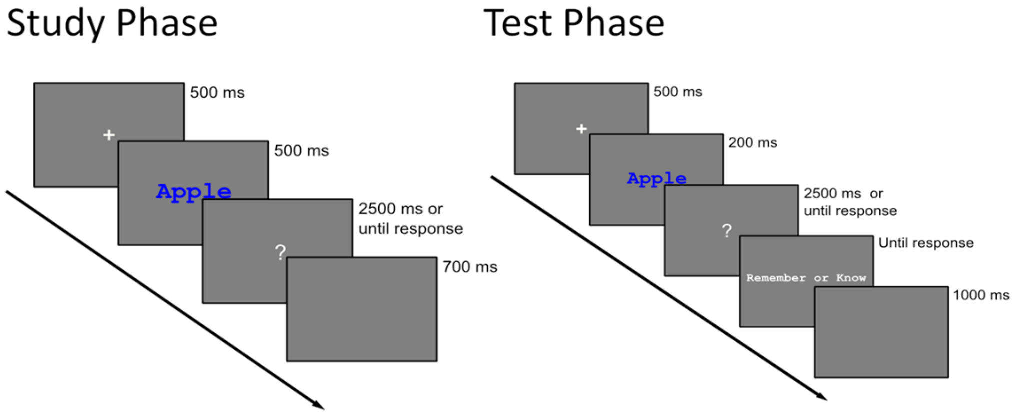

Figure 4.4 Trial procedure for the study phase and the test phase.

Test phase. In the test phase, 160 old (baseline, physical isolated, semantic isolated) and 160 new items (from the same categories as study items) were presented in the non-isolated color in order to increase familiarity-based recognition.

Each trial began with the presentation of a fixation cross for 500 ms followed by the test word which was presented for 200 ms (see Figure 4.4). First, participants were instructed to make an old/new recognition judgment within a time window of 2500 ms. After an old response, they were required to make a remember/know judgment without a time limit. Each trial ended with the presentation of a blank screen for 1000 ms.

4.2.2.4 EEG Recording and ERP Preprocessing

An elastic cap (Easycap, Herrsching, Germany) with 58 embedded silver/silverchloride EEG electrodes was attached to the participant's head. EEG was continuously recorded from Fp1, Fpz, Fp2, AF3, AF4, F7, F5, F3, F1, Fz, F2, F4, F6, F8, FT7, FC5, FC3, FC1, FCz, FC2, FC4, FC6, FT8, T7, C5, C3, C1, Cz, C2, C4, C6, T8, TP7, CP5, CP3, CP1, CPz, CP2, CP4, CP6, TP8, P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO3, POz, PO4, PO8, O1, Oz, O2, as well as from the right mastoid (M2). The reference was placed on the left mastoid (M1) and the ground electrode was placed on AFz. EOG activity was recorded with two electrodes placed on the outer canthi and by a pair of electrodes placed above and below the right eye. Electrode impedances were kept below 5 kΩ. Data were sampled at 500 Hz and filtered online from 0.016 to 250 Hz.

Offline, EEG data were processed with the Brain Vision Analyzer 2.0.3 (Brain Products, Gilching, Germany). Data were downsampled to 200 Hz and a high-pass filter at 0.1 Hz was applied. ECG, muscle, and ocular artifacts were removed using an ICA. After the ICA correction, data were re-referenced to linked mastoids and a low-pass filter at 30 Hz (48 dB/oct) was applied. Next, data were segmented into epochs of 2200 ms (including 200 ms pre-stimulus baseline) and a baseline correction was performed. Data were screened for remaining artifacts and all segments that contained amplitudes outside the range of -100 to 100 µV or voltage steps exceeding 50 µV/ms were removed.

4.2.2.5 Data Analysis

Preliminary data analyses did not reveal any differences between semantic isolated words at the first position and semantic isolates at the positions 5-8. Therefore, the data for semantic isolates were collapsed across position.

We analyzed the data using a one-way ANOVA with the factor Condition (physical isolate vs. semantic isolate vs. baseline item), separately for the ratings and the reaction times. Since expectancy-violating events are often evaluated less favorably than expectancy-consistent events (e.g. Bartholow et al., 2001; Mendes, Blascovich, Hunter, Lickel, & Jost, 2007), we tested whether baseline items received higher ratings and were associated with lower reaction times than the isolated items with contrasts. We refrained from including subsequent memory performance as a factor, because preliminary analyses did not reveal any effects for subsequent memory performance.

In order to test whether memory was superior for the isolated items, we analyzed the data from the R/K task using the approach by Yonelinas and Jacoby (1995; Recollection = R; Familiarity = K/(1 – R)). These data were subjected to an ANOVA with the factors Condition (physical isolate vs. semantic isolate vs. baseline item) and R/K Judgment (Remember vs. Know).

Reaction times for the test phase were subjected to an ANOVA with the factors Condition (physical isolate vs. semantic isolate vs. baseline item) and R/K Judgment (Remember vs. Know vs. Forgotten). Reaction times for all conditions in the test phase, however, were available for only 22 participants.

A priori, we defined the time window from 300 to 500 ms and from 500 to 700 ms as the time windows of interest and Fz, Cz, and Pz as the electrodes of interest for our ERP analysis. We expected the N400 to be primarily present in the early time window from 300 to 500 ms. The P300 should be mainly present in the late time window from 500 to 700 ms, because the P300 peak is often around 600 ms for words (e.g. Kutas, McCarthy, & Donchin, 1977). The first ERP analysis was intended as manipulation check, i.e. we wanted to check whether the semantic isolates elicited an N400 and the physical isolates elicited a P300. For this purpose, a repeated-measure ANOVA with the factors Time Window (early vs. late), Condition (semantic isolate vs. physical isolate vs. baseline item), and Electrode (Fz vs. Cz vs. Pz) and the mean amplitude as dependent variable was conducted. The second ERP analysis tested whether there was an N400 SME and a P300 SME using a repeated-measure ANOVA with the factors Time Window (early vs. late), Response (remembered vs. know vs. forgotten), and Electrode (Fz vs. Cz vs. Pz) and the mean amplitude as dependent variable. Please note that main effects for electrode or interactions involving only electrode and another factor were not followed-up.

In order to provide a more comprehensive picture of the subsequent memory effect, we also analyzed the N400 and P300 separately for the semantic isolates, the physical isolates, and the baseline items. Again, we conducted a repeated-measure ANOVA with the factors Time Window (early vs. late), Response (remembered vs. know vs. forgotten), and Electrode (Fz vs. Cz vs. Pz) and the mean amplitude as dependent variable with those participants who had six or more trials per condition. Therefore, 20 participants were included in the analysis of the semantic isolates, 21 participants were included in the analysis of the physical isolates, and 21 participants were included in the analysis of the baseline items.

All statistical analyses were conducted using IBM SPSS 24 (IBM Corp., Armonk, NY, USA). For all repeated-measure ANOVAs, sphericity was assessed using Mauchly’s Test and Greenhouse-Geißer correction was applied if necessary. The significance level was set to p = .050 for all analyses.

4.2.3 Results

4.2.3.1 Behavioral Results

Study phase. The analysis of the rating data in the study phase (Table 4.4.) revealed a marginally significant effect for Condition (F(2, 46) = 2.95, p = .062, η_p^2 = .11). As expected, the participants rated the baseline words more positive than the semantic and physical isolates (F(1, 23) = 5.01, p = .035, η_p^2 = .18). The ratings did not differ between semantic and physical isolates (F(1, 23) = 0.10, p = .751, η_p^2 = .00).

The analysis of the reaction times in the study phase (Table 4.4) revealed a main effect for Condition (F(2, 46) = 21.04, p < .001, η_p^2 = .48). The participants were faster in rating baseline items as compared to semantically or physically isolated words (F(1, 23) = 19.25, p < .001, η_p^2 = .456) and they were also faster in rating physically isolated words as compared to semantically isolated words (F(1, 23) = 23.02, p < .001, η_p^2 = .50).

Table 4.4 Mean and standard deviations for valence ratings (N = 24) and reaction times (N = 22) in the study phase.

	Valence Ratings		Reaction Items	
	M	SD	M	SD
Semantic Isolates	2.17	0.13	628	223
Physical Isolates	2.16	0.13	560	226
Baseline Items	2.21	0.14	538	192

Test phase. The analysis of the corrected R/K data (Table 4.5) revealed that there was neither a main effect for Condition (*F*(2, 46) = 0.74, *p* = .483, η_p^2 = .03), nor a main effect for R/K Judgment (*F*(1, 23) = 0.06, *p* = .809, η_p^2 = .00), nor an interaction between Condition and R/K Judgment (*F*(2, 46) = 0.40, *p* = .675, η_p^2 = .02). Thus, no Von Restorff effect was found.

The analysis of the reaction times to old items in the test phase (Table 4.6.) revealed a main effect for R/K Judgment (*F*(2, 42) = 12.30, *p* < .001, η_p^2 = .37). Contrary to our expectations, the participants' old/new judgments were faster for remembered items than for known items or forgotten items (*F*(1, 21) = 29.38, *p* < .001, η_p^2 = .58). There was no difference in reaction times between known and forgotten items (*F*(1, 21) = 0.34, *p* = .565, η_p^2 = .02). However, neither the main effect for condition nor the interaction between Condition and R/K Judgment were significant (*F*(2, 42) = 0.93, *p* = .404, η_p^2 = .04 and *F*(2.35, 49.27) = 0.75, *p* = .499, η_p^2 = .03, respectively).

Table 4.5 Mean and standard deviation for remembered and known items in the test phase. Please note that the know responses were corrected using the procedure by Yonelinas and Jacoby (1995).

	Semantic isolates		Physical isolates		Baseline items	
	M	SD	M	SD	M	SD
Remember	.55	.22	.53	.22	.57	.23
Know (corrected)	.56	.20	.56	.21	.57	.24

Table 4.6 Reaction times for the old/new response in the test phase (N = 22).

	Semantic isolates		Physical isolates		Baseline items	
	M	SD	M	SD	M	SD
Remember	761	130	781	122	781	115
Know	931	216	928	248	931	243
Forgotten	952	214	994	311	919	284

4.2.3.2 ERP Results

4.2.3.2.1 Manipulation Check

As expected and as can be seen in Figure 4.5, a P300 to physical and semantic isolates and an N400 to semantic isolates were observed in the ERPs of the study phase. This impression was corroborated by the statistical analysis. There was a main effect for Time Window ($F(1, 23) = 26.78$, $p < .001$, $\eta_p^2 = .54$) and for Condition ($F(2, 46) = 26.35$, $p < .001$, $\eta_p^2 = .53$), as well as an interaction between Time Window and Condition ($F(2, 46) = 3.93$, $p = .026$, $\eta_p^2 = .15$) and between Electrode and Condition ($F(2.66, 61.14) = 6.82$, $p = .001$, $\eta_p^2 = .23$). No other effects were statistically significant (all *F*s < 2.47, all *p*s > .126).

The follow-up ANOVA for the early time window revealed a significant effect for Condition ($F(2, 46) = 26.65$, $p < .001$, $\eta_p^2 = .54$). As expected for the N400, the semantic isolates were significantly more negative than the physical isolates and the baseline items ($F(1, 23) = 56.72$, $p < .001$, $\eta_p^2 = .71$). However, the baseline items were also more negative than the physical isolates ($F(1, 23) = 4.89$, $p = .037$, $\eta_p^2 = .18$), most likely due to onset of the P300 at around 400 ms.

The follow-up ANOVA for the late time window revealed a significant effect for Condition ($F(2, 46) = 19.03$, $p < .001$, $\eta_p^2 = .45$). As expected for the P300, the physical isolates were significantly more positive than the semantic isolates and the baseline items ($F(1, 23) = 52.54$, $p < .001$, $\eta_p^2 = .70$). There was no difference between the baseline items and the semantic isolates ($F(1, 23) = 0.31$, $p = .585$, $\eta_p^2 = .01$).

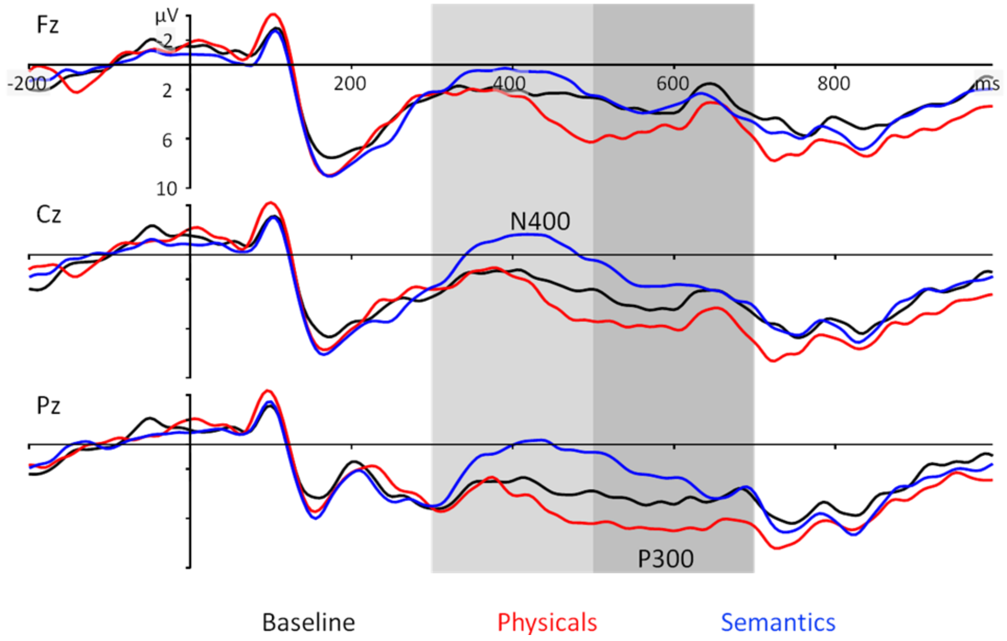

Figure 4.5 ERPs of the study phase to semantic isolates (red), physical isolates (blue), and baseline items (black). The bright grey bar marks the early time window from 300 to 500 ms. The dark grey bar marks the late time window from 500 to 700 ms. The N400 was most pronounced for the semantic isolates in the early time window, whereas the P300 was most pronounced in the physical isolates in the end of the early time window and most of the late time window.

4.2.3.2.2 Subsequent Memory Effects Across Word Categories

As shown in Figure 4.6., subsequently remembered words were more positive than subsequently forgotten or known words. This pattern was also reflected in the statistical results. There was a main effect for Time Window (F(1, 23) = 18.66, p < .001, η_p^2 = .45). The ERP amplitudes were in general more positive in the late time window as compared to the early time window. Furthermore, there was a main effect for Response (F(2, 46) = 6.83, p = .003, η_p^2 = .23). As indicated by follow-up contrasts, remembered words were more positive than forgotten and known words (F(1, 23) = 14.61, p = .001, η_p^2 = .39). There was no difference between forgotten and known words (F(1, 23) = 0.45, p = .510, η_p^2 = .02). No other effects were statistically significant (all Fs <2.43, all ps > .106).

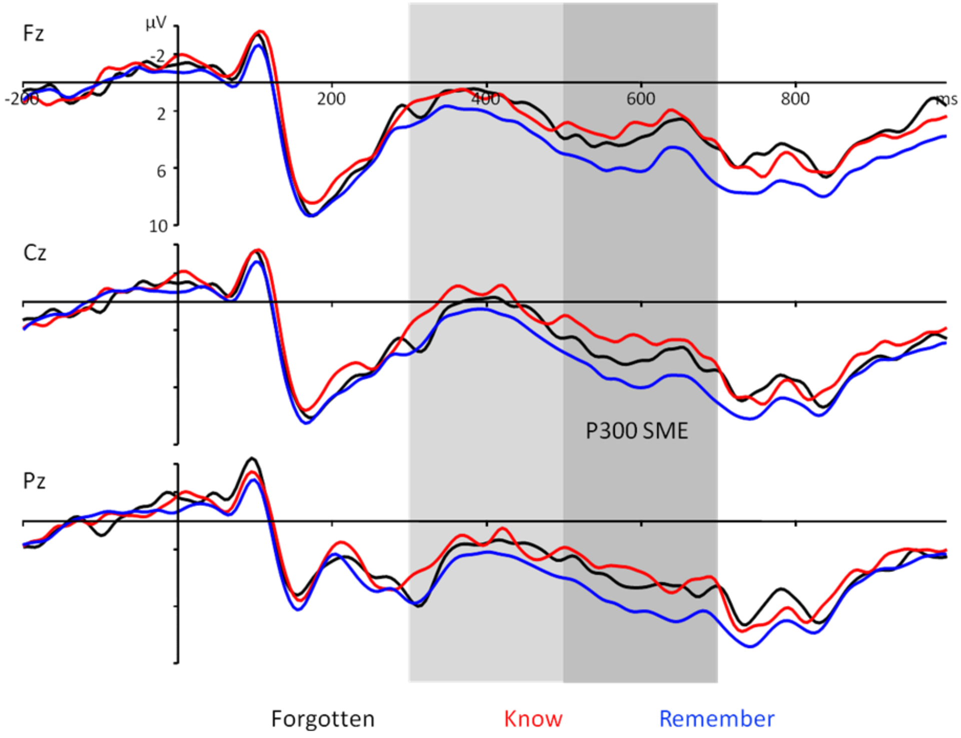

Figure 4.6 ERP subsequent memory effects for remembered (blue), known (red), and forgotten (black) words. The bright grey bar marks the early time window from 300 to 500 ms. The dark grey bar marks the late time window from 500 to 700 ms. A P300 subsequent memory effect was observed in the late time window. No N400 subsequent memory effect emerged.

In order to get a more comprehensive picture of the influence of isolation on the SMEs, separate analyses on subsamples with sufficient trials for semantic (N = 20) and physical isolates (N = 21) as well as baseline items (N = 21) were conducted.

4.2.3.2.3 N400 and P300 Subsequent Memory Effect in Semantically Distinctive Words

As shown in Figure 4.7 (Left column), no N400 SME could be observed in the early time window. However, there was a SME in the late time window. This SME differentiated between the remembered words on the one hand and the known or forgotten words on the other hand. This pattern was also found in the results of the statistical analysis.

The analysis of the semantic isolates revealed a main effect for Time Window (F(1, 19) = 23.36, p < .001, η_p^2 = .55), a main effect for Response (F(2, 38) = 3.62, p = .036, η_p^2 = .16), and an interaction between Time Window and Response (F(1.43, 27.08) = 5.40, p = .018, η_p^2 = .22). No other effects were statistically significant (all Fs <2.58, all ps > .120). Follow-up analyses for the interaction revealed that there was subsequent memory effect in the late time window (F(2, 38) = 6.35, p = .004, η_p^2 = .25), but not in the early time window (F(2, 38) = 0.97, p = .387, η_p^2 = .05). The ERPs to remembered words were more positive than the ERPs to known or forgotten words (F(1, 19) = 10.80, p = .004, η_p^2 = .36). There was no difference between known and forgotten words (F(1, 19) = 0.90, p = .354, η_p^2 = .05).

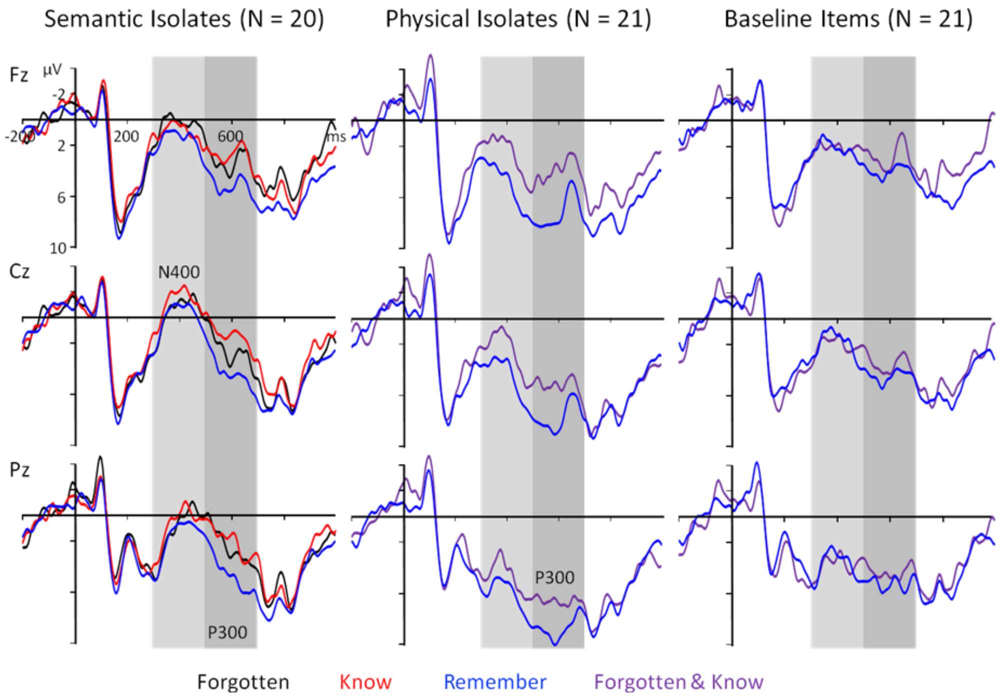

Figure 4.7 ERP subsequent memory effects for remembered (blue), known (red), and forgotten (black) words in the semantic isolate condition (left column) and for remembered (blue) and known or forgotten words (purple) in the physical isolate (middle column) and baseline (right columns) condition. Please note that forgotten and known words (purple) were collapsed for physical isolates and baseline items due to insufficient number of trials. The bright grey bar marks the early time window from 300 to 500 ms. The dark grey bar marks the late time window from 500 to 700 ms.

4.2.3.2.4 P300 Subsequent Memory Effect in Physically Distinctive Words

As shown in Figure 4.7 (middle column), the P300 amplitude at encoding was significantly larger for remembered than for known and forgotten items (P300 SME). Due to insufficient trial numbers, ERPs were collapsed across forgotten and known items for physical isolates and for baseline items.

The analysis of the physical isolates revealed a main effect for Time Window ($F(1, 20) = 15.49$, $p = .001$, $\eta_p^2 = .44$). The ERP mean amplitudes were more positive in the late time window as compared to the early time window. There was a marginally significant main effect for Electrode ($F(1.09, 21.75) = 3.69$, $p = .065$, $\eta_p^2 = .16$) indicating that the amplitudes increased from Fz to Pz. Furthermore, there was a main effect for Response ($F(1, 20) = 6.83$, $p = .017$, $\eta_p^2 = .25$). Critically, the interaction between Time Window and Response was not significant ($F(1, 20) = 1.23$, $p = .281$, $\eta_p^2 = .06$). Thus, the subsequent memory effect was independent of the time window – most likely due to the early onset of the P300. The ERPs to remembered words were more positive than the ERPs to known or forgotten words. No other effects were statistically significant (all *F*s <1.90, all *p*s > .180).

4.2.3.2.5 Subsequent Memory Effects in the Baseline Items

As shown in Figure 4.7 (right column), there was no SME for the baseline items. The results from the statistical analysis are consistent with this impression. There was only a significant main effect for Time Window ($F(1, 20) = 12.37$, $p = .002$, $\eta_p^2 = .38$) and a marginally significant main effect for Electrode ($F(1.17, 23.47) = 2.88$, $p = .098$, $\eta_p^2 = .13$). No other effects were statistically significant (all *F*s <1.78, all *p*s > .196).

4.2.4 Discussion

In the present ERP study, we investigated whether the N400 at encoding was related to familiarity-based recognition and whether the P300 at encoding was related to recollection-based recognition. Even though no Von Restorff effect was found on a behavioral level, the ERP data revealed P300 SMEs for remembered semantically and physically isolates. However, no N400 SME was found.

High overall memory performance and intentional learning may account for the absence of a Von Restorff effect in the behavioral data and of an N400 SME. Meyer et al. (2007) used an incidental encoding task to prevent strategic processing from changing semantic processing. Another difference between the study by Meyer et al. (2007) and our study was that we used the same nouns we presented in the study phase, whereas Meyer et al. (2007) presented sentences in the study phase, but only tested nouns and verbs in the test phase. In addition, we used the non-distinctive color of the study phase for all items in the test phase to encourage familiarity-based recognition for the semantic isolates and recollection-based recognition for the physical isolates. However, this procedure might have resulted in an enhanced recognition for the baseline items and reduced familiarity-based recognition for the physical isolates. Still another explanation for the absence of the Von Restorff effect might be the use of elaborative encoding strategies. Even though our task requires only rote semantic encoding, which produces robust Von Restorff effects (Fabiani & Donchin, 1995; Fabiani et al., 1990), 17 out of 24 participants reported to have formed associations at encoding. Elaborated encoding leads to different behavioral and ERP effects (Fabiani et al., 1990). Thus, incidental study tasks which encourage rote encoding and selection of a new font color for the test phase should be used to increase the difficulty of the recognition task in future studies.

However, the results from the rating task provide some evidence that distinctiveness affected the processing of the items at encoding. The more positive evaluation of the baseline words relative to the isolated words might be a manifestation of the mere exposure effect (Zajonc, 1968), i.e. the participants developed a preference for the typical and frequently presented category members. Furthermore, the longer reaction times for the rating of the physical and semantic isolates relative to the baseline items indicate that distinctiveness disrupted the processing of these items and more time was necessary for encoding (Stroessner et al., 1992). In addition, the longer reaction times for rating the semantic isolates relative to physical isolates together with the N400 component might index the additional semantic processing that was necessary for the semantic isolates. Furthermore, participants reported higher perceived distinctiveness for semantic isolates relative to the physical isolates in the post-experiment questionnaire (z = -3.16, p = .002).

The reaction time results from the test phase may also shed some light on the absence of a difference in the K judgments between the conditions. The reaction times for K were considerably slower than the reaction times for R responses. Since we did not offer a Guess response option, the reaction times might indicate that the K response might be confounded with guessing. Using the Remember-Know-Guess procedure (Gardiner, Ramponi, & Richardson-Klavehn, 1998) might alleviate this problem in future studies. An alternative explanation for the reaction time results is that participants waited for recollection to occur and only pressed the K button after they were sure that they would not be able to recollect the item.

The ERP results provide support for the view that P300 activity at encoding is linked to subsequent recollection-based recognition. Our study contributes to the literature that relates the P300 with subsequent recall performance especially for distinctive items (see Fabiani, 2006, for a review). These results are in line with other studies which also reported a link between positive ERPs and subsequent memory performance (e.g. Otten & Rugg, 2001; Paller, McCarthy, & Wood, 1988; Sanquist et al., 1980). Friedman and Trott (2000) used the R/K procedure and also obtained an SME for subsequently remembered, but not for subsequently known items.

The hypothesis that N400 at encoding is related to subsequent familiarity-based recognition could, in contrast, not be supported by the present findings. In fact, there was no N400 SME at all in the encoding phase. This result was surprising in light of the evidence reviewed in the introduction. Interestingly, some hints for an N400-like SME with a frontocentral topography can be found in the physical isolate condition. A post-hoc analysis of the N400-like component (using the time window 280-420 ms in order to avoid an overlap with the P300) revealed a significant difference between subsequently remembered words and subsequently known or forgotten words (F(1, 20) = 4.94, p = .038, η_p^2 = .20). Kamp et al. (2017) found a similar N400-like SME in a condition that prompted participants to integrate multiple items into a unitized concept. Our participants might have tried to integrate the deviant color and the word for the subsequent memory test. However, further studies are needed to test this notion.

4.3 Conclusions from Experiment 4 and 5

Experiment 4 assessed the relative merit of distinctiveness and accentuation within the AT framework. The results from the P300 indicate that distinctiveness rather than accentuation seems to matter at encoding. As expected from a distinctiveness perspective, the P300 was higher for shared distinctiveness than for distinctiveness on a single dimension. The ERP results from Experiment 5 provide support for the view that P300 activity at encoding is linked to subsequent recollection-based recognition. The hypothesis that N400 at encoding is related to subsequent familiarity-based recognition could, in contrast, not be supported by the findings of Experiment 5.

The Experiments 3 and 4 provided some support for the SDA and Experiment 5 revealed that the P300 is related to recollection-based recognition. On the basis of these findings we could derive detailed predictions for an ERP experiment that tests the effect of shared distinctiveness on memory. Since shallow encoding conditions lead to both, a strong IC (e.g. Fiedler, 2000) and a strong P300 SME (e.g. Fabiani, 2006), the P300 seems ideally suited for the investigation of shared distinctiveness in the IC paradigm. The main objective of Experiment 6, which will be presented in the next chapter, was to find out, whether source memory is enhanced by shared distinctiveness and whether the behavioral effects can be referred back to the P300 SME.

5 Electrophysiological Investigation of the Distinctiveness-Based Illusory Correlation

5.1 Introduction

Memory for extraordinary events is often superior to memory for ordinary events (e.g. Schmidt, 1991, 2012; von Restorff, 1933). According to Schmidt (2012) distinctiveness can arise from four different sources (primary and secondary distinctiveness, emotional significance, or high-priority stimuli) and all four sources have been shown to make an event more memorable. Stimuli or events become even more memorable, if they have two or more distinctive features (e.g. Hunt & Mitchell, 1982; Kuhbandner & Pekrun, 2013). Furthermore, distinctiveness can also affect primarily memory-based frequency judgments (Tversky & Kahneman, 1973). Tversky and Kahneman (1973) proposed that people use the availability (i.e. the ease of retrieval) of memories at the time of judgment to gauge the frequency of occurrence. This is also the case for covariation judgments (Tversky & Kahneman, 1973). Since distinctive memories are highly available, these memories have a strong impact on frequency judgments and on covariation judgment (e.g. Rothbart, Fulero, Jensen, Howard, & Birrell, 1978; Tversky & Kahneman, 1973). Distinctiveness plays a particular important role in the IC (e.g. Chapman, 1967; Hamilton & Gifford, 1976).

According to the SDA (Hamilton et al., 1985; Hamilton & Gifford, 1976; see section 2.1.2.1), shared distinctiveness should promote encoding and, as a consequence, increases the availability of the most infrequent category combinations. Since availability contributes to frequency judgments in addition to the actual frequency of occurrence (Tversky & Kahneman, 1973), the increased availability of rare category combinations in memory leads to their overestimation in frequency and, consequently, to ICs (e.g. Hamilton, 1981; Hamilton et al., 1985; Hamilton & Gifford, 1976). As outlined in section 2.1.2.1, the empirical evidence for the SDA from behavioral experiments so far is equivocal. In Experiment 3, however, we found that source memory for negative behavior was elevated for the minority even after controlling for response bias and primacy or recency effects as predicted by the SDA. Furthermore, source memory predicted the extent of IC in the other measures.

The P300 is an ERP component associated with the processing of (primary) distinctiveness (e.g. Fabiani, 2006; Polich, 2007). The P300 at encoding is related to subsequent memory performance and reflects the encoding of item-specific information (e.g. Fabiani et al., 1990; Kamp et al., 2017). Furthermore, the P300 SME seems to be confined to subsequent recognition that is based on recollection, i.e. the retrieval of contextual details (Mangels et al., 2001; see also Experiment 5). Since distinctive events typically elicit a P300 and the P300 amplitude is predictive for subsequent memory performance, the P300 seems ideally suited for the investigation of shared distinctiveness in the illusory correlation paradigm.

In the present ERP study, we investigated the behavioral effects of shared distinctiveness on the P300 at encoding and on source memory at later testing. For this purpose, we used a methodologically optimized version of the Hamilton and Gifford (1976) IC paradigm introduced in Experiment 3. This paradigm not only takes into account primacy and recency effects, but also eliminates confounds between discrimination and response bias that arises from the skewed frequency distributions. We hypothesized that shared distinctiveness should lead to an enhanced P300 subsequent memory effect, better source memory performance, and an IC.

5.2 Methods

5.2.1 Participants

Forty healthy, right-handed students of the Saarland University (31 female; median age: 22.5 years; range: 19-30 years) participated in this study for partial course credit. Four additional participants had to be excluded due to lack of compliance (three participants reported napping during the experiment and one participant took a break during an experimental task). All participants gave written informed consent prior to participation.

5.2.2 Materials

320 positive and 160 negative words were selected from a list of word norms provided by Lahl, Göritz, Pietrowsky, and Rosenberg (2009). Word frequency information was taken from the database dlexDB (Heister et al., 2011). Positive and negative words could be matched for arousal, concreteness, word length, and word frequency. However, items could not be matched for intensity (or extremity; i.e. the valence ratings that were converted to a common scale), a factor known to affect memory for emotional words (Kamp et al., 2015). Thus, positive items were more positive than negative items were negative. The descriptive statistics for the material can be found in Appendix C.

For the study phase, 240 positive and 120 negative words were divided into ten lists of 36 words. Each list contained 16 positive and 8 negative words presented in the majority source color (e.g. purple) and 8 positive and 4 negative words presented in the minority source color (e.g. orange; see Table 1). Critically, the three words at the beginning and the three words at the end of a study list (i.e. 4 positive and 2 negative words from the majority source) were always items in the majority color and not used in the test list in order to prevent primacy and recency effects (see Experiment 3).

For the test phase, ten lists of 36 words were presented, each list containing 24 positive and 12 negative words. For the positive words, eight words were from the majority source color, eight words were from the minority source color, and eight words were new (see Table 5.1). For the negative words, four words were from the majority source color, four words were from the minority source color, and four words were new.

Table 5.1 Distribution of positive and negative words in the study phase and the test phase.

	Study phase		Test phase	
	Positive	Negative	Positive	Negative
Majority	16	8	8	4
Minority	8	4	8	4
New	-	-	8	4

5.2.3 Procedure

The experiment consisted of ten study-test cycles (Figure 5.1). Each cycle began with a study phase, followed by a distracter 2-back task and the test phase. At the end of the experiment, participants estimated the relative frequency of negative items for each source in the last cycle and in the whole experiment in a post-experimental questionnaire.

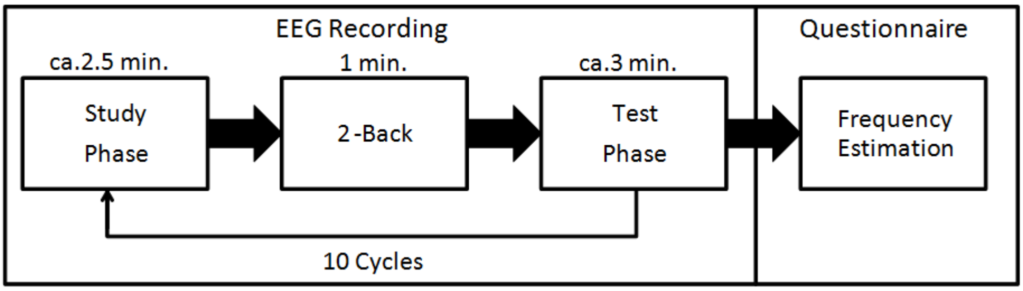

Figure 5.1 The procedure of the whole experimental session. The questionnaire with the frequency estimation task was handed to the participants only after they completed all 10 study-test cycles.

Study phase. Each trial in the study phase had the following structure (Figure 5.2 left): The trial began with a fixation cross presented for 500 ms. Next, a word was presented either in purple (R: 128, G: 0, B: 128) or orange (R: 255, G: 165, B: 0) for 500 ms followed by a blank screen for 1000 ms. We conducted a rating study prior to the experiment and chose purple and orange, because the ratings indicated that these colors have the least affective or semantic connotations. For half of the participants, words in purple were the majority source and words in orange were the minority source. For the other half, the colors were reversed. The participants were instructed to remember the word and its color for the subsequent memory test. Next, a screen prompted the participants with the question “Remember word and color”. The participants were instructed to make a judgment of learning (JOL; e.g. Nelson & Narens, 1990), i.e. to rate how likely they would remember the word and its color on a scale from 1 (“definitely will not remember”) to 6 (“definitely will remember”). When the participants chose their answer, the next trial began.

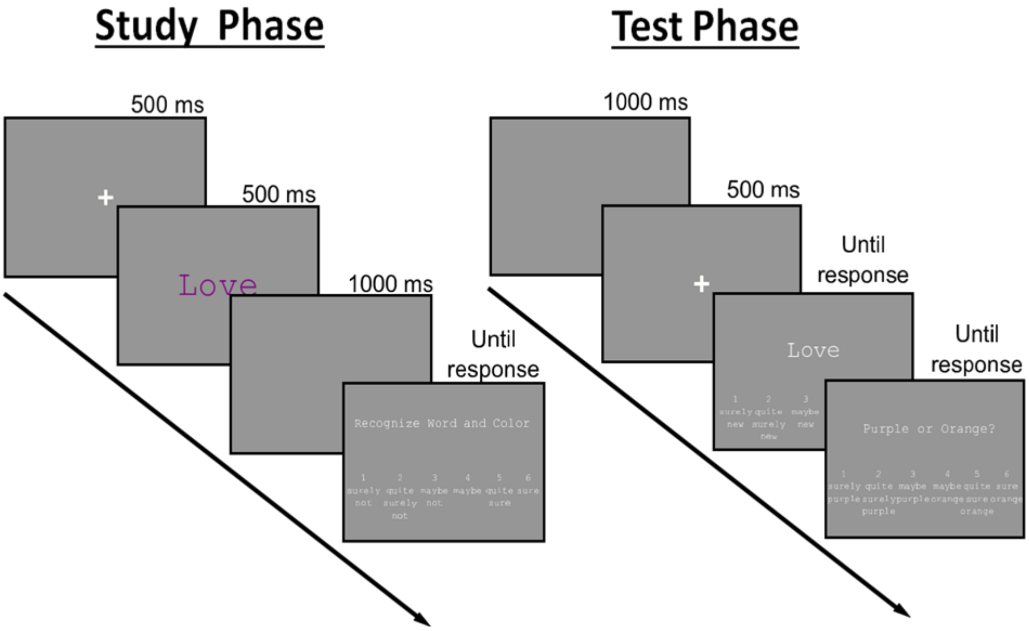

Figure 5.2 Procedure of the study phase (left) and test phase (right).

2-back task. Each trial of the 2-back task had the following structure: A number between one and four was presented for 500 ms followed by a fixation cross for 1500 ms. Participants were required to press the space bar when the presented number matched the number presented two trials before. The stimuli which required responses was treated as targets (p = .25) and the other stimuli as standards (p = .75). The 2-back task was introduced to prevent effects from immediate perceptual repetition (e.g. Grillon, Johnson, Krebs, & Huron, 2008) and to clear working memory. Because the IC paradigm as introduced by Hamilton and Gifford (1976) has not yet been investigated with ERPs and the

2-back task has been shown to elicit a P300 (e.g. Zaehle, Sandmann, Thorne, Jäncke, & Herrmann, 2011), it also served as a control that a typical P300 response can be observed in our sample.

Test phase. Each trial in the test phase had the following structure (Figure 5.2 right): The trial began with a blank screen presented for 1000 ms followed by a fixation cross presented for 500 ms. Next, a word was presented in white color. This word was either a new word or a word from the study phase (either from the majority source or the minority source). As shown in Table 5.1, the frequency for majority, minority, and new items were equated, thereby reducing response bias (see Experiment 3). Furthermore, words at the beginning or the end of a study list were excluded from the test list in order to eliminate primacy and recency effects as in Experiment 3. The participants had to make an old/new judgment on a six-point scale ranging from “surely new” to “surely old”. If the participants chose old, they then had to make a judgment on the source (majority color or minority color) on a six-point scale. There was no time limit for the old/new and source judgments. After their response, the next trial started.

5.2.4 EEG Recording and Preprocessing

An elastic cap (Easycap, Herrsching, Germany) with 58 embedded Ag/AgCl EEG electrodes was attached to the participant’s head. EEG was continuously recorded from Fp1, Fpz, Fp2, AF3, AF4, F7, F5, F3, F1, Fz, F2, F4, F6, F8, FT7, FC5, FC3, FC1, FCz, FC2, FC4, FC6, FT8, T7, C5, C3, C1, Cz, C2, C4, C6, T8, TP7, CP5, CP3, CP1, CPz, CP2, CP4, CP6, TP8, P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO3, POz, PO4, PO8, O1, Oz, O2, as well as from the right mastoid (M2). The reference as placed on the left mastoid (M1) and the ground electrode was placed on AFz. EOG activity was recorded with two electrodes placed on the outer canthi and by a pair of electrodes placed above and below the right eye. Electrode impedances were kept below 5 kΩ. Data were sampled at 500 Hz and filtered online from 0.016 to 250 Hz.

Offline, EEG data were processed with the Brain Vision Analyzer 2.0.3 (Brain Products, Gilching, Germany). Data were down-sampled to 200 Hz and a high-pass filter at 0.1 Hz was applied. Cardiovascular, muscle and ocular artifacts were removed using an ICA. After the ICA correction, data were re-referenced to linked mastoids and a low-pass filter at 30 Hz (48 dB/oct) was applied. Next, data of the study phase were segmented into epochs of 2200 ms (including 200 ms pre-stimulus baseline) and baseline correction was performed. Data were screened for remaining artifacts and all segments that contained amplitudes outside the range of -100 to 100 µV or voltage steps exceeding 50 µV/ms were removed.

5.2.5 Data Analysis

Distinctiveness as perceived during encoding was assessed by analyzing the JOLs, because previous studies (Dunlosky et al., 2000; Geraci & Manzano, 2010) indicated that JOLs were suitable as an on-line measure of perceived distinctiveness. The JOLs were averaged for each category of items and a 2 x 2 repeated-measure ANOVA was calculated with the factors Source (majority vs. minority) and Valence (positive vs. negative) and the mean JOL as dependent variable.

For the analysis of source memory performance, we treated the three points "surely new", "quite surely new", and "maybe new" of the confidence rating as new response and the three points "surely old", "quite surely old", and "maybe old" as old response. A similar procedure was used for the source judgments. Source memory performance was measured by calculating the unbiased hit rates (Wagner, 1993) for each type of item (see Experiment 3 and Appendix A.). The resulting values were then arcsine transformed for statistical analysis (Wagner, 1993). A Source (majority vs. minority vs. new) x Valence (positive vs. negative) repeated-measure ANOVA was calculated for the unbiased hit rates. In order to provide a more comprehensive picture of the memory performance, we also analyzed the confidence ratings of the source judgment for old items and the reaction times for correctly identified items. The confidence ratings were subjected to a Source (majority vs. minority) x Valence (positive vs. negative) repeated-measure ANOVA. The reaction times were subjected to a Source (majority vs. minority vs. new) x Valence (positive vs. negative) repeated-measure ANOVA.

In order to assess the extent of IC, we compared the estimated frequency of negative words for the majority and the minority source in the last block and across the whole experiment using dependent t-tests.

Based on Experiment 5, we chose the a priori defined time window from 500 to 700 ms and the electrode Pz for the analysis of the P300 in the study phase. Due to the imbalanced design resulting from the experimental manipulation and from the dependence between remembered and forgotten items (i.e. high memory performance led to fewer forgotten trials and vice versa; see Table S3 for the descriptive statistics on the trial numbers), we decided to analyze the P300 on the single trial level with multilevel linear modeling (MLM; see Finch, Bolin, & Kelley, 2014, for a general introduction). MLM is an alternative to the repeated-measure ANOVA which is especially useful for unbalanced designs (Field, Miles, & Field, 2012) and dependence of trial numbers as in subsequent memory experiments (Tibon & Levy, 2015). The main advantage of MLM over the repeated-measure ANOVA is that there is no need to exclude participants due to low trial numbers in specific experimental conditions and the analysis can be based on the whole sample (Tibon & Levy, 2015).

All data except the single-trial analysis were analyzed using SPSS 24. Significance level was set to p = .050 for all analyses. For all repeated-measure ANOVAs, the sphericity assumption was tested with Mauchly’s test and the Greenhouse-Geisser correction was applied when necessary. For the MLM, we used R 3.4.1 and the package nlme 3.1-131 (Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2018).

5.3 Results

5.3.1 Behavioral Results

Judgments of Learning. The analysis of the JOLs (see Table 5.2 for descriptive statistics) revealed a significant main effect for valence ($F(1, 39) = 15.96$, $p <$ .001, $\eta_p^2 = .29$). Positive words received higher JOLs than negative words. Neither the main effect for source nor the interaction between source and valence were significant (F(1, 39) = 1.44, p = .237, η_p^2 = .04 and F(1, 39) = 0.02, p = .885, η_p^2 < .01, respectively).

Table 5.2 Mean judgment of learning ratings (± SD) for positive and negative words.

	Positive	Negative
Majority	4.14 (0.49)	3.80 (0.46)
Minority	4.23 (0.63)	3.90 (0.65)

Source memory: The analysis of the unbiased hit rates (see Figure 5.3 Top) revealed a significant main effect for source (F(2, 78) = 149.67, p < .001, η_p^2 = .79), but no main effect for valence (F(1, 39) = 0.22, p = .642, η_p^2 = .01). Unbiased hit rates were higher for new items than for old items (F(1, 39) = 257.06, p < .001, η_p^2 = .87). There was no difference between the majority source and the minority source (F(1, 39) = 0.88, p = .354, η_p^2 = .02). Furthermore, there was an interaction between source and valence (F(1.69, 65.78) = 4.29, p = .023, η_p^2 = .10). The follow-up interaction contrasts for the old items revealed that the unbiased hit rates were higher in the majority than in the minority for positive items, but lower in the majority than in the minority for negative items (F(1, 39) = 4.11, p = .050, η_p^2 = .10). The interaction contrast for comparing old and new items was also significant indicating that the difference in unbiased hit rates between positive and negative valence for new items differed from the average difference in unbiased hit rates between positive and negative valence for old items (F(1, 39) = 4.71, p = .036, η_p^2 = .11). Follow-up t-tests did not reveal any significant differences between positive and negative items for the majority source (t(39) = 1.55, p = .065, one-sided, Cohen’s d = 0.24) or minority source (t(39) = -1.25, p = .109, one-sided, Cohen’s d = -0.20). Negative new items, however, were better identified as new than positive new items (t(39) = -1.76, p = .046, one-sided, Cohen’s d = -0.28).

The analysis of the confidence ratings for the source judgments (see Figure 5.3 Middle) revealed a main effect for source (F(1, 39) = 5.21, p = .028, η_p^2 = .12) and a trend for valence (*F*(1, 39) = 3.12, *p* = .085, η_p^2 = .07). There was also a significant interaction between source and valence (*F*(1, 39) = 10.05, *p* = .003, η_p^2 = .21). One-sided t-tests revealed that the participants were more confident in their source judgment to positive majority items than to negative majority items (*t*(39) = 3.70, *p* < .001, one-sided, Cohen's *d* = 0.59). Consistent with our prediction based on the SDA, participants were more confident in their source judgment to negative minority items than to positive minority items (*t*(39) = -1.99, *p* = .027, one-sided, Cohen's *d* = -0.32).

The analysis of reaction times to items attributed to the correct source (see Figure 5.3 Bottom) revealed a main effect for source (*F*(1.58, 61.47) = 10.80, *p* < .001, η_p^2 = .22), but not for valence (*F*(1, 39) = 2.27, *p* = .140, η_p^2 = .06). Reaction times were faster for new items than for old items (*F*(1, 39) = 14.07, *p* = .001, η_p^2 = .27). No differences emerged between majority source and minority source (*F*(1, 39) = 0.77, *p* = .387, η_p^2 = .02). Critically, the interaction between source and valence was significant (*F*(2, 78) = 5.01, *p* = .009, η_p^2 = .11). Participants reacted faster to negative majority items and positive minority items than to positive majority items or negative minority items (*F*(1, 39) = 6.62, *p* = .014, η_p^2 = .15). However, the difference between positive and negative items was similar for old and new items (*F*(1, 39) = 1.90, *p* = .175, η_p^2 = .05).

To sum up, a consistent pattern was observed for all three dependent variables – unbiased hit rates, source confidence ratings, and reaction times – in the source monitoring task. For the majority color, participants were more accurate, more confident, and slower for positive items than for negative items. For the minority items, the reverse pattern was observed. The pattern for the reaction times is a bit surprising as the reverse pattern would have been expected. The observed pattern might reflect a speed-accuracy trade-off.

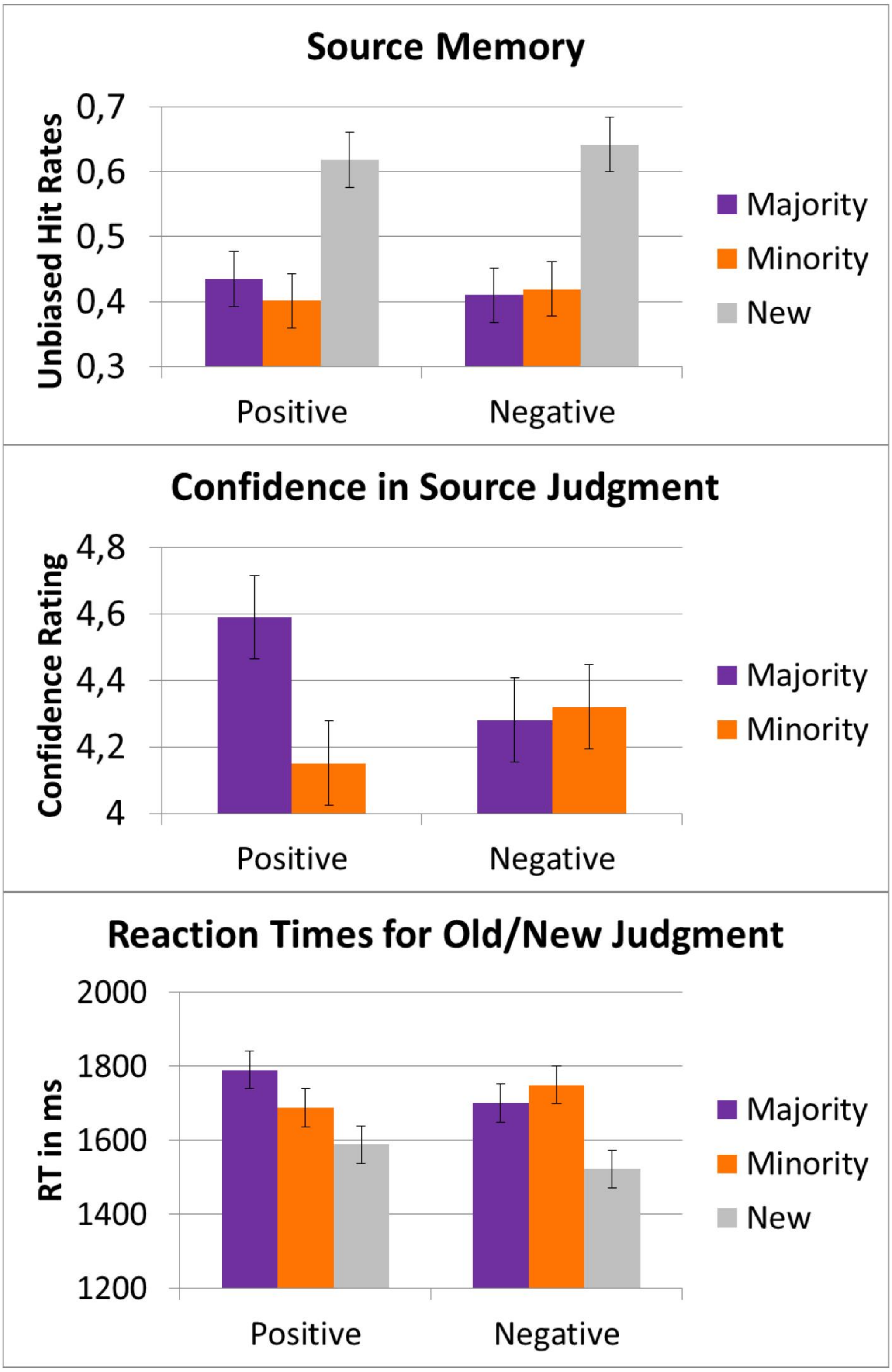

Figure 5.3 Overview over the behavioral results from the test phase. Top: Unbiased hit rates. Middle: Source confidence ratings. Bottom: Reaction times. The error bars represent within-subject 95% confidence intervals for the source x valence interaction.

Frequency judgment: Consistent with our prediction, the analysis of the frequency judgments revealed that the frequency of negative words in the minority source were overestimated relative to the frequency of negative words in the majority source across the whole experiment ($t(39) = - 2.02$, $p = .025$, one-sided, Cohen's $d = 0.32$; see Table 5.3). However, the frequency was not overestimated when participants made judgments for the last block ($t(39) = -1.12$, $p = .136$, one-sided, Cohen's $d = 0.18$). Furthermore, the participants correctly rated the majority source as highly frequent in both, the last block ($t(39) = 4.03$, $p < .001$, one-sided, Cohen's $d = 0.64$) and across the whole experiment ($t(39) = 4.72$, $p < .001$, one-sided, Cohen's $d = 0.75$). However, their estimates were lower than the actual frequency (0.67). We also calculated a phi coefficient from the frequency ratings and found a significant IC ($M = .07$, $SD = .21$; $t(39) = 2.00$, $p = .026$, one-sided, Cohen's $d = 0.36$)

Next, we ran separate multiple regressions for the unbiased hit rates and the confidence ratings in order to assess whether the IC is related to memory performance. To our surprise, neither regression model was significant (unbiased hit rats: $R^2 = .13$, $F(4, 35) = 1.26$, $p = .303$; confidence ratings: $R^2 = .09$, $F(4, 35) = 0.82$, $p = .521$). However, given that the assumption of normality of the residuals and the assumption of homoscedasticity were violated, the results from the regression analysis have to be treated with caution. Replicating the current study with a larger sample size might clarify why source memory performance did not predict the IC.

Table 5.3 Descriptive statistics for the frequency estimation task for the last block and the whole session

	Last block		Whole session	
	M	SD	M	SD
Negative majority words	.36	.16	.39	.15
Negative minority words	.39	.18	.46	.16
Overall frequency of the majority	.60	.15	.61	.15

5.3.2 ERP Results: P300 Subsequent Memory Effect

5.3.2.1 Model Selection

The P300 SME was analyzed with MLM (Field et al., 2012; Finch et al., 2014) at the electrode Pz for the time window from 500 to 700 ms. In order to assess the need for a MLM, we conducted an intercept test (Field et al., 2012). For this purpose, we included participants as random intercept (see Table 5.4 for information on model fit). A comparison of the random intercept model with the intercept only model revealed that the P300 amplitude at Pz varied across participants ($\chi^2(1) = 541.39$, $p < .001$; ICC = .07). Next, we included study block as another random intercept nested in the participants. Again, the P300 differed across study blocks ($\chi^2(1) = 25.17$, $p < .001$; ICC = .07 for participants, ICC = .02 for the blocks nested within participants). We defined this three level model as our baseline model.[4]

In the first analysis, we used weighted effect coding for the categorical variables Valence (Negative: -0.67, Positive: 0.33), Source (Minority: -0.50, Majority: 0.50), and Memory (Forgotten: -0.81, Remember: 0.19) and grand-mean centering for the variable Intensity. Next, we included these variables as well as all interactions between these variables as fixed effects to the baseline model. Intensity was included, because the positive and negative items differed in intensity (see section 5.2.2). The inclusion of these variables significantly improved the model fit ($\chi^2(15) = 35.34$, $p = .002$). Since inclusion of random slopes for Memory did not improve model fit ($\chi^2(4) = 1.57$, $p = .813$), we decided to use the random intercept model for interpretation.

[4] Some might argue that the study block should be treated as a level 1 variable rather than as a level 2 variable. Therefore, we conducted an additional multilevel analysis with study block as Helmert contrast coded level 1 variable. This model rendered virtually identical results as our three-level model. One noteworthy result from this analysis was that the P300 was lower in the first block relative to all other blocks. For the sake of clarity, we decided to use the three-level model for interpretation.

Table 5.4 Information on model fit for the MLM analyses.

	AIC	BIC	Log-Likelihood
Intercept only	70159.63	70173.93	-35077.81
Random intercept for participants	69620.24	69641.70	-34807.12
Random intercept for block nested within participants (baseline model)	69597.08	69625.69	-34794.54
Model 1	69591.74	69727.65	-34776.87
Model 2	69598.17	69762.69	-34776.08

5.3.2.2 Model Interpretation

Information on the coefficients of the final model can be found in Table 5.5. The analysis revealed a significant effect for Memory indicating that the P300 was larger for subsequently remembered items as compared to subsequently forgotten items and a significant effect for Source indicating that the P300 was larger for the rare color than for the common color. In contrast to our hypothesis, there was no interaction between Valence, Source, and Memory. However, there were marginally significant interactions between Valence, Source, and Intensity and between Intensity and Memory.

In order to follow-up the interactions, we used the terciles for Intensity to divide the data set into three subsets, a low, a medium and a high intensity subset (see Figure 5.4 for the ERP waveforms). For the high intensity subset, the follow-up analyses revealed a significant interaction between Valence and Source (*b* = 2.15, *t*(2673) = 2.61, *p* = .009). The P300 was larger for positive items than for negative items of the majority color (*b* = 1.15, *t*(1137) = 2.02, *p* = .044). For the minority color, a trend in the reverse direction was found, i.e. the P300 amplitudes tended to be larger for negative items than for positive items (*b* = -1.00, *t*(1138) = -1.67, *p* = .096). Furthermore, the P300 was significantly larger for negative items in the minority color than for negative items in the majority color (*b* = -2.47, *t*(366) = -3.26, *p* = .001). For positive items, however, no difference between the majority color and the minority color in P300 amplitude was observed (*b* = -0.19, *t*(1990) = -0.49, *p* = .627). Furthermore, there was also a significant effect for Memory in the high intensity subset (*b* = 1.46, *t*(2675) = 3.19, *p* = .002).

However, the follow-up analyses revealed that the interaction between Valence and Source was not significant for the low and medium intensity subset (*b* = -0.54, *t*(2698) = -0.78, *p* = .434 and *b* = -1.06, *t*(2865) = -1.31, *p* = .189, respectively) or when collapsing data across low and medium intensity items (*b* = -0.41, *t*(5966) = -0.84, *p* = .403). Furthermore, the effect for Memory was on a trend level for the low intensity subset (*b* = 0.78, *t*(2700) = 1.71, *p* = .087) and absent in the medium intensity subset (*b* = 0.75, *t*(2867) = 1.64, *p* = .100). However, the effect for Memory became significant when collapsing across low and medium intensity items (*b* = 0.76, *t*(5968) = 2.34, *p* = .019) indicating that the SME was weaker for low and medium intensity items than for high intensity.

To sum up, the ERP data revealed that the P300 amplitude predicted subsequent memory. Even though the SME was not affected by our distinctiveness manipulation, the SME was stronger for high intensity stimuli. Furthermore, shared distinctiveness affected the P300 amplitude for highly intense stimuli. The P300 amplitude was higher for positive majority and negative minority stimuli than for negative majority and positive minority stimuli.

Table 5.5 Coefficients from the selected model

	B	SE B	t(9030)
(Intercept)	3.05	0.42	7.21 (p < .001)
Valence	0.03	0.23	0.15 (p = .880)
Source	-0.76	0.21	-3.59 (p < .001)
Intensity	0.15	0.11	1.34 (p = .179)
Memory	0.96	0.28	3.41 (p < .001)
Valence x Source	0.27	0.46	0.60 (p = .547)
Valence x Intensity	-0.03	0.21	-0.13 (p = .897)
Source x Intensity	0.22	0.23	0.96 (p = .336)
Valence x Memory	-0.30	0.58	-0.51 (p = .610)
Source x Memory	-0.05	0.54	-0.09 (p = .927)
Intensity x Memory	0.58	0.29	1.96 (p = .051)
Valence x Source x Intensity	0.81	0.43	1.88 (p = .060)
Valence x Source x Memory	-0.48	1.16	-0.41 (p = .682)
Valence x Intensity x Memory	0.05	0.55	0.08 (p = .934)
Source x Intensity x Memory	0.21	0.59	0.39 (p = .700)
Valence x Source x Intensity x Memory	0.91	1.10	0.82 (p = .412)

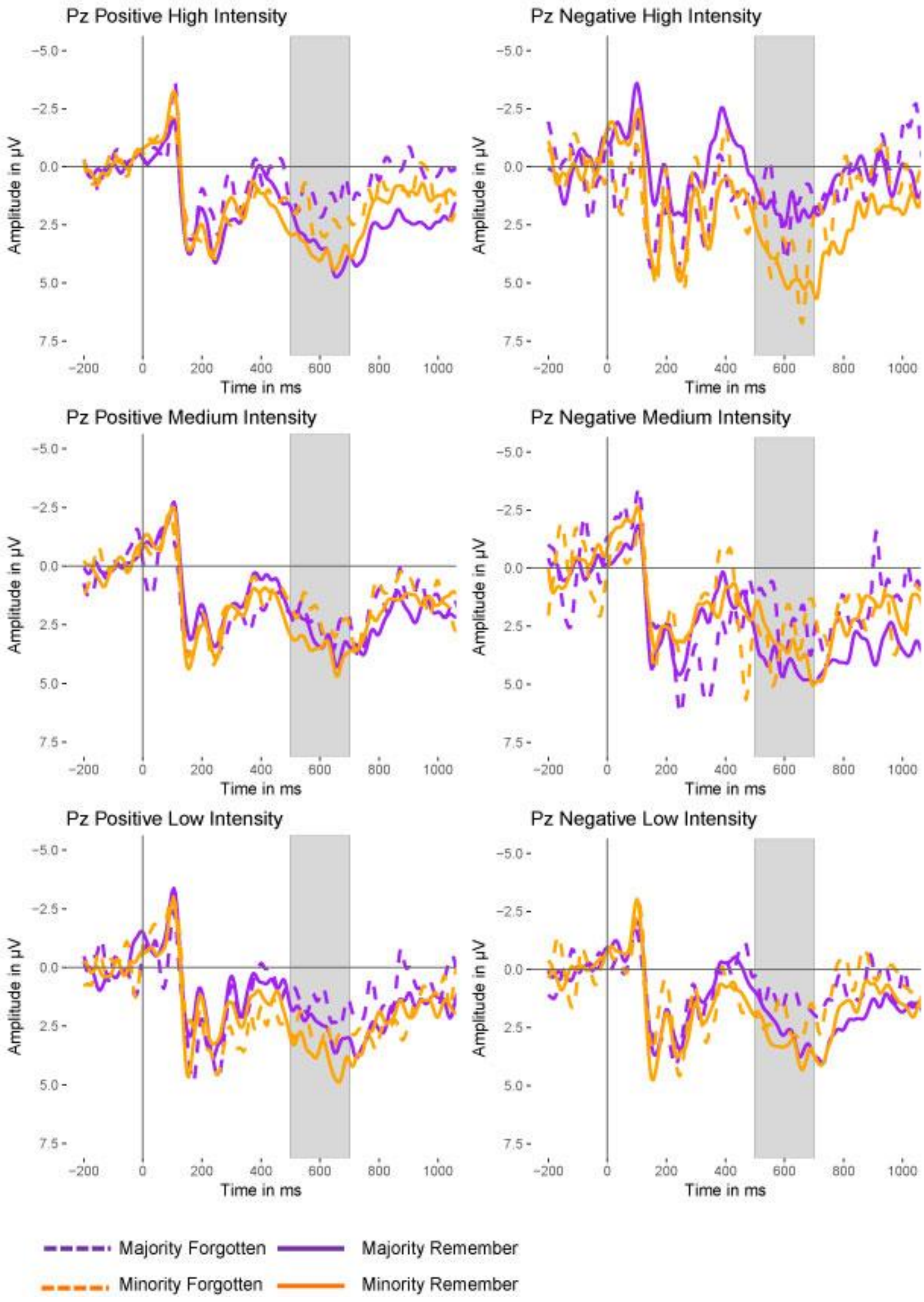

Figure 5.4 P300 at the electrode Pz in the study phase for positive and negative items of high, medium or low intensity. Dashed lines denote forgotten items and solid lines denote remembered items. The grey bar indicates the 500-700 ms time window which was used for statistical analysis.

5.4 Discussion

5.4.1 The Effects of Shared Distinctiveness

The present study used ERPs to investigate the effect of shared distinctiveness on source memory and its neural correlates. Shared distinctiveness at encoding was created by presenting frequent, positive words and infrequent, negative words either in a frequent color or in an infrequent color. We found better source memory for negative minority items than for positive minority items. However, positive items were better remembered than negative items for the majority source. Reaction times were faster to negative majority items and positive minority items than to positive majority items and negative minority items indicating a speed-accuracy trade-off. Interestingly, the cross-over interaction pattern which we observed for the unbiased hit rates, source confidence ratings, and reaction times was also found in the ERPs for highly intense stimuli. The P300 at encoding was larger for positive majority items and negative minority items than for negative majority items and positive minority items. Furthermore, the frequency of negative minority items was estimated to be higher than the frequency of negative majority items indicating the presence of an IC.

In brief, our study revealed a highly consistent pattern of results indicating that shared distinctiveness indeed affects encoding and retrieval. The patterns observed for the minority items are largely consistent with the SDA. However, the pattern of results for the majority – especially for the P300 – is inconsistent with the SDA and our results from Experiment 4. According to the SDA, source memory should have been better for the minority than for the majority and best for the negative minority items. In our case, memory performance was similar for the majority and the minority. Furthermore, we were unable to replicate the findings by Johnson and Mullen (1994) and McConnell et al. (1994) that participants are faster in assigning negative minority items to the correct source. However, these results are in line with our memory results from Experiment 3.

In contrast to studies showing that the P300 is predictive for subsequent memory especially for distinctive items (see Fabiani, 2006, for a review), the present study revealed only a generic P300 SME which was modulated by intensity, but not by shared distinctiveness. This is in line with a study by Kamp et al. (2015) who found that subsequently remembered emotional words were more extreme than subsequently forgotten emotional words. Since shared distinctiveness did not modulate the P300 SME, but affected source memory performance, our results might indicate that the immediate processing of distinctive features might be only indirectly related to better encoding (e.g. Fernández et al., 1998). This would be consistent with evidence indicating that absolute rather than immediate distinctiveness is responsible for superior memory (Hunt, 1995; von Restorff, 1933) and ICs (McConnell et al., 1994).

There is a large agreement that the P300 reflects the updating of mental schemata and that the P300 amplitude is indirectly proportional to subjective probability (e.g. Duncan-Johnson & Donchin, 1977; see also Polich, 2007). Consistent with this view, we found that the P300 for extreme negative minority items was larger than the P300 for extreme positive minority items and largest as compared to all other category combinations. However, contrary to our prediction, the pattern was reversed for intense majority items. This result cannot be attributed to the participants' misperception of the frequencies for majority items. The participants not only correctly identified the preponderance of the majority color, but also the preponderance of positive items in the frequency estimation task. Moreover, the presence of an IC can be taken as evidence that positive items presented in the majority color were perceived as the most frequent category combination. Therefore, it is even more surprising that the P300 amplitude was larger for positive majority items than for negative majority items. There are several potential explanations for this unexpected result.

This relation was found only for the highly intense stimuli indicating that these items might have been perceived as less characteristic for the majority and, consequently, as more distinctive than the more moderate positive items. Indeed, no differences between the four category combinations were found for low or moderate intense items. This might account for the fact that effects on the P300 were only observed for the extreme items. Furthermore, the JOL task at encoding might have prompted participants to pay attention especially to the positive items of the majority. Since a post-hoc analysis indicated that high intensity positive items received the highest JOL as compared to all other categories (all ps < .001), the JOL task might also explain why the increased P300 amplitude was only observed for the extremely positive majority items. An alternative interpretation is that the higher JOLs for the positive items as compared to the negative items reflect the difference in intensity between positive and negative items. In this sense, the JOLs might reflect that the positive items were perceived as more distinctive.

To sum up, the interpretations presented in this section can all account only for some part of the results. An alternative, more integrative perspective on the results is offered by AT which will be discussed in the next section.

5.4.2 Attention Theory: Shared Distinctiveness, Accentuation, and Illusory Correlations

As outlined in section 2.1.2.3.2, AT claims that the two diametrically opposed category combinations, the most frequent and the least frequent in the IC paradigm, receive more attention than the remaining two category combinations for the purpose of category accentuation and differentiation (Sherman et al., 2009). In our study, the positivity of the majority source color should have been learned first and attention should have been shifted to the negative items in the minority source color for differentiation. Consistent with this idea, we found stronger effects for the positive majority items and the negative minority items than for the other category combinations in the unbiased hit rates, source confidence ratings, reaction times, and in the P300 for intense items. Moreover, the presence of an IC provides further evidence that the majority source is associated with positivity, whereas the minority source is associated with negativity.

Thus, our results imply that attention is allocated to diametrical category combinations (Tajfel & Wilkes, 1963) leading to the accentuation of differences between the categories (Sherman et al., 2009). The effects for the P300 could reflect contrast enhancement for the most informative category combinations (i.e. majority positive and minority negative).

Even though AT (Sherman et al., 2009) does not make specific predictions regarding the contribution of episodic memory to IC, it seems plausible that the attention shifts contributed to better encoding. Since we found only a generic

P300 SME in the present task, the P300 might reflect attention shifts or the updating of the mental representation, which indirectly contribute to memory rather than direct effects of successful encoding. This might also account for the fact that we found an IC and superior memory for positive majority and negative minority items, even though they were not correlated.

Murphy et al. (2011) reported that the IC reached a maximum at intermediate trial numbers, but disappeared after extended learning (i.e. 90 trials in Murphy et al., 2011). In contrast, we found an IC even after 360 study trials. Even though it can be argued that we measured the IC in a pre-asymptotic phase, this explanation seems highly unlikely for the following reasons. The IC began to disappear after 48 trials in Murphy et al. (2011), whereas we measured a reliable IC even after more than seven times the number of trials of the asymptote in Murphy et al. (2011). Furthermore, Kutzner et al. (2011) also reported that the IC can be observed even after 300 learning trials. It might rather be the case that increasing transparency of the study task, a factor known to affect the IC (Chapman, 1967) might have been responsible for the disappearance of the IC in Murphy et al. (2011). Thus, the presence of an IC even after extended learning is completely in line with AT or the SDA, but inconsistent with basic learning approaches.

In our study, accentuation was considered only post-hoc as a potential explanation to account for the unexpected interaction pattern found in the behavioral and ERP data, especially for the findings related to the majority. Furthermore, the critical interaction in the ERP data was only significant on a trend level. Therefore, the accentuation interpretation has to be considered preliminary. So far, only a few studies have investigated the IC from an accentuation perspective (e.g. Berndsen et al., 2001; McGarty et al., 1993) and such frameworks fail to account for some findings in the literature (e.g. Van Rooy et al., 2013) and the present thesis (e.g. Experiment 1 and 2). Therefore, more systematic investigations of the effect of accentuation on the IC are necessary.

Future studies should aim at testing the accentuation effect by explicitly focusing on the diagonals of the contingency table. If there is a genuine ERP correlate of the accentuation effect, the diagonal that contains the most and the least frequent category combinations should together elicit a larger P300 than the diagonal that contains the category combinations of intermediate frequency. Experiment 5 provides some preliminary results on this topic. We found that the P300 amplitude linearly increased with decreasing objective probability. This result is in line with common conceptions of the P300 as an index of subjective probability, but contradicts an interpretation based on category accentuation. However, since we failed to obtain an IC in the frequency estimation task, these results have to be treated with caution.

5.4.3 Strengths and Limitations

The current study improved on prior studies on the IC. For this ERP study, we adapted the optimized version of the Hamilton and Gifford (1976) paradigm from Experiment 3. These optimizations especially increase the reliability of the source memory task.

Equating the frequency of the new items allowed us to better discriminate between real performance differences and response biases which arise from the skewed frequency distribution (Table 5.1; see also Experiment 3). Moreover, this procedure enabled us to preclude primacy and recency effects by systematically excluding majority items at the beginning and the end of each study list.

In addition, the positive or negative word itself was the only cue provided for recalling the color at encoding in our study. Most studies on the IC, in contrast, use person descriptions which include the name of the person, his/her group membership, and a description of his/her behavior (e.g. Hamilton & Gifford, 1976). Thus, the group membership could be remembered not only based on the behavioral description, but also on the basis of the individual’s name.

Furthermore, we rigorously matched the stimulus material for factors known to affect memory, namely arousal, concreteness, word length, and word frequency. However, given the limited range of highly positive or negative words provided by Lahl et al. (2009), we had to include less intense words in order to achieve a high trial number and, consequently, a high signal-to-noise ratio. The P300 results clearly indicate that intensity needs to be taken into account in order to obtain unambiguous results.

However, our study has some limitations, especially concerning the distinctiveness manipulation, which need to be considered. First of all, we had too many distinctive items per list. Right now there are 16 items in the most frequent category combination (i.e. positive majority items), but 20 distinctive items in the remaining three category combinations (see Table 1). Thus, there are fewer items in the most frequent category combination than in the combination of all distinctive items. This problem might also have contributed to the unusual finding for the P300. In most studies on distinctiveness and subsequent memory, only a single deviant item per distinctive category is presented in each list (e.g. Fabiani & Donchin, 1995; see also Experiment 5). A control experiment with fewer items per list in the least frequent category would provide more conclusive evidence on the effect of shared distinctiveness.

Second, the experimental design includes several types of distinctiveness due to the structure of the Hamilton and Gifford (1976) IC paradigm (see section 3.1, for an extensive discussion of this point). Primary distinctiveness was deliberately induced by the infrequency of the minority source and negative items. However, negative items are also distinctive due to emotionality and secondary distinctiveness, because negative events are less frequent in real life than positive events (Schmidt, 2012; see also Alves et al., 2017; Baumeister et al., 2001). In addition, participants needed more time to encode negative items of the minority (Stroessner et al., 1992). Lower source memory performance for the negative items overall might have been caused by insufficient encoding time. Using only physical distinctiveness (i.e. size and color) might be more suitable to test the influence of shared distinctiveness on ERPs and memory than the Hamilton and Gifford (1976) paradigm.

Third, the valence manipulation might have been subtler than the color manipulation. This might explain why the P300 effect for valence was only obtained for highly intense items, whereas a main effect was found for color. This is consistent with the fact that participants rated the minority color to be more salient than the majority color in the post-experimental questionnaire (Minority: M = 3.46, SD = 0.85; Majority: M = 3.08, SD = 0.74; t(38) = -1.96, p = .029, one-sided). Due to the necessary high number of trials the selected positive and negative stimuli might have been too heterogeneous. Thus, only the extreme exemplars elicited strong ERP responses. Still another reason for the heterogeneous results for valence might be that arousal rather than valence drives the effect of the emotional stimuli on ERPs. However, arousal was matched for positive and negative stimuli and therefore cannot account for the findings.

Stroessner et al. (1992) reported that positive and negative mood reduces or even erases differential attention allocation at encoding and thereby the IC effect. Our experiment is very long and the 2-back task was reported to be exhausting. Exhausted subjects might have had a negative mood which, as a consequence, reduced the shared distinctiveness effect and the IC. Furthermore, depressive mood has been noted to reduce ICs in the literature (e.g. Alloy & Abramson, 1979). In contrast, people who are optimistic are more prone to cognitive illusions, including the IC (e.g. Taylor & Brown, 1988). Future studies should, therefore, include measures of depression or optimism and more rigorously assess the mood of the participants in different stages of the experiment along with using a less taxing filler task.

5.5 Conclusion

Experiment 6 showed that shared distinctiveness indeed leads to better source memory and ICs, thereby extending previous studies (Hamilton et al., 1985; McConnell et al., 1994; Risen et al., 2007; see also Experiment 3). In contrast to predictions based on the SDA, memory performance was not related to the extent of IC. However, memory was also enhanced for positive items in the frequent color. This pattern was also reflected in the P300 for highly positive and negative items. Our results imply that the processing of distinctiveness might lead to attention allocation to diametrical category combinations (Tajfel & Wilkes, 1963), thereby accentuating the differences between the categories (Sherman et al., 2009). However, shared distinctiveness did not modulate the P300 SME indicating that the processing of distinctive features might be only indirectly related to better encoding (Fernández et al., 1998). AT (Sherman et al., 2009) provided an integrative perspective for the behavioral and ERP results.

6 General Discussion

The aim of the present thesis was to determine the role of episodic memory in the distinctiveness-based IC by using behavioral methods and ERPs. Proponents of the SDA claim that superior memory for shared distinctive group-behavior combinations is responsible for the IC (e.g. Hamilton & Gifford, 1976, McConnell et al., 1994), whereas proponents of the ILA claim that ICs result from regression to the mean which especially affects the representation of the infrequent group (e.g. Fiedler, 1991, 2000). Three behavioral studies (Experiment 1-3) were conducted to determine the relative merit of the SDA and the ILA. Experiment 1 and 2 investigated whether an IC could be observed under conditions with skewed and equated category frequency conditions for valence, respectively. Via a computer simulation with BIAS (Fiedler, 1996, 2000), we identified that a comparison of these conditions would provide a critical test of the ILA. The results from Experiment 1 and 2 clearly rule out the ILA (Fiedler, 2000) as a plausible account for the distinctiveness-based IC, because ICs were observed irrespective of the frequency condition. Experiment 3 introduced several methodological refinements to Experiment 1 and 2 as well as previous studies. It revealed that memory for distinctive category combinations was enhanced as predicted by the SDA. Furthermore, source memory performance predicted the extent of IC. The three behavioral studies, thus, established that 1) ICs are driven by distinctiveness rather than infrequency and 2) there is in fact a memory advantage for shared distinctive items.

The P300 is an ERP component closely linked to the processing of distinctiveness (e.g. Fabiani, 2006). Therefore, we conducted three ERP studies in order to further illuminate the relationship between the P300, (shared) distinctiveness, memory, and IC. According to accentuation approaches, attention should lie on both, the most frequent and the least frequent category combination, in order to maximize the differentiation between the categories (Sherman et al., 2009). Therefore, Experiment 4 was designed to compare the SDA with the accentuation approach to ICs by using an oddball paradigm (Polich, 2007). Consistent with the SDA, we found that the P300 amplitude increased as a function of distinctiveness. In Experiment 5, we applied the subsequent memory paradigm to test whether the N400 and the P300 at encoding predicted recognition based on familiarity and recollection, respectively. The results imply that the P300 predicts recollection-based recognition, whereas the N400 does not predict subsequent memory.

In the final study, Experiment 6, we combined the insights gained from the previous five experiments in order to test whether perceived distinctiveness as indexed by the P300 can be related to subsequent memory and the amount of perceived covariation between two features. The behavioral results of Experiment 6 replicated the results of Experiment 3. Unlike Experiment 3, memory performance did not predict the extent of IC in Experiment 6. For highly intense items, the P300 amplitude was larger for the most frequent and the least frequent category combination – a result which is more in line with the accentuation approach than with the SDA. However, shared distinctiveness did not modulate the P300 SME, indicating that the processing of distinctive features might be only indirectly related to better encoding.

In the following sections, the contribution of all six experiments to the research literature will be critically evaluated. First, we will discuss how the experiments contribute to our understanding of the relationship between episodic memory and the IC (section 6.1). Then, the implications of our experiments for models of the distinctiveness-based IC will be assessed (section 6.2). Finally, a discussion of the strengths and limitations of the present studies as well as possible future directions will be provided (section 6.3).

6.1 The Relationship between Episodic Memory and the Distinctiveness-Based Illusory Correlation

ICs are thought to arise at retrieval (e.g. Hamilton & Gifford, 1976). Therefore, research has rarely focused on whether ICs are formed already during encoding and existing data are ambiguous. Fiedler (1985) reported that the encoding measures of IC were significantly correlated with the extent of IC at retrieval in four out of six experiments and concluded that ICs might already arise during encoding. In contrast to these studies, the Experiments 1-3 in the present thesis as well as Weigl et al. (2015) showed that the actual correlation is accurately perceived during encoding, even though an IC can be observed at test. These results are consistent with evidence indicating that ICs do not arise under conditions that prompt participants to form an impression of the groups during encoding (e.g. Meiser, 2003; Pryor, 1986). Furthermore, Fiedler's (1985) experiments included real-world groups and, therefore, might be compromised by pre-existing expectations. This leads us to the conclusion that distinctiveness-based ICs can be assumed to result primarily from post-encoding processing, even though this does not yet clarify the role of episodic memory in the IC paradigm.

As outlined in section 2.4, the literature on the contribution of (episodic) memory to ICs is equivocal. At least part of the heterogeneous results can be attributed to the fact that most IC studies assess memory with paradigms and measurement techniques which are not optimal from the perspective of memory research (Klauer & Meiser, 2000). In the Experiments 1 and 2, we relied on the source memory task, which is typically used in research on the distinctiveness-based IC (Mullen & Johnson, 1990), and found that memory performance was better for the majority than for the minority and best for positive traits of the majority. A methodologically optimized paradigm, which was applied in Experiment 3 and 6, not only replicated the effect for positive items of the majority, but also revealed a memory advantage for shared distinctiveness as predicted by the SDA. In contrast to most IC studies, we exerted rigorous control for potentially confounding factors like valence, arousal, concreteness, or word length with word norms in our experiments and provided a strict definition of distinctiveness based on the memory research literature (Schmidt, 1991, 2012).

According to Fiedler (2000), ICs are most likely to be obtained under suboptimal encoding conditions like incidental encoding or high memory load. The present experiments found ICs of similar size under incidental (Experiment 3) and intentional encoding (Experiment 1, 2, and 6). Furthermore, the effect of shared distinctiveness on memory can be obtained irrespective of whether encoding was incidental (Experiment 3) or intentional (Experiment 6) as long as primary and recency effects are considered and the impact of response bias is reduced by equating the frequencies of majority and minority items at test.

One potential explanation to accommodate the discrepancy between Fiedler's (2000) position and our experiments might be that studies which compare optimal with suboptimal encoding conditions often also differ in the degree of transparency about the goal of the study (e.g. Lilli & Rehm, 1983, 1984). Chapman (1967) found that the extent of IC was larger in the first block than in the subsequent blocks. An IC study by Pryor (1986) points in a similar direction. Pryor (1986) used the Hamilton and Gifford (1976) paradigm, but gave the participants two different instructions. Participants in the memory condition were instructed to memorize each statement, whereas participants in the impression formation condition were instructed to form an impression of each group (Pryor, 1986). An IC was found in the memory condition, but not in the impression formation condition. The participants in the impression formation condition could re-use their impression formed at encoding and did not show an IC. In contrast, the participants in the memory condition were unaware of the subsequent liking rating and, therefore, had to form an impression at retrieval. Nevertheless, the memory performance was roughly the same in the impression formation condition and the memory condition.

Despite the fact that the shared distinctiveness effect was found in Experiment 3 and 6, we obtained a direct relationship between source memory performance and the IC only in Experiment 3. Memory did not predict the IC in the multiple regression analysis in Experiment 6. Moreover, differential encoding did not contribute to the IC or the memory effect, because the P300 SME in Experiment 6 was not affected by our distinctiveness manipulation. This was surprising, especially with regard to the results of Experiment 5. There, we found a P300 SME only in the conditions with distinctive items as was expected based on the ERP literature (e.g. Fabiani, 2006; Kamp et al., 2015). Although there was a relationship between source memory and IC in Experiment 1 and 2, the interpretation was complicated by the suppression effect and the suboptimal source memory task. Together, this pattern of results cast serious doubt on whether there exists a straightforward relationship between memory and the IC.

One interpretation of this pattern of results could be that the effects in memory are an epiphenomenon rather than a causal agent in ICs. Van Rooy et al. (2013) observed an IC irrespective of the presence or absence of negative items in the least frequent group. Nevertheless, memory was enhanced in the condition with negative items in the least frequent group. In a similar vein, Pryor (1986) reported comparable memory results in an impression formation condition and a memory condition, even though an IC was found only in the memory condition. Thus, ICs might not be directly related to the memory benefit for shared distinctiveness. Rather shared distinctiveness might produce both, the memory benefit and the IC, but via different causal routes.

An alternative interpretation of the results is also conceivable. In contrast to the typical IC experiment, participants had to learn 360 items in Experiment 6. Thus, it might be the case that participants use all available information if they have to deal with a small amount of items. However, this strategy might be inefficient for a large amount of items. Rather than using all items retrieved from memory for their frequency estimation, it seems likely that the participants based their estimates on a small sample of the retrieved items. The memory task in Experiment 6 was an old/new recognition task followed by a source memory task. Therefore, participants might have recognized more items than they were able to actively retrieve during the frequency estimation task. Further evidence for the notion that the IC is generated directly during retrieval (e.g. frequency estimation, group assignment) stems from a study by Ratliff and Nosek (2010), who reported that an IC, which was found in explicit attitude measures (i.e. trait ratings), was absent in implicit measures (i.e. the Implicit Association Test; Greenwald, McGhee, & Schwartz, 1998). From this perspective, the absence of a correlation between source memory and the IC in Experiment 6 is not so surprising.

A critical test to decide between the two interpretations would be to use a free recall task in an IC paradigm with large numbers of items, because free recall better captures what items are available to the participants than the source memory task in Experiment 6. If the epiphenomenon interpretation were true, then the free recall should not correlate with the extent of IC. However, if the sampling interpretation were true, then the amount of recalled items for each category combination should predict the extent of IC.

In conclusion, the experiments in the present thesis provided evidence for a memory benefit for shared distinctive items. Nevertheless, a direct relationship between episodic memory and the IC was found only in a subset of experiments. A closer look at the differences between the experiments led to the thought that the question might not be if participants use episodic memory to guide their subjective covariation assessment. Rather, the right question might be under which circumstances episodic memory is used for covariation assessment. Unless further studies are conducted, these thoughts remain pure speculations.

6.2 Theretical Implications of Experiment 1-6 for Accounts of the Illusory Correlation

In this section, the merit of the SDA, the ILA, and accentuation account or AT as an explanation for the IC is critically evaluated with respect to the results obtained from the IC experiments of the present thesis. At the end of this section, the implications of the present experiments for other accounts of the IC will be discussed.

6.2.1 Implications for the Shared Distinctiveness Account

The SDA was the first account offered to explain the distinctiveness-based IC (Hamilton & Gifford, 1976). The main assumptions of the SDA are 1) that infrequent items are more distinctive than more common ones, 2) that these distinctive items receive additional processing at encoding, which increases their availability at retrieval, and 3) that these items are, as a consequence, more easily retrieved from memory (Hamilton & Gifford, 1976, Tversky & Kahneman, 1973).

Experiment 4 provided a critical test of the first assumption of the SDA. As predicted on the basis of the ERP literature on the P300 (e.g. Donchin, 1981; Polich, 2007) and consistent with the SDA, we found a linear increase of the P300 amplitude as a function of infrequence in Experiment 4. The highest P300 amplitude was elicited by the least frequent category combination indicating that shared distinctiveness indeed leads to more processing than distinctiveness on just one dimension. This result is also consistent with the finding by Stroessner et al. (1992) that shared distinctiveness leads to prolonged encoding times. However, in contrast to the predictions by the SDA, participants did not overestimate the frequency of the least frequent category combination. Instead, they overestimated the frequency of the most frequent category combination.

Some evidence for the second assumption of the SDA was found in Experiment 5 and 6. Participants in Experiment 5 were exposed to semantically and physically isolated items in the midst of regular items at the encoding phase. Even though no behavioral von Restorff effect was obtained, the ERPs provided evidence that the semantically and physically isolated items were not only perceived as distinctive as indexed by the P300 amplitude, but also that the additional processing led to more successful encoding as indexed by the P300 SME. These findings are consistent with the literature on the P300 SME (see Fabiani, 2006, for a review). Experiment 6 combined the subsequent memory paradigm with the IC paradigm. However, in contrast to our predictions based on the SDA, we found only a generic P300 SME, i.e. subsequently remembered items elicited a larger P300 amplitude than subsequently forgotten items. Rather, the P300 amplitude was modulated by shared distinctiveness irrespective of subsequent memory. In contrast to our findings in Experiment 4, the P300 amplitude was larger for both, positive majority items and negative minority items, i.e. the most and the least frequent category combination. Since the effects for shared distinctiveness in the P300 were found only for highly intense stimuli, further studies on the subsequent memory effect for shared distinctiveness are needed to allow firm conclusion concerning the second assumption of the SDA.

The results from our experiments provide only mixed support for the third assumption of the SDA. Even though we predicted enhanced memory for the negative minority items in Experiment 1 and 2, we only found that memory was better for the majority than the minority and best for positive majority items. It seems likely that methodological problems with the source memory task in Experiment 1 and 2 are responsible for the absence of a memory effect for shared distinctiveness, because we found enhanced memory for the least frequent category combination, the negative items of the minority, in

Experiment 3 and 6. Contrary to our predictions based on the SDA, we found that memory for positive majority items, the most frequent category combination, was also enhanced. It is interesting to note that Hamilton and Gifford (1976) found more accurate cued recall for both, the most and the least frequent category combination, even though the result is not extensively discussed.

In short, the results from the present experiments provide only mixed support for the three main assumptions of the SDA. Most convincing evidence was obtained for the first assumption of the SDA, but further studies are needed to test the remaining two assumptions. Shared distinctiveness as a mechanism underlying the IC can account for some results like the enhanced memory for negative minority items (Experiment 3 and 6) or the linear increase of the P300 as a function of infrequency as in Experiment 4, but at the same time fails to explain the unexpected pattern of the P300 in Experiment 3 or the superior memory for the majority.

6.2.2 Implications for the Information Loss Account

The ILA was proposed as a major theoretical alternative to the SDA (Hamilton & Gifford, 1976; Hamilton et al., 1985; McConnell et al., 1994), because it can explain a variety of results found in the literature on the IC and related phenomena (e.g. baseline neglect, accentuation) without the need to postulate biased processing of incoming information (Fiedler, 1991, 1996, 2000). Regression to the mean which results from the aggregation over noisy exemplars is proposed to be the driving force underlying ICs (Fiedler, 1991, 1996, 2000). Nevertheless, the ILA fails to account for most of the results from the experiments of the present thesis.

First of all, the claim by proponents of the ILA that memory is best for the most frequent category combination and worst for the least frequent category combination (Fiedler, 1991; Fiedler et al., 1993) was disconfirmed by Experiment 3 and 6. In these experiments we found that memory was better for the most and the least frequent category combination as compared to the moderately frequent category combinations.

Second, the ILA can account for the presence of an IC after extended learning of category combinations in Experiment 6, but it fails to account for the findings in Experiment 4. Even though the ILA could account for the absence of an IC in Experiment 4 by assuming complete information loss, this explanation seem unlikely for the following reasons. In Experiment 4, participants were presented a total of 360 category combinations with the categories' color and symbol being negatively correlated (φ = -.125), i.e. 63% of the frequent letters were presented in the frequent color, but 75% of the infrequent letters were presented in the frequent color. According to the ILA, participants should be highly accurate in their estimation for the frequent letter, but overestimate the frequency of the infrequent letters presented in the infrequent color due to regression to the mean (Fiedler, 1991, 1996, 2000). However, the reverse pattern was found. Participants overestimated the frequency of the frequent letters presented in the frequent color, but were fairly accurate in their estimation for the least frequent category combination.

Finally, the results of Experiment 1 and 2, which were intentionally designed to test the ILA by presenting skewed and equated category frequency conditions for valence, directly contradict the ILA. More precisely, in contrast to the prediction of the ILA that an IC should be found for skewed, but not for equated frequencies, we found an IC in both conditions. To sum up, the ILA can be ruled out as a viable account for the results of our IC experiments.

6.2.3 Implications for the Accentuation Account and Attention Theory

The accentuation approach is the third major approach to the IC (McGarty et al., 1993). According to this approach, participants search for meaning in the material presented in an IC experiment. The most sensible hypothesis in this scenario would be that one group is better than the other (McGarty et al., 1993). Participants treat the greater absolute difference between the number of positive and negative items for the majority than the minority as a real group difference and enhance this difference by accentuation. A more recent account based on accentuation, AT, claims that participants first learn about the positivity of the majority and their attention then shifts towards negative minority items for the purpose of category accentuation and differentiation (Sherman et al., 2009).

Accentuation can account for the patterns in the source memory task in Experiment 3 and 6. Since positive information about the majority and negative information about the minority would be consistent with the participants' hypothesis (McGarty et al., 1993) and memory for hypothesis-consistent information is superior than for information inconsistent with the participants' hypothesis (Gilboa & Marlatte, 2017; van Kesteren, Ruiter, Fernández, & Henson, 2012; but Stangor & McMillan, 1992), the accentuation account offers an explanation for the superior source memory for positive majority items and negative minority items as compared to the remaining items. Since category members which heighten between-category differences receive more attention at encoding (Krueger, 1991; Krueger & Rothbart, 1990), attention shifts might also promote successful encoding for positive majority items and negative minority items.

In Experiment 6, stronger effects were found for the positive majority items and the negative minority items than for the other category combinations in the P300 for highly intense items. These results can be interpreted as evidence for the attention shift mechanism proposed by AT (Sherman et al., 2009). Moreover, the presence of an IC provides further evidence that the majority source is associated with positivity, whereas the minority source is associated with negativity.

However, accentuation is inconsistent with the linear increase of the P300 as a function of infrequency which we found in Experiment 4, because the two diametrically opposed category combinations, the most and the least frequent, should have received more attention than the remaining two category combinations (Sherman et al., 2009). This might indicate that the likelihood for finding accentuation effects and attention shifts increases if several forms of distinctiveness are implied (e.g. primary and secondary distinctiveness; emotional significance) as in Experiment 3 and 6. Primary distinctiveness alone as in Experiment 5 might not be sufficient for accentuation effects in the ERPs and overt behavior. An alternative explanation is that the simplistic material in Experiment 4 did not prompt the participants to search for a pattern between symbols and colors. Klauer and Meiser (2000), for example, found

ICs for meaningful group-behavior combinations, but not for more arbitrary group-gender combinations. Similar results were obtained by Haslam et al. (1996). Further studies are needed to decide between these hypotheses.

Furthermore, the results from Experiment 2 are inconsistent with the accentuation approach (McGarty et al., 1993) as well as with the AT by Sherman et al. (2009). In the skewed frequency condition, there are 24 items in favor for the majority and only 12 items in favor for the minority (see diagonals in Table 1). Participants accentuate this perceived difference between the groups, i.e. the information on the diagonal favoring the majority is emphasized. However, in the equated frequency condition the number of stimuli on the diagonals is 18 in both cases and, therefore, participants should not have a clear preference for one specific group as suggested by the accentuation approach. AT can only account for the present findings by assuming that positive features are learned before negative features even in the equated frequency condition. Under these circumstances, participants should still shift their attention towards the negative items when learning about the minority.

To sum up, accentuation seems to provide the best explanation for the results in the source memory task in Experiment 3 and 6 and the P300 in Experiment 6. However, both, the accentuation account and AT, fall short in explaining the linear relationship between the P300 and infrequency (Experiment 4) or the results from the equated frequency condition (Experiment 2) without adding yet to be tested assumptions.

6.2.4 Implications for other Accounts of the Illusory Correlation

The present thesis so far focused mainly on the three major accounts of the IC, namely the SDA, the ILA, and the accentuation account. However, the findings from our experiments also inform about the validity of other proposed accounts for ICs.

The mere exposure effect which is a preference for more often encountered stimuli over less often encountered stimuli (Zajonc, 1968; see also Hamilton, 1981) might offer an intriguingly simple alternative explanation as to why an IC can be observed in Experiment 1, 2 3, and 6. In each of these experiments, participants see more items referring to the majority than to the minority and therefore they might judge the majority group more favorably due to higher familiarity. Consistent with the mere-exposure effect, participants in Experiment 5 preferred the frequent non-isolated items over the infrequent isolated items. However, the mere-exposure effect cannot explain the absence of an IC in Experiment 4. Furthermore, the participants in this experiment did not develop a preference for the majority symbol or color (both ps > .20). In addition, the mere-exposure effect does not provide an explanation for the effects we observed in the source memory task. Last but not least, the mere-exposure effect cannot explain the standard finding that ICs, though slightly weakened, are reversed when negative items are more frequent than positive items (Hamilton & Gifford, 1976 Exp. 2; see also Mullen & Johnson, 1990, for a review).

ICs have been considered as a special case of pseudocontingencies (e.g. Fiedler, Freytag, & Meiser, 2009; Fiedler, Kutzner, & Vogel, 2013). Pseudocontingencies arise when a covariation judgment has to be made for two dimensions that have skewed frequency distributions as base rate (e.g. both dimensions have a 3:1 ratio). In contrast to ICs, joint observations of both dimensions are not necessary. Participants in a pseudocontingency paradigm use the information of the base rate to infer a correlation between two dimensions. Although the results of Experiment 1, 3, and 6 can be interpreted as a pseudocontingency, the results of Experiment 2 are clearly incompatible with this view, because the base rate for the dimension valence is 50:50 and no pseudocontingency should arise. The absence of a behavioral IC effect in Experiment 4 is also inconsistent with the pseudocontingency account, because the marginal distributions are the same as in typical IC experiments (2:1) and should therefore be sufficient to elicit a pseudocontingency effect.

Our experiments also have implications for accounts that refute the erroneous or biased character of ICs. For example, Smith (1991) postulated that subjects in an IC experiment rely on the absolute and not on the relative frequency in their judgment. In the original experiment of Hamilton and Gifford (1976), there is a surplus of 10 desirable behaviors for group A (18 desirable behaviors minus 8 undesirable behaviors). For group B, however, there is only a surplus of 5 desirable behaviors (9 desirable behaviors minus 4 undesirable behaviors).

From this perspective, it seems perfectly rational to rate the majority more favorably than the minority in the Experiments 1, 3, and 6. Smith’s (1991) approach can even account for the absence of an IC in Experiment 4. In Experiment 4, the difference between the frequent color and the infrequent color was 60 for both, the frequent and the infrequent letter (frequent letter: 150 – 90 = 60, infrequent letter: 90 – 30 = 60; see also Table 4.1). However, this logic cannot be applied to Experiment 2, because the surplus would be zero for both groups (12 positive traits – 12 negative traits for the majority or 6 positive traits – 6 negative traits for the minority; see Table 3.1). Nevertheless, an IC was observed in this experiment. Furthermore, frequency ratios similar to those in Experiment 4 have been applied in a previous study and led to an IC (Weigl et al., 2015). Since both, Experiment 2 and Weigl et al. (2015), used person descriptions which contained a name, the group membership and a positive or negative trait/behavior, it seems likely that secondary distinctiveness and emotional significance (Schmidt, 2012) played a role in addition to shared distinctiveness. These findings imply that Smith’s (1991) account might provide a fruitful perspective on IC paradigms which only use primary distinctiveness as in Experiment 4. But its logic might not extend to cases with secondary distinctiveness or emotional significance as in Experiment 2 and most other IC based on the paradigm by Hamilton and Gifford (1976).

A variant of this argument would be that participants might not pay attention to the complete contingency table when evaluating the groups, but instead restrict themselves to consider only the positive instances (Fiedler, 1985). Indeed, if one is explicitly asked about the number of positive instances in two classes, it is not at all erroneous to ignore the negative instances. From this perspective, the Experiments 1, 2, 3, and 6 would all provide evidence for the positivity of the majority (i.e. 16 vs. 8 in Exp. 1 and 3, 12 vs. 6 in Exp. 2, and 160 vs. 80 in Experiment 6; see Table 3.1 and Table 6.1). This might explain why participants evaluated the majority more favorable than the minority in the evaluative trait rating, because the scale is largely composed of positive traits. However, the frequency estimation task in the Experiments 1, 2, and 6 required the participant to explicitly estimate the proportion of negative traits in both groups. In this case, participants should evaluate the majority less favorable than the minority, because there are also more negative traits in the majority than in the minority. Moreover, the participants in Experiment 3 had to estimate the absolute frequency of all four category combinations and, therefore, should have not shown an IC at all in the frequency estimation task.

Our results from the source memory task are consistent with Rothbart's (1981) availability account of ICs. Availability depends not only on the actual frequency, but is also influenced by factors unrelated to frequency like distinctiveness, recency, or novelty (Tversky & Kahneman, 1973). But in contrast to Hamilton and Gifford (1976) who focus on distinctiveness, Rothbart (1981) argues that “[i]f we ask which pairs of instances are going to be most available to memory, we can reasonably assert that it would be the most frequent category” (Rothbart, 1981, p. 174). In the typical IC experiment, the positive items of the majority are the most frequent stimuli. Therefore, these items should be most available in memory. Consistent with this idea, we found that source memory was better for the majority than the minority and best for positive traits of the majority in Experiment 1. Our memory results from Experiment 3 and 6 are also in line with Rothbart et al. (1978) who not only found that extreme items are more available than less extreme items (Rothbart et al, 1978, Exp. 2 & 3), but also that positive items were remembered better than negative items, if positive items are more frequent than negative items (Rothbart et al, 1978, Exp. 1). However, Rothbart’s account is somewhat difficult to reconcile with the equated frequency condition in which only superior source memory for the majority, but (due to the equated frequencies) no difference between positive and negative traits would have been expected.

Finally, our experiments have also implications for models relying on associative learning mechanisms, namely the associative learning account by Murphy et al. (2011) and the multi-component model (MCM) by Van Rooy et al. (2013). Murphy et al. (2011) proposed that associative learning models like the Rescorla-Wagner model (Rescorla & Wagner, 1972) can explain the development of ICs. They reasoned that IC was only a transitory phenomenon in the acquisition stage and that the contingency judgment would be quite accurate after extended learning. According to the associative learning account, the IC can be observed prior to sufficient learning, because learning reaches the asymptote faster for the majority than the minority. The results of Murphy et al. (2011; see also Spiers et al., 2016, for a similar experiment) support the predictions. Due to the additional distinctiveness of negative items (e.g. Alves et al., 2015, 2017b), it seems reasonable to assume that learning differs between positive and negative items. In this case, the associative learning account by Murphy et al. (2011) would also predict an IC in both, the skewed and the equated frequency condition (Experiment 1 and 2, respectively). Critically, the IC was maximal between 36 and 54 trials in the experiments by Murphy et al. (2011). Most studies on the IC, including our Experiments 1, 2, and 3, only use between 36 and 48 trials (see Mullen & Johnson, 1990, for a review). At first glance, it seems to be the case that our studies measured the pre-asymptotic state and that no IC would have been observed with a larger number of trials. However, the IC in the study by Murphy et al. (2011) has already vanished after 90 trials, whereas the IC was still present even after 300 trials in the study by Kutzner et al. (2011) as well as in our Experiment 6. Given that Kutzner et al. (2011) also used social groups in their extended learning experiment, it seems unlike that the IC in our Experiment 6 persisted solely due to the use of non-social stimuli. Since Murphy et al. (2011) assessed the attitude towards the majority and the minority at several time points during the acquisition phase, it might be rather the case that rapid decline of the IC reflects increasing transparency of the study purpose (see Lilli & Rehm, 1983, 1984) rather than the asymptotic phase of learning. Another problem of the associative learning account is that it does not give any specific information about the time point when the asymptotic phase is reached. Furthermore, the associative learning account does not make any specific prediction regarding episodic memory. In short, there is a need for more studies which explore the distinctiveness-based IC at higher trial numbers. Such studies could determine whether the IC disappears with extended learning. Comparing the extent of IC between the skewed and the equated frequency condition at the asymptotic stage might provide a critical test for the validity of the associative learning account.

The multi-component model (MCM) is a connectionist model by Van Rooy et al. (2013) which assumes that mental representations are created which are connected to the social group and include a global evaluative impression as well as episodic information. The evaluative connections for the groups grow stronger over the course of learning. But at the same time, memory for discrete episodes is impaired due to competition with stronger evaluative connections. Since this effect is stronger for the majority than the minority, frequency or likability judgments which do not (necessarily) require episodic information are biased in favor for the majority. In contrast, episodic retrieval is better for the minority, especially for negative items. The effect for shared distinctiveness, which we found in Experiment 3 and 6, is consistent with the MCM, because the episodic trace should be strongest for the least frequent category combination. Van Rooy et al. (2013, Exp. 2) reported a similar finding. However, the enhanced source memory for positive majority items we found in Experiment 1, 2, 3, and 6 challenges the basic tenet of the MCM, namely that episodic traces are weakened by increased learning. As with the associative learning account by Murphy et al. (2011), the MCM can only account for the equated frequency condition of Experiment 2 by assuming different learning curves for positive and negative stimuli.

6.2.5 Concluding Remarks

From the discussion in the previous sections it becomes clear that neither theoretical account is supported entirely by our six experiments. All above mentioned accounts provide reasonable explanations for the occurrence of an IC in the standard paradigm as implemented in Experiment 1. However, each account fails to account for portions of the results in the subsequent five experiments without the inclusion of additional assumptions. The ILA and the accentuation account, for example, fail to account for Experiment 2 and 4, whereas the SDA is challenged by the absence of a correlation between memory and the extent of IC in Experiment 6. An implication from our results is that ICs are determined by multiple causes and accounts which rely on a single mechanism are unable to encompass all findings in our experiments and in the research literature (see also Sherman et al., 2009, for a similar discussion). Therefore, the more promising research strategy might be to determine the conditions which prompt participants to choose one mechanism over another instead of focusing on testing each account.

6.3 Strengths and Limitations

In the present thesis, the IC was investigated using different experimental procedures (computer experiments, online questionnaires, ERPs) and materials (traits, behavior descriptions, nouns). Furthermore, Experiment 4 and 6 investigated ICs under extended encoding conditions – an area which is still understudied to date (for notable exceptions, see Kutzner et al., 2011; Murphy et al., 2011). A reliable IC was obtained in all experiments but one, indicating that the IC is a robust phenomenon in line with the conclusions drawn by Mullen and Johnson (1990). Furthermore, applying Schmidt's (2012) classification of extraordinary events to ICs can be considered as a strength of the present thesis. In previous research, the term distinctiveness was rarely explicitly defined and scholars differed in their use of the term. However, Schmidt's (2012) classification allowed us to conceptually disentangle distinctiveness and infrequency in Experiment 1 and 2, thereby providing a critical test of the ILA.

Another methodological contribution of the present thesis is that participants in Experiment 3 and 4 had the possibility to make frequency estimations for all available categories or category combinations. Most IC studies allow participants only to estimate the frequency for negative items (Haslam & McGarty, 1994). As a consequence, it remains unclear whether participants overestimate the frequency of negative minority items, underestimate the frequency of positive minority items, or both. Furthermore, such an assessment can bias the estimation of the extent of IC (Haslam & McGarty, 1994). Allowing participants to estimate the frequency for all category combinations in Experiment 3 led to the novel result that participants not only overestimate the frequency of negative items in the minority, but also underestimate the frequencies of positive and negative items in the majority. The more thorough assessment and analysis of the subjective frequency estimates in Experiment 4 allowed us to identify that participants were highly accurate in estimating the marginal relative frequencies, but failed at estimating the conditional relative frequencies. Even though these results are promising, more research is needed to obtain a more complete understanding of subjective frequency estimation.

In Experiment 6, we used HLM to analyze single trial ERP data, because the traditional repeated-measure ANOVA based on mean amplitude ERPs proved to be ill-suited for designs with unbalanced trial numbers as in IC research (Tibon & Levy, 2015). Research designs for the distinctiveness-based IC necessarily involve skewed frequency distributions and, as a result, unbalanced trial numbers. This problem was further exacerbated, because Experiment 6 was subsequent memory experiments, in which the trial numbers can vary tremendously as a function of memory performance (e.g. Fabiani & Donchin, 1995; Sanquist et al., 1980). Due to the mutual dependence between trial number and memory performance (i.e. the more encoding trials were remembered, the less the number of forgotten trials), it would have been necessary to exclude participants due to insufficient trials in a single condition as a consequence of high memory performance (see Tibon & Levy, 2015 for a discussion). HLM not only allowed us to use the whole sample for analysis, but also to build a statistical model that more accurately captures the actual design structure (i.e. trials were nested within blocks, blocks were nested within participants).

Many ERP researchers discuss methodological problems with the ANOVA (e.g. Luck & Gaspelin, 2017), but rarely promote more modern methodological approaches like HLM or mixed effects models (e.g. Baldwin, 2017; Tibon & Levy, 2015). Experiment 6 contributes to a growing literature which goes beyond the ANOVA/MANOVA approach by applying more sophisticated statistical methods like HLM or mixed effects models (e.g. Davidson & Indefrey, 2007; Rosburg, Mecklinger, & Frings, 2011; Tibon & Levy, 2014). A major benefit of HLM for ERP research is the possibility to address research questions which cannot be adequately addressed with typical general linear model procedures like cross-level interactions (Finch et al., 2014). In Experiment 6, we were able to account for differences in intensity of the emotional words by the use of HLM. Future studies which use the subsequent memory paradigm to investigate the effects of distinctiveness should also consider HLM as a powerful alternative to the traditional GLM approach.

However, the heterogeneous material used across the different experiments is also one of the major caveats of the current thesis, because it can hamper the generalizability of the results and the transfer of the results to typical designs.

For example, conflicting results on the robustness of the IC after extended learning were obtained. While no IC was observed in Experiment 4, a robust IC was found in Experiment 6. An oddball task was used in Experiment 4 in order to maximize the likelihood of obtaining a reliable P300 with a satisfying signal-to-noise ratio (Duncan et al., 2009; Polich, 2007). In the designing of the task, we heavily profited from several decades of research on the P300 and the oddball task (e.g. Polich, 2007), but at the same time had to deviate from the typical IC paradigm established by Hamilton and Gifford (1976). The three main differences are the use of simplistic neutral stimuli instead of positive or negative person descriptions, the use of more than 300 stimuli instead of 36 or 48 stimuli, and the presentation of a negative correlation instead of a zero-correlation. The failure to replicate the IC could be attributed to any of the three differences or a combination thereof. In Experiment 6, we tried to create an experimental design which would be optimal not only for obtaining reliable P300 SMEs, but also for obtaining the distinctiveness-based IC, and were able to replicate the IC in the frequency estimation task and the SME in the P300 time window. Nevertheless, further studies are needed to systematically examine the potential reasons which led to the discrepant results.

Another limitation of our experiments is the absence of a condition without shared distinctiveness. In previous experiments which did not present any negative minority items, participants still reported an IC (Fiedler, 1991; Shavitt et al., 1999; Van Rooy et al., 2013). These experiments challenge the SDA and AT, because neither can account for the finding. The SDA per definition relies on the presence of shared distinctiveness. For AT, attention shifts to the negative items of the minority are critical for accentuation and differentiation. Such attention shifts, however, cannot occur in experiments without negative minority items. The ILA, in contrast, can easily account for ICs in the absence of shared distinctiveness with random noise and regression to the mean. Thus, an explanation which focuses strongly on the least frequent category combination will most likely not be insufficient with respect to the entire research literature. Of course, there are also plausible methodological explanations for the IC despite the absence of shared distinctiveness (e.g. the affordance of the experimental task). Thus, future studies which more systematically investigate conditions without shared distinctiveness would be beneficial for the further theoretical development.

In a similar vein, it has been criticized that most studies approach the IC with a 2 x 2 contingency table in mind (McGarty & de la Haye, 1997; Meiser, 2011). The experiments in the present thesis are no exception. There are only few studies which investigated ICs with other designs. The experimental design by Van Rooy et al. (2013), for example, was based on a 4 x 2 contingency table and Meiser (2003) demonstrated that illusory correlations can even arise when the two variables are correlated with a third variable. This limitation also has implications for theorizing, because the absence of such studies impedes the integration of the IC into broader frameworks of subjective covariation assessment (Meiser, 2011). Although it was helpful to restrict our experimental designs to the 2 x 2 contingency table in order to make our results more comparable with other IC studies, a more complex design might have allowed a more critical test of the assumptions of the SDA. For example, the use of four groups with different group sizes instead of two groups as in the study by Van Rooy et al. (2013) would enable us to test more systematically whether a decreasing group size indeed leads to better memory and a less favorable evaluation. The research literature can so far only provide tentative answers to these questions (e.g. Sanbonmatsu, Sherman, & Hamilton, 1987; Van Rooy et al., 2013).

Moreover, mood or depressiveness was not considered in any of our experiments. However, there is evidence that both can affect the propensity to ICs (e.g. Alloy & Abramson, 1979; Stroessner et al., 1992) or to the illusion of control (Langer, 1975). Taylor and Brown (1988) noted that healthy people exhibit a stronger propensity to cognitive illusion than (mildly) depressed people and this propensity is even increased when the self is implied as in the illusion of control. In these cases, healthy or optimistic people’s judgments systematically deviate from the objectively presented contingencies. Since perceived control is relevant for the self, people might be motivated to overestimate their control over their environment (see also Seligman, 1992). In a similar vein, superstitious behavior (e.g. Skinner, 1992) is most often observed in appetitive paradigms, indicating that the desirability of the outcome influences the perception of contingency (see also Alloy &

Abramson, 1979, Exp. 3). We did not include any manipulations of mood or strongly appetitive material and tested only healthy participants. Therefore, the impact of mood or depressiveness on our data should be low and unsystematic. Nevertheless, future studies should consider mood or depressiveness in order to preclude confounds.

Until now researchers have rarely asked how participants arrive at their covariation estimates (see Berndsen et al., 2001, for a notable exception). Our experiments are no exception. We used the free responses at the end of the experiment only for ensuring the integrity of the data. However, the use of methods like the think-aloud method (a technique, in which participants have to verbalize all their thoughts; e.g. Fonteyn, Kuipers, & Grobe, 1993) might provide insights into covariation assessment which cannot be obtained by other more indirect behavioral measures. A huge benefit of this approach could be that it informs about whether participants discount certain types of evidence in their evaluation (e.g. "He helped the blind man only, because he wanted to impress his girlfriend."). The major drawbacks of this method is, of course, that participants themselves might not be aware of the factors which drive their frequency estimation and that paying attention to the own inferential processes might alter the very nature of these processes. Berndsen et al. (2001) used the think-aloud method, but restricted their analysis to the study phase and the group assignment task, because the search for meaning was the prime interest of their study. Their study does not inform whether and how participants use memory during the frequency estimation and evaluative trait rating.

Finally, a more general limitation in IC research is the absence of a control condition in almost all IC studies. Introducing a control condition would put further constraints on theorizing. In some sense, Experiment 2 of the present thesis could be considered as a control condition, because an equal distribution instead of a skewed distribution for valence was used – thereby precluding primary distinctiveness. However, the best control condition for the distinctiveness-based IC would be an equal distribution for group and behavior. Such a control group would allow determining to what extent positive and negative behaviors, the labels "majority" and "minority" mentioned in the instruction by Hamilton and Gifford (1976), or the participants' inclination to search for meaning in the stimulus material (McGarty et al., 1993) affect their covariation estimate. To the best of our knowledge, equal distributions for both categories have so far been only used in the expectancy-based IC (e.g. Spears et al., 1987).

7 Conclusion and Outlook

Research on the distinctiveness-based IC has mostly been conducted in isolation from research on the cognitive neuroscience of memory. Even the few studies that attempted to bridge the gap mostly focused on pre-existing stereotypes or on person perception rather than on the formation of new intergroup attitudes (see Spiers et al., 2016, for an exception). The present thesis went a step further and translated the core features of the IC paradigm from the social cognition literature into paradigms suitable for ERP methods in order to provide a more integrative perspective on the distinctiveness-based IC.

The results from our experiments clearly rule out the ILA (Fiedler, 2000) as a plausible account for the distinctiveness-based IC. Rather, the experiments partially support the SDA (Hamilton & Gifford, 1976), the accentuation account (McGarty et al., 1993) and other accounts of the IC. The overall pattern of results is mostly in line with AT (Sherman et al., 2009), which encompasses both, accentuation and distinctiveness, in a single framework for the IC. Future studies should investigate boundary conditions for the AT framework and further elucidate the relationship between attention shifts and episodic memory. As some of our results indicate that the effect of episodic memory on ICs depends on the context, a promising avenue for future studies might be to investigate under what circumstances participants rely on memory for their covariation assessment.

Moreover, we were able to show that ICs can persist even after prolonged learning. This implies that people might maintain their stereotypes even after extended exposure to different social groups. On a practical level, our studies add a further mosaic chip to our understanding, why mere intergroup contact is insufficient to promote attitude changes towards outgroups (e.g. Allport, 1979; Pettigrew, 1998).

Our studies further demonstrate that the concepts and methods developed in cognitive psychology and cognitive neuroscience can be fruitfully applied to phenomena in social psychology. So far, this neurocognitive approach to social reality has been restricted to the distinctiveness-based IC. Future studies could expand this approach to the expectancy-based IC or the IC based on the positive-negative asymmetry. Another venue could be to complement our electrophysiological data with data obtained by functional neuroimaging. Such studies could help exploring the brain structures involved in the subjective assessment of covariations. First attempts in this direction have been made (Spiers et al., 2016). Other methodological approaches could include eye-tracking. This would allow obtaining more information on how much attention participants pay to certain items.

In short, the present thesis blended the neuroscience of learning and memory with the literature on social cognition and provided new insights in the neural underpinnings of stereotype acquisition. However, more research is needed in order to obtain a more integrative view of the influence of episodic memory on human covariation assessment.

Appendix A. Unbiased Hit Rates

Wagner (1993) proposed the unbiased hit rates as a measure of accuracy that corrects for response biases. The calculation of the unbiased hit rates depends on the number of sources in the experiment. We had a 2 x 2 design in Experiment 1 and 2 (Table A.1) and a 3 x 3 design in Experiment 3 and 6 (Table A.2). The formulas for the unbiased hit rates in a 2 x2 design and in a 3 x 3 design are given in Table A.3.

Table A.1 A 2 x 2 matrix. Each entry represents the absolute frequency for this cell.

	Judgment	
Stimulus	1 (Majority)	2 (Minority)
1 (Majority)	a	b
2 (Minority)	c	d

More generally, unbiased hit rates are calculated by multiplying the conditional probability of correctly classifying a stimulus given that it is present with the conditional probability of correctly applying a judgment category given that it is applied. The resulting values can range from 0 to 1 and can thus be interpreted like normal hit rates. The resulting values are then arcsine transformed for statistical analysis (Wagner, 1993).

Table A.2 A 3 x 3 matrix. Each entry represents the absolute frequency for this cell.

		Judgment	
Stimulus	1 (Majority)	2 (Minority)	3 (New)
1 (Majority)	a	b	c
2 (Minority)	d	e	f
3 (New)	g	h	i

Table A.3 Formulas for calculating the unbiased hit rates for 2 x 2 matrices (Experiment 1 and 2) and 3 x 3 matrices (Experiment 3 and 6). The letters in the formulas refer to the cells in the corresponding matrix.

Matrix	Majority	Minority	New
2 x 2	$H_{U1} = \frac{a}{a+b} \cdot \frac{a}{a+d}$	$H_{U2} = \frac{d}{c+d} \cdot \frac{d}{b+d}$	–
3 x 3	$H_{U1} = \frac{a}{a+b+c} \cdot \frac{a}{a+d+g}$	$H_{U2} = \frac{e}{d+e+f} \cdot \frac{e}{b+e+h}$	$H_{U3} = \frac{i}{g+h+i} \cdot \frac{i}{c+f+i}$

Appendix B. Stimulus Material for Experiment 1 and 2

A set of 256 adjectives was drawn from the Berlin Affective Word List Reloaded (BAWL-R; Võ et al., 2009). These items were rated by three raters (including the first author) on a 5-point scale whether they are applicable to a person and whether they denote a state or trait. From this set 24 positive and 24 negative adjectives describing traits were selected. The set of 48 adjectives served as item pool for Experiment 1 and Experiment 2 (see Table B.1 for statistical information and B.2 for the German trait words).

Table B.1 Descriptive and inferential statistics concerning the matching of the stimuli. Values for valence, arousal, imageability, and word length are taken from the BAWL-R. Values for the log-transformed lemma frequency are taken from the dlexDB online database (www.dlexdb.de).

Experiment 1	Positive	Negative	Statistics
Valence	1.89 (.34)	-1.84 (.37)	*t*(34) = 30.23, *p* < .001
Arousal	2.63 (.55)	2.77 (.48)	*t*(34) = -0.74, *p* = .465
Imageability	3.09 (.67)	3.01 (.39)	*t*(32.91) = 0.48, *p* = .637
Word length	6.21 (1.50)	6.75 (1.22)	*t*(26.76) = -1.16, *p* = .255
Lemma frequency	34.97 (80.53)	10.79 (19.47)	*t*(34) = 2.07, *p* = .046
Experiment 2	Positive	Negative	Statistics
Valence	1.89 (.35)	-1.81 (.35)	*t*(34) = 31.51, *p* < .001
Arousal	2.84 (.48)	2.91 (.62)	*t*(34) = -0.41, *p* = .688
Imageability	2.96 (.54)	3.20 (.66)	*t*(34) = -1.16, *p* = .253
Word length	6.61 (1.29)	6.89 (1.37)	*t*(34) = -0.63, *p* = .535
Lemma frequency	32.72 (90.49)	8.00 (16.22)	*t*(34) = 2.37, *p* = .024

Table B.2 German trait words used in Experiment 1 and 2.

	Positive traits	Negative traits
Experiment 1	aktiv	asozial
	beliebt	dekadent
	brillant	herzlos
	ehrlich	korrupt
	flexibel	labil
	human	peinlich
	kreativ	planlos
	lieb	primitiv
	loyal	schwach
	mutig	stur
	nett	treulos
	pfiffig	unfair
	reizvoll	
	sanft	
	schlau	
	sinnlich	
	spontan	
	stark	
	stilvoll	
	taktvoll	
	tolerant	
	vital	
	warm	
	weise	
Experiment 2	aktiv	asozial
	beliebt	dekadent
	brillant	gierig
	ehrlich	herzlos
	flexibel	humorlos
	kreativ	labil
	loyal	launisch
	mutig	mies
	pfiffig	militant
	reizvoll	neidisch
	schlau	peinlich
	sinnlich	planlos
	spontan	primitiv
	stark	schwach
	stilvoll	stur
	tolerant	treulos
	vital	unfair
	weise	weltfern

Appendix C. Stimulus Material for Experiment 6

Table C.1 Mean, standard deviation (in brackets), and inferential statistics for the positive and negative words. Please note that only the log-transformed word frequencies were subjected to statistical analysis.

	Positive (N = 320)	Negative (N = 160)	Comparison
Valence	7.54 (0.53)	2.90 (0.62)	F(1, 478) = 7298.08, p < .001
Intensity	2.54 (0.53)	2.10 (0.62)	F(1, 478) = 65.52, p < .001
Arousal	4.67 (1.31)	4.73 (1.21)	F(1, 478) = 0.25, p = .618
Concreteness	6.49 (1.81)	6.38 (1.55)	F(1, 478) = 0.43, p = .510
Word length	7.49 (2.86)	7.33 (2.54)	F(1, 478) = 0.39, p = .535
Word frequency	5701 (9180)	3141 (4252)	-
Log(Word frequency)	7.66 (1.59)	7.56 (0.92)	F(1, 478) = 0.62, p = .431

References

Aggleton, J. P., & Brown, M. W. (1999). Episodic memory, amnesia, and the hippocampal–anterior thalamic axis. *Behavioral and Brain Sciences*, *22*(3), 425–444.

Alloy, L. B., & Abramson, L. Y. (1979). Judgment of contingency in depressed and nondepressed students: Sadder but wiser? *Journal of Experimental Psychology: General*, *108*(4), 441–485. https://doi.org/10.1037/0096-3445.108.4.441

Allport, G. W. (1979). *The nature of prejudice* (Unabridged, 25th anniversary ed). Reading, Mass: Addison-Wesley Pub. Co.

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*(Supplement C), 103–117. https://doi.org/10.1016/j.jesp.2015.10.011

Alves, H., Koch, A., & Unkelbach, C. (2017). Why Good Is More Alike Than Bad: Processing Implications. *Trends in Cognitive Sciences*, *21*(2), 69–79. https://doi.org/10.1016/j.tics.2016.12.006

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*(6), 896–909. https://doi.org/10.3758/s13421-015-0515-5

Bader, R., Mecklinger, A., Hoppstädter, M., & Meyer, P. (2010). Recognition memory for one-trial-unitized word pairs: Evidence from event-related potentials. *NeuroImage*, *50*(2), 772–781. https://doi.org/10.1016/j.neuroimage.2009.12.100

Baldwin, S. A. (2017). Improving the rigor of psychophysiology research. *International Journal of Psychophysiology*, *111*, 5–16. https://doi.org/10.1016/j.ijpsycho.2016.04.006

Bartholow, B. D., Fabiani, M., Gratton, G., & Bettencourt, B. A. (2001). A Psychophysiological Examination of Cognitive Processing of and Affective Responses to Social Expectancy Violations. *Psychological Science*, *12*(3), 197–204. https://doi.org/10.1111/1467-9280.00336

Bartholow, B. D., Pearson, M. A., Gratton, G., & Fabiani, M. (2003). Effects of alcohol on person perception: A social cognitive neuroscience approach. *Journal of Personality and Social Psychology*, *85*(4), 627–638. https://doi.org/10.1037/0022-3514.85.4.627

Battig, W. F., & Montague, W. E. (1969). Category norms of verbal items in 56 categories A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology*, *80*(3p2), 1. https://doi.org/10.1037/h0027577

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323–370. https://doi.org/10.1037/1089-2680.5.4.323

Bell, R., Buchner, A., Erdfelder, E., Giang, T., Schain, C., & Riether, N. (2012). How specific is source memory for faces of cheaters? Evidence for categorical emotional tagging. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *38*(2), 457–472. https://doi.org/10.1037/a0026017

Berndsen, M., McGarty, C., Van der Pligt, J., & Spears, R. (2001). Meaning-seeking in the illusory correlation paradigm: The active role of participants in the categorization process. *British Journal of Social Psychology*, *40*(2), 209–233. https://doi.org/10.1348/014466601164821

Berndsen, M., Spears, R., Van der Pligt, J., & McGarty, C. (1999). Determinants of intergroup differentiation in the illusory correlation task. *British Journal of Psychology*, *90*(2), 201–220. https://doi.org/10.1348/000712699161350

Bless, H., Fiedler, K., & Strack, F. (2004). *Social cognition: how individuals construct social reality*. Hove, East Sussex, UK ; New York: Psychology Press.

Bradley, M. M., Greenwald, M. K., Petry, M. C., & Lang, P. J. (1992). Remembering pictures: Pleasure and arousal in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(2), 379–390. https://doi.org/10.1037/0278-7393.18.2.379

Bruce, D., & Gaines, M. T. (1976). Tests of an organizational hypothesis of isolation effects in free recall. *Journal of Verbal Learning and Verbal Behavior*, *15*(1), 59–72. https://doi.org/10.1016/S0022-5371(76)90007-4

Budd, T. W., Barry, R. J., Gordon, E., Rennie, C., & Michie, P. T. (1998). Decrement of the N1 auditory event-related potential with stimulus repetition: habituation vs. refractoriness. *International Journal of Psychophysiology*, *31*(1), 51–68. https://doi.org/10.1016/S0167-8760(98)00040-3

Bulli, F., & Primi, C. (2006). Illusory correlation and cognitive processes: a multinomial model of source-monitoring. *Review of Psychology*, *13*(2), 95–102.

Cahill, L., Prins, B., Weber, M., & McGaugh, J. L. (1994). β-Adrenergic activation and memory for emotional events. *Nature*, *371*(6499), 702–704. https://doi.org/10.1038/371702a0

Chapman, L. J. (1967). Illusory correlation in observational report. *Journal of Verbal Learning and Verbal Behavior,* *6*(1), 151–155. https://doi.org/10.1016/S0022-5371(67)80066-5

Chapman, L. J., & Chapman, J. P. (1967). Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology, 72*(3), 193–204. https://doi.org/10.1037/h0024670

Cohen, N., Pell, L., Edelson, M. G., Ben-Yakov, A., Pine, A., & Dudai, Y. (2015). Peri-encoding predictors of memory encoding and consolidation. *Neuroscience & Biobehavioral Reviews*, *50*, 128–142. https://doi.org/10.1016/j.neubiorev.2014.11.002

Davidson, D. J., & Indefrey, P. (2007). An inverse relation between event-related and time–frequency violation responses in sentence processing. *Brain Research, 1158*, 81–92. https://doi.org/10.1016/j.brainres. 2007.04.082

de Gelder, B., Böcker, K. B. E., Tuomainen, J., Hensen, M., & Vroomen, J. (1999). The combined perception of emotion from voice and face: early interaction revealed by human electric brain responses. *Neuroscience Letters*, *260*(2), 133–136. https://doi.org/10.1016/S0304-3940(98) 00963-X de Jong, P. J., & Merckelbach, H. (2000). Phobia-relevant illusory correlations: The role of phobic responsivity. *Journal of Abnormal Psychology*, *109*(4), 597–601. https://doi.org/10.1037/0021-843X.109.4.597

Doise, W., Deschamps, J.-C., & Meyer, G. (1978). The accentuation of intra-category similarities. In H. Tajfel (Ed.), *Differentiation between social groups: studies in the social psychology of intergroup relations* (pp. 159–168). London ; New York: Published in cooperation with European Association of Experimental Social Psychology by Academic Press.

Dolcos, F., & Cabeza, R. (2002). Event-related potentials of emotional memory: Encoding pleasant, unpleasant, and neutral pictures. *Cognitive, Affective, & Behavioral Neuroscience*, *2*(3), 252–263. https://doi.org/10.3758/CABN.2.3.252

Donchin, E. (1981). Surprise!… Surprise? *Psychophysiology*, *18*(5), 493–513. https://doi.org/10.1111/j.1469-8986.1981.tb01815.x

Duncan, C. C., Barry, R. J., Connolly, J. F., Fischer, C., Michie, P. T., Näätänen, R., … Van Petten, C. (2009). Event-related potentials in clinical research: Guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. *Clinical Neurophysiology*, *120*(11), 1883–1908. https://doi.org/10.1016/j.clinph.2009.07.045

Duncan-Johnson, C. C., & Donchin, E. (1977). On Quantifying Surprise: The Variation of Event-Related Potentials With Subjective Probability. *Psychophysiology*, *14*(5), 456–467. https://doi.org/10.1111/j.1469-8986.1977.tb01312.x

Dunlosky, J., Hunt, R. R., & Clark, E. (2000). Is perceptual salience needed in explanations of the isolation effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(3), 649–657. https://doi.org/10.1037/0278-7393.26.3.649

Ebbinghaus, H. (1966). *Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie*. Amsterdam: Bonset.

Ehrenberg, K., Cataldegirmen, H., & Klauer, K. C. (2001). Valenz und Geschlechtstypikalität von 330 Verhaltensbeschreibungen - Eine Normierung für studentische Stichproben. *Zeitschrift Für Sozialpsychologie*, *32*(1), 13–28. https://doi.org/10.1024//0044-3514.32.1.13

Elger, C. E., Grunwald, T., Lehnertz, K., Kutas, M., Helmstaedter, C., Brockhaus, A., … Heinze, H. J. (1997). Human temporal lobe potentials in verbal learning and memory processes. *Neuropsychologia*, *35*(5), 657–667. https://doi.org/10.1016/S0028-3932(96)00110-8

Fabiani, M. (2006). Multiple electrophysiological indices of distinctiveness. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and memory* (pp. 339–360). Oxford ; New York: Oxford University Press.

Fabiani, M., & Donchin, E. (1995). Encoding processes and memory organization: A model of the von Restorff effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(1), 224–240. https://doi.org/10.1037/0278-7393.21.1.224

Fabiani, M., Karis, D., & Donchin, E. (1986). P300 and Recall in an Incidental Memory Paradigm. *Psychophysiology*, *23*(3), 298–308. https://doi.org/10.1111/j.1469-8986.1986.tb00636.x

Fabiani, M., Karis, D., & Donchin, E. (1990). Effects of mnemonic strategy manipulation in a Von Restorff paradigm. *Electroencephalography and Clinical Neurophysiology*, *75*(1–2), 22–35. https://doi.org/10.1016/0013-4694(90)90149-E

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

Feldman, J. M., Camburn, A., & Gatti, G. M. (1986). Shared distinctiveness as a source of illusory correlation in performance appraisal. *Organizational Behavior and Human Decision Processes*, *37*(1), 34–59. https://doi.org/10.1016/0749-5978(86)90043-9

Fernández, G., Effern, A., Grunwald, T., Pezer, N., Lehnertz, K., Dümpelmann, M., … Elger, C. E. (1999). Real-Time Tracking of Memory Formation in the Human Rhinal Cortex and Hippocampus. *Science*, *285*(5433), 1582–1585. https://doi.org/10.1126/science.285.5433.1582

Fernández, G., Weyerts, H., Tendolkar, I., Smid, H. G. o. m., Scholz, M., & Heinze, H.-J. (1998). Event-related potentials of verbal encoding into episodic memory: Dissociation between the effects of subsequent memory performance and distinctiveness. *Psychophysiology*, *35*(06), 709–720. https://doi.org/null

Fiedler, K. (1985). *Kognitive Strukturierung der sozialen Umwelt: Untersuchungen zur Wahrnehmung kontingenter Ereignisse*. Göttingen: Hogrefe.

Fiedler, K. (1991). The tricky nature of skewed frequency tables: An information loss account of distinctiveness-based illusory correlations. *Journal of Personality and Social Psychology*, *60*(1), 24–36. https://doi.org/10.1037/0022-3514.60.1.24

Fiedler, K. (1996). Explaining and simulating judgment biases as an aggregation phenomenon in probabilistic, multiple-cue environments. *Psychological Review*, *103*(1), 193–214. https://doi.org/10.1037/0033-295X.103.1.193

Fiedler, K. (2000). Illusory correlations: A simple associative algorithm provides a convergent account of seemingly divergent paradigms. *Review of General Psychology*, *4*(1), 25–58. https://doi.org/10.1037/1089-2680.4.1.25

Fiedler, K., & Armbruster, T. (1994). Two halfs may be more than one whole: Category-split effects on frequency illusions. *Journal of Personality and Social Psychology*, *66*(4), 633–645. https://doi.org/10.1037/0022-3514.66.4.633

Fiedler, K., Freytag, P., & Meiser, T. (2009). Pseudocontingencies: An integrative account of an intriguing cognitive illusion. *Psychological Review*, *116*(1), 187–206. https://doi.org/10.1037/a0014480

Fiedler, K., Kutzner, F., & Vogel, T. (2013). Pseudocontingencies: Logically Unwarranted but Smart Inferences. *Current Directions in Psychological Science*, *22*(4), 324–329. https://doi.org/10.1177/0963721413480171

Fiedler, K., Russer, S., & Gramm, K. (1993). Illusory Correlations and Memory Performance. *Journal of Experimental Social Psychology*, *29*(2), 111–136. https://doi.org/10.1006/jesp.1993.1006

Fiedler, K., & Walther, E. (2004). *Stereotyping as inductive hypothesis testing*. Hove (UK) ; New York: Psychology Press.

Field, A. P., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. London ; Thousand Oaks, Calif: Sage.

Finch, W. H., Bolin, J. E., & Kelley, K. (2014). *Multilevel modeling using R*. Boca Raton, FL: CRC Press, Taylor & Francis Group.

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

Fiske, S. T., & Taylor, S. E. (2013). *Social Cognition: from brains to culture* (2nd edition). Los Angeles: SAGE.

Fonteyn, M. E., Kuipers, B., & Grobe, S. J. (1993). A Description of Think Aloud Method and Protocol Analysis. *Qualitative Health Research*, *3*(4), 430–441. https://doi.org/10.1177/104973239300300403

Friedman, D., & Johnson, R. (2000). Event-related potential (ERP) studies of memory encoding and retrieval: A selective review. *Microscopy Research and Technique*, *51*(1), 6–28. https://doi.org/10.1002/1097-0029(20001001)51:1<6::AID-JEMT2>3.0.CO;2-R

Friedman, D., & Trott, C. (2000). An event-related potential study of encoding in young and older adults. *Neuropsychologia*, *38*(5), 542–557. https://doi.org/10.1016/S0028-3932(99)00122-0

Garcia-Marques, L., & Hamilton, D. L. (1996). Resolving the apparent discrepancy between the incongruency effect and the expectancy-based illusory correlation effect: The TRAP model. *Journal of Personality* *and Social Psychology*, *71*(5), 845–860. https://doi.org/10.1037/0022-3514.71.5.845

García-Larrea, L., & Cézanne-Bert, G. (1998). P3, Positive slow wave and working memory load: a study on the functional correlates of slow wave activity. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *108*(3), 260–273. https://doi.org/10.1016/S0168-5597(97)00085-3

Gardiner, J. M., Ramponi, C., & Richardson-Klavehn, A. (1998). Experiences of Remembering, Knowing, and Guessing. *Consciousness and Cognition*, *7*(1), 1–26. https://doi.org/10.1006/ccog.1997.0321

Geraci, L., & Manzano, I. (2010). Distinctive items are salient during encoding: Delayed judgements of learning predict the isolation effect. *The Quarterly Journal of Experimental Psychology*, *63*(1), 50–64. https://doi.org/10.1080/17470210902790161

Gilboa, A., & Marlatte, H. (2017). Neurobiology of Schemas and Schema-Mediated Memory. *Trends in Cognitive Sciences*, *21*(8), 618–631. https://doi.org/10.1016/j.tics.2017.04.013

Green, R. T. (1958). Surprise, isolation and structural change as factors affecting recall of a temporal series. *British Journal of Psychology; London, Etc.*, *49*(1). Retrieved from https://search.proquest.com/docview/1293574324/citation/9CA1C52F4BAE4904PQ/1

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480. https://doi.org/10.1037/0022-3514.74.6.1464

Griffin, T. R. (2005). *The effect of fearful stimuli on attention and memory.* Unpublished master's thesis, Middle Tennessee State University.

Grillon, M.-L., Johnson, M. K., Krebs, M.-O., & Huron, C. (2008). Comparing effects of perceptual and reflective repetition on subjective experience during later recognition memory. *Consciousness and Cognition*, *17*(3), 753–764. https://doi.org/10.1016/j.concog.2007.09.004

Hajcak, G., MacNamara, A., & Olvet, D. M. (2010). Event-Related Potentials, Emotion, and Emotion Regulation: An Integrative Review. *Developmental Neuropsychology*, *35*(2), 129–155. https://doi.org/10.1080/87565640903526504

Hamilton, D. L. (1981). Illusory Correlation as a Basis for Stereotyping. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 115–144). Hillsdale, N.J.: L. Erlbaum Associates.

Hamilton, D. L., Dugan, P. M., & Trolier, T. K. (1985). The formation of stereotypic beliefs: Further evidence for distinctiveness-based illusory correlations. *Journal of Personality and Social Psychology*, *48*(1), 5–17. https://doi.org/10.1037/0022-3514.48.1.5

Hamilton, D. L., & Gifford, R. K. (1976). Illusory correlation in interpersonal perception: A cognitive basis of stereotypic judgments. *Journal of Experimental Social Psychology*, *12*(4), 392–407.

Hamilton, D. L., & Rose, T. L. (1980). Illusory correlation and the maintenance of stereotypic beliefs. *Journal of Personality and Social Psychology*, *39*(5), 832–845. https://doi.org/10.1037/0022-3514.39.5.832

Haslam, S. A., & McGarty, C. (1994). Problems with the measurement of illusory correlation. *European Journal of Social Psychology*, *24*(5), 611–621. https://doi.org/10.1002/ejsp.2420240507

Haslam, S. A., McGarty, C., & Brown, P. M. (1996). The Search for Differentiated Meaning is a Precursor to Illusory Correlation. *Personality and Social Psychology Bulletin*, *22*(6), 611–619. https://doi.org/10.1177/0146167296226006

Heister, J., Würzner, K.-M., Bubenzer, J., Pohl, E., Hanneforth, T., Geyken, A., & Kliegl, R. (2011). dlexDB – eine lexikalische Datenbank für die psychologische und linguistische Forschung. *Psychologische Rundschau*, *62*(1), 10–20. https://doi.org/10.1026/0033-3042/a000029

Hunt, R. R., & McDaniel, M. A. (1993). The Enigma of Organization and Distinctiveness. *Journal of Memory and Language*, *32*(4), 421–445. https://doi.org/10.1006/jmla.1993.1023

Hunt, R. R. (1995). The subtlety of distinctiveness: What von Restorff really did. *Psychonomic Bulletin & Review*, *2*(1), 105–112. https://doi.org/10.3758/BF03214414

Hunt, R. R. (2009). Does salience facilitate longer-term retention? *Memory*, *17*(1), 49–53. https://doi.org/10.1080/09658210802524257

Hunt, R. R., & Elliot, J. M. (1980). The role of nonsemantic information in memory: Orthographic distinctiveness effects on retention. *Journal of Experimental Psychology: General*, *109*(1), 49–74. https://doi.org/10.1037/0096-3445.109.1.49

Hunt, R. R., & Mitchell, D. B. (1982). Independent effects of semantic and nonsemantic distinctiveness. *Journal of Experimental Psychology:*

Learning, Memory, and Cognition, *8*(1), 81–87. https://doi.org/10.1037/0278-7393.8.1.81

Hunt, R. R., & Smith, R. E. (1996). Accessing the particular from the general: The power of distinctiveness in the context of organization. *Memory & Cognition*, *24*(2), 217–225. https://doi.org/10.3758/BF03200882

Hunt, R. R., & Worthen, J. B. (Eds.). (2006). *Distinctiveness and memory*. Oxford ; New York: Oxford University Press.

Hurlemann, R., Hawellek, B., Matusch, A., Kolsch, H., Wollersen, H., Madea, B., … Dolan, R. J. (2005). Noradrenergic Modulation of Emotion-Induced Forgetting and Remembering. *Journal of Neuroscience*, *25*(27), 6343–6349. https://doi.org/10.1523/JNEUROSCI.0228-05.2005

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*(5), 513–541. https://doi.org/10.1016/0749-596X(91)90025-F

James, W. (1902). *The varieties of religious experience: A study in human nature*. New York, London, Toronto: Longmans, Green and Co.

Jenkins, H. M., & Ward, W. C. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs: General and Applied*, *79*(1), 1–17. https://doi.org/10.1037/h0093874

Jenkins, W. O., & Postman, L. (1948). Isolation and “Spread of Effect” in Serial Learning. *The American Journal of Psychology*, *61*(2), 214–221. https://doi.org/10.2307/1416967

Johnson, C., & Mullen, B. (1994). Evidence for the Accessibility of Paired Distinctiveness in Distinctiveness-Based Illusory Correlation in

Stereotyping. *Personality and Social Psychology Bulletin*, *20*(1), 65–70. https://doi.org/10.1177/0146167294201006

Kamiya, S. (1997). Emotion as a factor in the von Restorff phenomenon. *Japanese Journal of Research on Emotions*, *5*(1), 24–35. https://doi.org/10.4092/jsre.5.24

Kamp, S.-M., Bader, R., & Mecklinger, A. (2017). ERP Subsequent Memory Effects Differ between Inter-Item and Unitization Encoding Tasks. *Frontiers in Human Neuroscience*, *11*. https://doi.org/10.3389/fnhum.2017.00030

Kamp, S.-M., Potts, G. F., & Donchin, E. (2015). On the roles of distinctiveness and semantic expectancies in episodic encoding of emotional words. *Psychophysiology*, *52*(12), 1599–1609. https://doi.org/10.1111/psyp.12537

Kanouse, D. E. (1984). Explaining Negativity Biases in Evaluation and Choice Behavior: Theory and Research. In T. C. Kinnear (Ed.), *Advances in Consumer Research* (Vol. 11, pp. 703–708). Provo, UT: Association for Consumer Research.

Karis, D., Fabiani, M., & Donchin, E. (1984). “P300” and memory: Individual differences in the von Restorff effect. *Cognitive Psychology*, *16*(2), 177–216. https://doi.org/10.1016/0010-0285(84)90007-0

Karpicke, J. D., & Roediger, H. L. (2008). The Critical Importance of Retrieval for Learning. *Science*, *319*(5865), 966–968. https://doi.org/10.1126/science.1152408

Kirsch, I., Lynn, S. J., Vigorito, M., & Miller, R. R. (2004). The role of cognition in classical and operant conditioning. *Journal of Clinical Psychology, 60*(4), 369–392. https://doi.org/10.1002/jclp.10251

Klauer, K. C., & Meiser, T. (2000). A Source-Monitoring Analysis of Illusory Correlations. *Personality and Social Psychology Bulletin*, *26*(9), 1074–1093. https://doi.org/10.1177/01461672002611005

Kleinsmith, L. J., Kaplan, S., & Trate, R. D. (1963). The relationship of arousal to short- and long-term verbal recall. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *17*(4), 393–397. https://doi.org/10.1037/h0083278

Kline, S., & Groninger, L. D. (1991). The imagery bizarreness effect as a function of sentence complexity and presentation time. *Bulletin of the Psychonomic Society*, *29*(1), 25–27. https://doi.org/10.3758/BF03334758

Knight, R. T. (1984). Decreased response to novel stimuli after prefrontal lesions in man. *Electroencephalography and Clinical Neurophysiology*, *59*(1), 9–20.

Knight, Robert T. (1996). Contribution of human hippocampal region to novelty detection. *Nature*, *383*(6597), 256–259. https://doi.org/10.1038/383256a0

Koch, A., Alves, H., Krüger, T., & Unkelbach, C. (2016). A general valence asymmetry in similarity: Good is more alike than bad. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *42*(8), 1171–1192. https://doi.org/10.1037/xlm0000243

Kramer, D. A., & Schmidt, S. R. (2007). Alcohol beverage cues impair memory in high social drinkers. *Cognition and Emotion*, *21*(7), 1535–1545. https://doi.org/10.1080/02699930701193369

Krueger, J. (1991). Accentuation effects and illusory change in exemplar-based category learning. *European Journal of Social Psychology*, *21*(1), 37–48. https://doi.org/10.1002/ejsp.2420210104

Krueger, J., & Rothbart, M. (1990). Contrast and accentuation effects in category learning. *Journal of Personality and Social Psychology*, *59*(4), 651–663. https://doi.org/10.1037/0022-3514.59.4.651

Krueger, J., Rothbart, M., & Sriram, N. (1989). Category learning and change: Differences in sensitivity to information that enhances or reduces intercategory distinctions. *Journal of Personality and Social Psychology*, *56*(6), 866–875. https://doi.org/10.1037/0022-3514.56.6.866

Kruschke, J. K. (2003). Attention in Learning. *Current Directions in Psychological Science*, *12*(5), 171–175. https://doi.org/10.1111/1467-8721.01254

Kruschke, J. K., Kappenman, E. S., & Hetrick, W. P. (2005). Eye Gaze and Individual Differences Consistent With Learned Attention in Associative Blocking and Highlighting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 830–845. https://doi.org/10.1037/0278-7393.31.5.830

Kuhbandner, C., & Pekrun, R. (2013). Joint effects of emotion and color on memory. *Emotion*, *13*(3), 375–379. https://doi.org/10.1037/a0031821

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: the P300 as a measure of stimulus evaluation time. *Science*, *197*(4305), 792–795. https://doi.org/10.1126/science.887923

Kutas, Marta, & Federmeier, K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology*, *62*(1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, Marta, & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205. https://doi.org/10.1126/science.7350657

Kutzner, F., Vogel, T., Freytag, P., & Fiedler, K. (2011). A Robust Classic: Illusory Correlations are Maintained under Extended Operant Learning. *Experimental Psychology*, *58*(6), 443–453. https://doi.org/10.1027/1618-3169/a000112

Lahl, O., Göritz, A. S., Pietrowsky, R., & Rosenberg, J. (2009). Using the World-Wide Web to obtain large-scale word norms: 190,212 ratings on a set of 2,654 German nouns. *Behavior Research Methods*, *41*(1), 13–19. https://doi.org/10.3758/BRM.41.1.13

Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology*, *32*(2), 311–328. https://doi.org/10.1037/0022-3514.32.2.311

Lilli, W., & Rehm, J. (1983). Theoretische und empirische Untersuchungen zum Phänomen der “illusorischen Korrelation” (illusory correlation). I. Ableitung von Randbedingungen für das Auftreten von Effekten der illusorischen Korrelation aus dem Konzept der Verfügbarkeits-

(availability-)Heuristik. *Zeitschrift für Sozialpsychologie*, *14*(3), 251–261.

Lilli, W., & Rehm, J. (1984). Theoretische und empirische Untersuchungen zum Phänomen der Zusammenhangstäuschung. II. Entwicklung eines Modells zum quantitativen Urteil und Diskussion seiner Implikationen für die soziale Urteilsbildung. *Zeitschrift Für Sozialpsychologie*, *15*, 60–73.

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology*, *54*(1), 146–157. https://doi.org/10.1111/psyp.12639

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*(3), 252–271. https://doi.org/10.1037/0033-295X.87.3.252

Mangels, J. A., Picton, T. W., & Craik, F. I. M. (2001). Attention and successful episodic encoding: an event-related potential study. *Cognitive Brain Research*, *11*(1), 77–95. https://doi.org/10.1016/S0926-6410(00)00066-5

Mather, M. (2007). Emotional Arousal and Memory Binding: An Object-Based Framework. *Perspectives on Psychological Science*, *2*(1), 33–52. https://doi.org/10.1111/j.1745-6916.2007.00028.x

McConnell, A. R., Sherman, S. J., & Hamilton, D. L. (1994). Illusory correlation in the perception of groups: An extension of the distinctiveness-based account. *Journal of Personality and Social Psychology*, *67*(3), 414–429. https://doi.org/10.1037/0022-3514.67.3.414

McDaniel, M. A., DeLosh, E. L., & Merritt, P. S. (2000). Order information and retrieval distinctiveness: Recall of common versus bizarre material. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(4), 1045–1056. https://doi.org/10.1037/0278-7393.26.4.1045

McDaniel, M. A., & Geraci, L. (2006). Encoding and Retrieval Processes in Distinctiveness Effects: Towards an Integrative Framework. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and memory* (pp. 65–88). Oxford ; New York: Oxford University Press.

McEvoy, C. L., & Nelson, D. L. (1982). Category Name and Instance Norms for 106 Categories of Various Sizes. *The American Journal of Psychology*, *95*(4), 581–634. https://doi.org/10.2307/1422189

McGarty, C., & de la Haye, A.-M. (1997). Stereotype formation: Beyond illusory correlation. In R. Spears, P. J. Oakes, N. Ellemers, & S. A. Haslam (Eds.), *The social psychology of stereotyping and group life* (pp. 144–170). Malden: Blackwell Publishing.

McGarty, C., Haslam, S. A., Turner, J. C., & Oakes, P. J. (1993). Illusory correlation as accentuation of actual intercategory difference: Evidence for the effect with minimal stimulus information. *European Journal of Social Psychology*, *23*(4), 391–410. https://doi.org/10.1002/ejsp.2420230406

Medin, D. L., & Edelson, S. M. (1988). Problem structure and the use of base-rate information from experience. *Journal of Experimental Psychology: General*, *117*(1), 68–85. https://doi.org/10.1037/0096-3445.117.1.68

Meiser, T. (2003). Effects of processing strategy on episodic memory and contingency learning in group stereotype formation. *Social Cognition*, *21*(2), 121–156. https://doi.org/10.1521/soco.21.2.121.21318

Meiser, T. (2008). Illusorische Korrelation. In L.-E. Petersen & B. Six (Eds.), *Stereotype, Vorurteile und soziale Diskriminierung: Theorien, Befunde und Interventionen* (1. Aufl., pp. 53–61). Weinheim: Beltz, PVU.

Meiser, T. (2011). Much Pain, Little Gain? Paradigm-Specific Models and Methods in Experimental Psychology. *Perspectives on Psychological Science*, *6*(2), 183–191. https://doi.org/10.1177/1745691611400241

Meiser, T., & Hewstone, M. (2001). Crossed categorization effects on the formation of illusory correlations. *European Journal of Social Psychology*, *31*(4), 443–466. https://doi.org/10.1002/ejsp.55

Mendes, W. B., Blascovich, J., Hunter, S. B., Lickel, B., & Jost, J. T. (2007). Threatened by the unexpected: Physiological responses during social interactions with expectancy-violating partners. *Journal of Personality and Social Psychology*, *92*(4), 698–716. https://doi.org/10.1037/0022-3514.92.4.698

Meyer, P., Mecklinger, A., & Friederici, A. D. (2007). Bridging the gap between the semantic N400 and the early old/new memory effect: *NeuroReport*, *18*(10), 1009–1013. https://doi.org/10.1097/WNR.0b013e32815277eb

Michelon, P., & Snyder, A. Z. (2006). Neural correlates of incongruity. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and memory* (pp. 361–380). Oxford ; New York: Oxford University Press.

Monfort, V., & Pouthas, V. (2003). Effects of working memory demands on frontals slow waves in time-interval reproduction tasks in humans. *Neuroscience Letters*, *343*(3), 195-199. https://doi.org/10.1016/S0304-3940(03)00385-9

Mullen, B., & Johnson, C. (1990). Distinctiveness-based illusory correlations and stereotyping: A meta-analytic integration. *British Journal of Social Psychology*, *29*(1), 11–28. https://doi.org/10.1111/j.2044-8309.1990.tb00883.x

Murphy, R. A., Schmeer, S., Vallée-Tourangeau, F., Mondragón, E., & Hilton, D. (2011). Making the illusory correlation effect appear and then disappear: The effects of increased learning. *The Quarterly Journal of Experimental Psychology*, *64*(1), 24–40. https://doi.org/10.1080/17470218.2010.493615

Näätänen, R., Schröger, E., Karakas, S., Tervaniemi, M., & Paavilainen, P. (1993). Development of a memory trace for a complex sound in the human brain. *Neuroreport*, *4*(5), 503–506.

Näätänen, Risto, Pakarinen, S., Rinne, T., & Takegata, R. (2004). The mismatch negativity (MMN): towards the optimal paradigm. *Clinical Neurophysiology*, *115*(1), 140–144. https://doi.org/10.1016/j.clinph.2003.04.001

Nelson, T. O., & Narens, L. (1990). Metamemory: A Theoretical Framework and New Findings. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 26, pp. 125–173). Academic Press. https://doi.org/10.1016/S0079-7421(08)60053-5

Neville, H. J., Kutas, M., Chesney, G., & Schmidt, A. L. (1986). Event-related brain potentials during initial encoding and recognition memory of congruous and incongruous words. *Journal of Memory and Language, 25*(1), 75–92. https://doi.org/10.1016/0749-596X(86)90022-7

Nisbett, R. E., & Ross, L. (1980). *Human inference: strategies and shortcomings of social judgment*. Englewood Cliffs, N.J: Prentice-Hall.

Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. *Journal of Experimental Psychology: General, 130*(3), 466–478. https://doi.org/10.1037/0096-3445.130.3.466

Olichney, J. M., Petten, C. V., Paller, K. A., Salmon, D. P., Iragui, V. J., & Kutas, M. (2000). Word repetition in amnesia. *Brain, 123*(9), 1948–1963. https://doi.org/10.1093/brain/123.9.1948

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*(4), 725–729. https://doi.org/10.1037/0278-7393.9.4.725

Osterhout, L., Bersick, M., & Mclaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory & Cognition*, *25*(3), 273–285. https://doi.org/10.3758/BF03211283

Otten, L. J., & Donchin, E. (2000). Relationship between P300 amplitude and subsequent recall for distinctive events: Dependence on type of distinctiveness attribute. *Psychophysiology*, *37*(5), 644–661. https://doi.org/10.1111/1469-8986.3750644

Otten, L. J., & Rugg, M. D. (2001). When more means less: neural activity related to unsuccessful memory encoding. *Current Biology*, *11*(19), 1528–1530. https://doi.org/10.1016/S0960-9822(01)00454-7

Paller, K. A., McCarthy, G., & Wood, C. C. (1988). ERPs predictive of subsequent recall and recognition performance. *Biological Psychology*, *26*(1–3), 269–276. https://doi.org/10.1016/0301-0511(88)90023-3

Paller, K. A., & Wagner, A. D. (2002). Observing the transformation of experience into memory. *Trends in Cognitive Sciences*, *6*(2), 93–102. https://doi.org/10.1016/S1364-6613(00)01845-3

Pandey, S., & Elliott, W. (2010). Suppressor Variables in Social Work Research: Ways to Identify in Multiple Regression Models. *Journal of the Society for Social Work and Research*, *1*(1), 28–40. https://doi.org/10.5243/jsswr.2010.2

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, *2*, 10. https://doi.org/10.3389/neuro.11.010.2008

Pettigrew, T. F. (1998). Intergroup Contact Theory. *Annual Review of Psychology*, *49*(1), 65–85. https://doi.org/10.1146/annurev.psych.49.1.65

Peynircioğlu, Z. F., & Mungan, E. (1993). Familiarity, relative distinctiveness, and the generation effect. *Memory & Cognition*, *21*(3), 367–374. https://doi.org/10.3758/BF03208269

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2018). nlme: Linear and Nonlinear Mixed Effects Models (Version 3.1-137). Retrieved from https://CRAN.R-project.org/package=nlme

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Pryor, J. B. (1986). The Influence of Different Encoding Sets Upon the Formation of Illusory Correlations and Group Impressions. *Personality and Social Psychology Bulletin*, *12*(2), 216–226. https://doi.org/10.1177/0146167286122008

Ratliff, K. A., & Nosek, B. A. (2010). Creating distinct implicit and explicit attitudes with an illusory correlation paradigm. *Journal of Experimental Social Psychology*, *46*(5), 721–728. https://doi.org/10.1016/j.jesp.2010.04.011

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*(6), 855–863. https://doi.org/10.1016/S0022-5371(67)80149-X

Rescorla, R. A., & Wagner, A. R. (1972). A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Risen, J. L., Gilovich, T., & Dunning, D. (2007). One-Shot Illusory Correlations and Stereotype Formation. *Personality and Social*

Psychology Bulletin, *33*(11), 1492–1502. https://doi.org/10.1177/0146167207305862

Rosburg, T., Johansson, M., Weigl, M., & Mecklinger, A. (2015). How does testing affect retrieval-related processes? An event-related potential (ERP) study on the short-term effects of repeated retrieval. *Cognitive, Affective, & Behavioral Neuroscience*, *15*(1), 195–210. https://doi.org/10.3758/s13415-014-0310-y

Rosburg, T., Mecklinger, A., & Frings, C. (2011). When the Brain Decides A Familiarity-Based Approach to the Recognition Heuristic as Evidenced by Event-Related Brain Potentials. *Psychological Science*, *22*(12), 1527–1534. https://doi.org/10.1177/0956797611417454

Rosburg, T., Weigl, M., & Sörös, P. (2014). Habituation in the absence of a response decrease? *Clinical Neurophysiology*, *125*(1), 210–211. https://doi.org/10.1016/j.clinph.2013.06.011

Rothbart, M. (1981). Memory Processes and Social Beliefs. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 145–181). Hillsdale, N.J.: L. Erlbaum Associates.

Rothbart, M., Fulero, S., Jensen, C., Howard, J., & Birrell, P. (1978). From individual to group impressions: Availability heuristics in stereotype formation. *Journal of Experimental Social Psychology*, *14*(3), 237–255. https://doi.org/10.1016/0022-1031(78)90013-6

Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review*, *103*(4), 734. https://doi.org/10.1037/0033-295X.103.4.734

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences*, *11*(6), 251–257. https://doi.org/10.1016/j.tics.2007.04.004

Sable, J. J., Low, K. A., Maclin, E. L., Fabiani, M., & Gratton, G. (2004). Latent inhibition mediates N1 attenuation to repeating sounds. *Psychophysiology*, *41*(4), 636–642. https://doi.org/10.1111/j.1469-8986.2004.00192.x

Sanbonmatsu, D. M., Sherman, S. J., & Hamilton, D. L. (1987). Illusory correlation in the perception of individuals and groups. *Social Cognition*, *5*(1), 1–25.

Sanquist, T. F., Rohrbaugh, J. W., Syndulko, K., & Lindsley, D. B. (1980). Electrocortical Signs of Levels of Processing: Perceptual Analysis and Recognition Memory. *Psychophysiology*, *17*(6), 568–576. https://doi.org/10.1111/j.1469-8986.1980.tb02299.x

Schmidt, S. R. (1991). Can we have a distinctive theory of memory? *Memory & Cognition*, *19*(6), 523–542. https://doi.org/10.3758/BF03197149

Schmidt, S. R. (2006). Emotion, Significance, Distinctiveness, and Memory. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and memory* (pp. 47–64). Oxford ; New York: Oxford University Press.

Schmidt, S. R. (2012). *Extraordinary memories for exceptional events*. New York, NY: Psychology Press.

Schmidt, S. R., & Saari, B. (2007). The emotional memory effect: Differential processing or item distinctiveness? *Memory & Cognition*, *35*(8), 1905–1916. https://doi.org/10.3758/BF03192924

Schröder, A., Gemballa, T., Ruppin, S., & Wartenburger, I. (2011). German norms for semantic typicality, age of acquisition, and concept familiarity. *Behavior Research Methods, 44*(2), 380–394. https://doi.org/10.3758/s13428-011-0164-y

Schubotz, R. (1999). Instruction differentiates the processing of temporal and spatial sequential patterns: evidence from slow wave activity in humans. *Neuroscience Letters*, *265*(1), 1–4. https://doi.org/10.1016/S0304-3940(99)00152-4

Schupp, H. T., Cuthbert, B. N., Bradley, M. M., Cacioppo, J. T., Ito, T., & Lang, P. J. (2000). Affective picture processing: The late positive potential is modulated by motivational relevance. *Psychophysiology*, *37*(2), 257–261.

Seligman, M. E. P. (1971). Phobias and preparedness. *Behavior Therapy, 2*(3), 307–320. https://doi.org/10.1016/S0005-7894(71)80064-3

Seligman, M. E. P. (1992). *Helplessness: On depression, development, and death* ([Repr.]). New York: Freeman.

Shavitt, S., Sanbonmatsu, D. M., Smittipatana, S., & Posavac, S. S. (1999). Broadening the Conditions for Illusory Correlation Formation: Implications for Judging Minority Groups. *Basic and Applied Social Psychology*, *21*(4), 263–279. https://doi.org/10.1207/S15324834BASP2104_1

Shepherd, J. W., Gibling, F., & Ellis, H. D. (1991). The effects of distinctiveness, presentation time and delay on face recognition. *European Journal of Cognitive Psychology*, *3*(1), 137–145. https://doi.org/10.1080/09541449108406223

Sherman, J. W., Kruschke, J. K., Sherman, S. J., Percy, E. J., Petrocelli, J. V., & Conrey, F. R. (2009). Attentional processes in stereotype formation: A common model for category accentuation and illusory correlation. *Journal of Personality and Social Psychology*, *96*(2), 305–323. https://doi.org/10.1037/a0013778

Skinner, F. B. (1992). “Superstition” in the pigeon. *Journal of Experimental Psychology: General*, *121*(3), 273–274. https://doi.org/10.1037/0096-3445.121.3.273

Slusher, M. P., & Anderson, C. A. (1987). When reality monitoring fails: The role of imagination in stereotype maintenance. *Journal of Personality and Social Psychology*, *52*(4), 653–662. https://doi.org/10.1037/0022-3514.52.4.653

Smedslund, J. (1963). The Concept of Correlation in Adults. *Scandinavian Journal of Psychology*, *4*(1), 165–173. https://doi.org/10.1111/j.1467-9450.1963.tb01324.x

Smith, E. R. (1991). Illusory correlation in a simulated exemplar-based memory. *Journal of Experimental Social Psychology*, *27*(2), 107–123. https://doi.org/10.1016/0022-1031(91)90017-Z

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117*(1), 34–50. https://doi.org/10.1037/0096-3445.117.1.34

Sokolov, E. N. (1963). Higher Nervous Functions: The Orienting Reflex. *Annual Review of Physiology*, *25*(1), 545–580. https://doi.org/10.1146/annurev.ph.25.030163.002553

Spears, Russell, Eiser, J. R., & van der Pligt, J. (1987). Further evidence for expectation-based illusory correlations. *European Journal of Social Psychology*, *17*(2), 253–258. https://doi.org/10.1002/ejsp.2420170211

Spears, Russell, van der Pligt, J., & Eiser, J. R. (1986). Generalizing the illusory correlation effect. *Journal of Personality and Social Psychology*, *51*(6), 1127–1134. https://doi.org/10.1037/0022-3514.51.6.1127

Spiers, H. J., Love, B. C., Le Pelley, M. E., Gibb, C. E., & Murphy, R. A. (2016). Anterior Temporal Lobe Tracks the Formation of Prejudice. *Journal of Cognitive Neuroscience*, *29*(3), 530–544. https://doi.org/10.1162/jocn_a_01056

Stangor, C., & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and social developmental literatures. *Psychological Bulletin*, *111*(1), 42. https://doi.org/10.1037/0033-2909.111.1.42

Starr, B. J., & Katkin, E. S. (1969). The clinician as an aberrant actuary: Illusory correlation and the incomplete sentences blank. *Journal of Abnormal Psychology*, *74*(6), 670–675. https://doi.org/10.1037/h0028466

Strange, B. A., Henson, R. N. A., Friston, K. J., & Dolan, R. J. (2000). Brain Mechanisms for Detecting Perceptual, Semantic, and Emotional Deviance. *NeuroImage*, *12*(4), 425–433. https://doi.org/10.1006/nimg.2000.0637

Stroessner, S. J., Hamilton, D. L., & Mackie, D. M. (1992). Affect and stereotyping: The effect of induced mood on distinctiveness-based illusory correlations. *Journal of Personality and Social Psychology*, *62*(4), 564–576. https://doi.org/10.1037/0022-3514.62.4.564

Stroessner, S. J., & Plaks, J. E. (2001). Illusory correlation and stereotype formation: Tracing the arc of reseach over a quarter century. In G. B. Moskowitz (Ed.), *Cognitive social psychology: the Princeton Symposium on the Legacy and Future of Social Cognition* (pp. 247–259). Mahwah, NJ: Lawrence Erlbaum Associates.

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-Potential Correlates of Stimulus Uncertainty. *Science*, *150*(3700), 1187–1188. https://doi.org/10.1126/science.150.3700.1187

Suzuki, A., & Suga, S. (2010). Enhanced memory for the wolf in sheep's clothing:: Facial trustworthiness modulates face-trait associative memory. *Cognition*, *117*(2), 224–229. https://doi.org/10.1016/j.cognition.2010.08.004

Tajfel, H. (1959). Quantitative Judgement in Social Perception. *British Journal of Psychology*, *50*(1), 16–29. https://doi.org/10.1111/j.2044-8295.1959.tb00677.x

Tajfel, H., & Wilkes, A. L. (1963). Classification and Quantitative Judgement. *British Journal of Psychology*, *54*(2), 101–114. https://doi.org/10.1111/j.2044-8295.1963.tb00865.x

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, *103*(2), 193–210. https://doi.org/10.1037/0033-2909.103.2.193

Tibon, R., & Levy, D. A. (2014). The time course of episodic associative retrieval: Electrophysiological correlates of cued recall of unimodal and crossmodal pair-associate learning. *Cognitive, Affective, & Behavioral Neuroscience*, *14*(1), 220–235. https://doi.org/10.3758/s13415-013-0199-x

Tibon, R., & Levy, D. A. (2015). Striking a balance: analyzing unbalanced event-related potential data. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00555

Tomarken, A. J., Mineka, S., & Cook, M. (1989). Fear-relevant selective associations and covariation bias. *Journal of Abnormal Psychology*, *98*(4), 381–394. https://doi.org/10.1037/0021-843X.98.4.381

Tomarken, A. J., Sutton, S. K., & Mineka, S. (1995). Fear-relevant illusory correlations: What types of associations promote judgmental bias? *Journal of Abnormal Psychology*, *104*(2), 312–326. https://doi.org/10.1037/0021-843X.104.2.312

Tulving, E. (1969). Retrograde Amnesia in Free Recall. *Science*, *164*(3875), 88–90. https://doi.org/10.1126/science.164.3875.88

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie Canadienne*, *26*(1), 1–12. https://doi.org/10.1037/h0080017

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232. https://doi.org/10.1016/0010-0285(73)90033-9

van der Meulen, M. a., Anton, F., & Petersen, S. (2017). Painful decisions: How classifying sensations can change the experience of pain. *European Journal of Pain*, n/a-n/a. https://doi.org/10.1002/ejp.1061 van Kesteren, M. T. R., Ruiter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in Neurosciences*, *35*(4), 211–219. https://doi.org/10.1016/j.tins.2012.02.001

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language*, *50*(3), 289–335. https://doi.org/10.1016/j.jml.2003.10.003

Van Rooy, D., Van Overwalle, F., Vanhoomissen, T., Labiouse, C., & French, R. (2003). A recurrent connectionist model of group biases. *Psychological Review*, *110*(3), 536–563.

Van Rooy, D., Vanhoomissen, T., & Van Overwalle, F. (2013). Illusory correlation, group size and memory. *Journal of Experimental Social Psychology*, *49*(6), 1159–1167. https://doi.org/10.1016/j.jesp.2013.05.006

Verleger, R., Heide, W., Butt, C., & Kömpf, D. (1994). Reduction of P3b in patients with temporo-parietal lesions. *Cognitive Brain Research*, *2*(2), 103–116. https://doi.org/10.1016/0926-6410(94)90007-8

Võ, M. L. H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin Affective Word List Reloaded (BAWL-R). *Behavior Research Methods*, *41*(2), 534–538. https://doi.org/10.3758/BRM.41.2.534

von Restorff, H. (1933). Über die Wirkung von Bereichsbildungen im Spurenfeld. *Psychologische Forschung*, *18*(1), 299–342. https://doi.org/10.1007/BF02409636

Voss, J. L., & Federmeier, K. D. (2011). FN400 potentials are functionally identical to N400 potentials and reflect semantic processing during recognition testing. *Psychophysiology*, *48*(4), 532–546. https://doi.org/10.1111/j.1469-8986.2010.01085.x

Waddill, P. J., & McDaniel, M. A. (1998). Distinctiveness effects in recall: Differential processing or privileged retrieval? *Memory & Cognition*, *26*(1), 108–120. https://doi.org/10.3758/BF03211374

Wagner, H. L. (1993). On measuring performance in category judgment studies of nonverbal behavior. *Journal of Nonverbal Behavior*, *17*(1), 3–28. https://doi.org/10.1007/BF00987006

Wallace, W. P. (1965). Review of the historical, empirical, and theoretical status of the von Restorff phenomenon. *Psychological Bulletin*, *63*(6), 410–424. https://doi.org/10.1037/h0022001

Weigl, M., Mecklinger, A., & Rosburg, T. (2015). Accurately perceived, falsely retrieved: Illusory correlations originate from biased retrieval of accurately encoded contingencies. In *Abstracts of the 57th Conference of Experimental Psychologists* (p. 271). Lengerich: Pabst Science Publishers.

Weigl, M., Mecklinger, A., & Rosburg, T. (2016). Transcranial direct current stimulation over the left dorsolateral prefrontal cortex modulates auditory mismatch negativity. *Clinical Neurophysiology*, *127*(5), 2263–2272. https://doi.org/10.1016/j.clinph.2016.01.024

Weigl, M., Mecklinger, A., & Rosburg, T. (2018). Illusory correlations despite equated category frequencies: A test of the information loss account.

Consciousness and Cognition, *63*, 11–28. https://doi.org/10.1016/j.concog.2018.06.002

Worthen, J. B., Marshall, P. H., & Cox, K. B. (1998). List length and the bizarreness effect: Support for a hybrid explanation. *Psychological Research*, *61*(2), 147–156. https://doi.org/10.1007/s004260050021

Yonelinas, A. P., & Jacoby, L. L. (1995). The Relation between Remembering and Knowing as Bases for Recognition: Effects of Size Congruency. *Journal of Memory and Language*, *34*(5), 622–643. https://doi.org/10.1006/jmla.1995.1028

Yonelinas, Andrew P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(6), 1341–1354. https://doi.org/10.1037/0278-7393.20.6.1341

Yonelinas, Andrew P. (2002). The Nature of Recollection and Familiarity: A Review of 30 Years of Research. *Journal of Memory and Language*, *46*(3), 441–517. https://doi.org/10.1006/jmla.2002.2864

Yonelinas, Andrew P., Aly, M., Wang, W.-C., & Koen, J. D. (2010). Recollection and familiarity: Examining controversial assumptions and new directions. *Hippocampus*, *20*(11), 1178–1194. https://doi.org/10.1002/hipo.20864

Zaehle, T., Sandmann, P., Thorne, J. D., Jäncke, L., & Herrmann, C. S. (2011). Transcranial direct current stimulation of the prefrontal cortex modulates working memory performance: combined behavioural and electrophysiological evidence. *BMC Neuroscience*, *12*(2), 1–11. https://doi.org/10.1186/1471-2202-12-2

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*(2, Pt.2), 1–27. https://doi.org/10.1037/h0025848

Curriculum Vitae

Personal Data

Name:	Michael Weigl
Date of Birth:	03/03/1987
Place of Birth:	Brackenheim (Baden-Wuerttemberg, Germany)
Nationality:	German

Education

2012 – 2019	PhD student at the Experimental Neuropsychology Unit, Saarland University, Germany
2012	Certificate in East Asia Studies at the University Trier, Rhineland-Palatinate, Germany
2012	Diploma in Psychology, Saarland University, Germany
2007 – 2012	Studies in Psychology, Saarland University, Saarland, Germany
1998 – 2007	Annette-Kolb-Gymnasium Traunstein, Bavaria, Germany

Career/Employment

Since 2019	Post-Doc in the DFG project “The aging episodic memory and its plasticity: A cross-cultural approach” at the Experimental Neuropsychology Unit, Saarland University, Germany
2018 – 2019	Lecturer at the Institute of Psychology, Saarland University, Germany
2017	Research stay at the Chinese Academy of Science in Beijing, China
2012 – 2018	Research assistant at the Experimental Neuropsychology Unit, Saarland University
2012	Student research assistant at the Social Psychology Unit, Saarland University, Saarland, Germany
2011 – 2012	Part-time research internship at the Experimental Neuropsychology Unit, Saarland University, Saarland, Germany
2010 – 2012	Tutor for the course “Social Psychology” at the Social Psychology Unit, Saarland University, Germany

2010 – 2010 Tutor for the course “Empirical Training II” at the Social Psychology Unit, Saarland University, Germany

2010 – 2011 Student research assistant for the DFG project “Affective Science” at the Social Psychology Unit, Saarland University, Germany

2009 – 2010 Part-time research internship at the Social Psychology Unit, Saarland University, Saarland, Germany

2007 Full-time internship at the Chiemgau Lebenshilfe Werkstätten, Bavaria, Germany

Publications

Journal Articles (peer-reviewed)

Rosburg, T., Weigl, M., & Deuring, G. (2019). Enhanced processing of facial emotion for target stimuli. *International Journal of Psychophysiology, 146*, 190-200. https://doi.org/10.1016/j.ijpsycho.2019.08.010

Cai, X., Weigl, M., Liu, B., Cheung, E., Ding, J., Chan, R. C. K. (2019). Delay discounting and affective priming in individuals with negative schizotypy. *Schizophrenia Research, 210*, 180-187. https://doi.org/ 10.1016/j.schres.2018.12.040

Weigl, M., Mecklinger, A., & Rosburg, T. (2018). Illusory correlations despite equated category frequencies: A test of the information loss account. *Consciousness and Cognition, 63*, 11-28. https://doi.org/10.1016/ 10.1016/j.concog.2018.06.002

Rosburg, T., Weigl, M., Thiel, R., & Mager, R. (2018). The event-related potential component P3a is diminished by identical deviance repetition, but not by non-identical repetitions. *Experimental Brain Reseach, 236*, 1519-1530. https://doi.org/ 10.1007/s00221-018-5237-z

Zhou, H., Cai, X., Weigl, M., Bang, P., Cheung, E. F. C., & Chan, R. C. K. (2018). Multisensory temporal binding window in autism spectrum disorders and schizophrenia spectrum disorders: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews, 86*, 66–76. https://doi.org/10.1016/j.neubiorev.2017.12.013

Weigl, M., Mecklinger, A., & Rosburg, T. (2016). Transcranial direct current stimulation over the left dorsolateral prefrontal cortex modulates auditory mismatch negativity. *Clinical Neurophysiology, 127(1)*, 2263-2272. https://doi.org/10.1016/j.clinph.2016.01.024

Rosburg, T., Johansson, M., Weigl, M., & Mecklinger, A. (2015). How does testing affect retrieval-related processes? An event-related potential (ERP) study on the short-term effects of repeated retrieval. *Cognitive, Affective & Behavioral Neuroscience, 15(1)*, 195-210. https://doi.org/10.3758/s13415-014-0310-y

Rosburg, T., Weigl, M., & Sörös, P. (2014). Habituation in the absence of a response decrease? *Clinical Neurophysiology, 125(1)*, 210-211. https://doi.org/10.1016/j.clinph.2013.06.011

Conference Contributions (posters and talks)

Weigl, M., Thiel, R., Mecklinger, A., & Rosburg, T. (2019, April). A comparison between distinctiveness and accentuation in the illusory correlation paradigm: An event-related potential study. Poster presented at the 61st Conference of Experimental Psychologists (TeaP) in London, United Kingdom.

Kamp, S.-M., Döring, J., Weigl, M., & Mecklinger, A. (2017, April). Influences of Novelty on Episodic Memory in Aging. Poster presented at the 4th International Conference on Aging & Cognition in Zurich, Switzerland.

Weigl, M., Pham, H. H., Mecklinger, A. & Rosburg, T. (2017, March). The effect of shared distinctiveness on source memory and illusory correlations: An event-relaed potential study. Poster presented at the 24th Annual Meeting of the Cognitive Neuroscience Society in San Francisco, CA, USA.

Weigl, M., Ehritt, A., Mecklinger, A., & Rosburg, T. (2016, April). Can event-related potentials at encoding predict whether subsequent recognition is based on familiarity or recollection? Poster presented at the 23rd Annual Meeting of the Cognitive Neuroscience Society in New York, USA.

Weigl, M., Mecklinger, A., & Rosburg, T. (2016). The role of episodic memory in illusory correlation - Evidence for the distinctiveness account [Abstract]. In J. Funke, J. Rummel, & A. Voß (Eds.), *Abstracts*

of the 58th Conference of Experimental Psychologists TeaP 2016, pp. 369-370, Lengerich: Pabst Science Publishing.

Weigl, M., Mecklinger, A., & Rosburg, T. (2015, July). Illusory correlations despite equated category frequencies: Evidence against the information loss account. Talk at the 14th European Congress of Psychology in Milan, Italy.

Weigl, M., Mecklinger, A., & Rosburg, T. (2015). Transkranielle Gleichstromstimulation über dem linken dorsolateralen präfrontalen Kortex moduliert die auditorische Mismatch Negativität, aber nicht die P3a und P3b. [Abstract]. *Kognitive Neurophysiologie des Menschen – Human Cognitive Neurophysiology*, *8*, 49-50.

Weigl, M., Loschelder, D. D., Friese, M., & Trötschel, R. (2015). Looking at my offer: Procedural framing of negotiation proposals affects the sender's reference point of a transaction. [Abstract]. In C. Bermeitinger, A. Mojzisch, & W. Greve (Eds.), *Abstracts of the 57th Conference of Experimental Psychologists TeaP 2015*, p. 270, Lengerich: Pabst Science Publishers.

Weigl, M., Mecklinger, A., & Rosburg, T. (2015). Accurately perceived, falsely retrieved: Illusory correlations originate from biased retrieval of accurately encoded contingencies. [Abstract]. In C. Bermeitinger, A. Mojzisch, & W. Greve (Eds.), *Abstracts of the 57th Conference of Experimental Psychologists TeaP 2015*, p. 271, Lengerich: Pabst Science Publishers.

Rosburg, T. Mecklinger, A., Weigl, M., & Johansson, M. (2013, April). The effects of immediate testing on neural correlates of recollection. Poster presented at the 20th Annual Meeting of the Cognitive Neuroscience Society in San Francisco, USA.

Schlemmer, B. A., Weigl, M., & Unz, D. (2011). The Story-Setting-Effect in the Light of Construal Level Theory [Abstract]. In O. Özen, M. Schreier, & Y. Thies-Brandner (Eds.), *Media Psychology. Focus Theme: Cognitive and Emotional Involvement during Media Reception. Proceedings of the 7th Conference of the Media Psychology Division of the German Psychological Society*, pp. 85-86. Lengerich: Pabst Science Publishers.

Unz, D., Schlemmer, B. A., & Weigl, M. (2010). Story-Setting-Effect revisited. Persuasion durch fiktionale Narrative? [Abstract]. In F. Peterman & U. Koglin (Hrsg.), *47. Kongress der Deutschen Gesellschaft für Psychologie*, p. 133. Lengerich: Pabst Science Publishers.