Behavioral and Psychophysiological Investigations into the Neurocognition of Social Influence

—*Kumulative Arbeit*—

Inaugural-Dissertation
zur Erlangung der Doktorwürde
der
Philosophischen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität
zu Bonn

vorgelegt von

Robert Schnürch

aus

Düsseldorf

Bonn, 2016

Gedruckt mit der Genehmigung der Philosophischen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn

Zusammensetzung der Prüfungskommission:
Prof. Dr. André Beauducel (Vorsitzender)
Prof. Dr. Henning Gibbons (Betreuer und Gutachter)
Prof. Dr. Martin Reuter (Gutachter)
Prof. Dr. Rainer Banse (weiteres prüfungsberechtigtes Mitglied)

Tag der mündlichen Prüfung: 28. September 2016

Contents

Acknowledgments

First and foremost, I would like to thank Henning for giving me the opportunity to work in his lab and under his supervision. He has provided me with an unimaginable combination of freedom (to explore my own ideas) and direction (to shape my ideas into measurable output). The time during my PhD could not have been any better. Henning's grounded, supportive, and amiable nature is unequaled in this business.

My sincere thanks go to Martin for generously taking the time to be my second reviewer. His help in the completion of my PhD is greatly appreciated.

Plenty of students helped me collect data for my experiments. In particular, I would like to thank our past and present student assistants and interns: Caroline, Charlotte, Christina, Hannah, Helena, Jan, Kristin, Lily, Nadja, Rafaela, Sabine, and Schazia. And for making work in this department a joyful, entertaining, and light endeavor, I want to thank my dear colleagues Anna and Judith.

Beyond doubt, I am indebted to my family for sharing with me their unwavering support and invigorating humor, their tender criticism and euphoric appreciation. I would like to extend my heartfelt gratitude to my allegiant and generous father as well as to my incredible siblings—and best friends in the world!—Anna, Lisa, and, in particular, Martin, who co-authored one of the papers and thoroughly proofread the synopsis. Yet ultimately, I owe it all to our wonderful mother, who is so sorely missed and still remains our guide and inspiration.

This thesis is for Carina, my one true love, closest friend, and partner in crime, whom I met in a dull, yet (in hindsight) invaluable psychology lecture more than seven years ago. Ever since, we have shared science just as much as we share everything else in our lives. I could not have done any of this work without her brilliant advice, her contagious encouragement, her unconditional love—for all of which I am eternally grateful.

Preface

Before laying out the actual synopsis of my research, I feel obliged to make a clarifying remark regarding the word *neurocognition*, a term that I frequently use throughout this thesis. Only recently, this very word has been mentioned among a list of 50 psychological and psychiatric terms that should be avoided due to their inherent imprecision (Lilienfeld et al., 2015). According to the authors of said list, all cognition is necessarily based upon neural activity, making this compound a pleonasm. Going even further, they have likened it to recently emerging neologisms, such as neuroeducation or neuropolitics, which—according to the authors—use the prefix “neuro” as a means of persuasion (Lilienfeld et al., 2015).

In my humble opinion, the word neurocognition is neither a pleonasm, nor is it comparable to other neuro-compounds that might appear similar at first glance. Cognitive processes are typically described at different levels of specification and precision. In many areas of psychology, researchers refer to cognition and the processes it comprises as coarse, broad, often overt aspects of behavior. For example, social psychologists might manipulate whether or not participants attend to a group norm (Cialdini, Reno, & Kallgren, 1990). However, attention is vaguely defined in this context, referring merely to an increased propensity to observe or consider certain cues. By applying different techniques, such as electrophysiological or neuroimaging methods, one can operationalize cognitive constructs more precisely. For example, in many neuroscientific studies, *all* stimuli are observed and consciously processed. Nevertheless, we might still be able to determine that some of them are processed more deeply than others. What would appear to be the same degree of attention in a behavioral sociopsychological context, where attention might be roughly equivalent to observance or interest, would be divided into different qualities of processing when taking a different methodological approach. While I would not suggest that either of the perspectives is more appropriate than the other, it seems helpful to openly acknowledge the level of analysis when talking about something as ill-defined as cognition, especially in interdisciplinary contexts such as the present thesis. Therefore, by using the term neurocognition, I clarify the level of resolution, the level of zoom into the mind, that I surmise in my own and other researchers’ work.

Specifically, my use of the term neurocognition conveys three relevant qualifications of how I interpret and conduct psychological investigations. First, I take a cognitive approach to behavior, believing that we can and need to disassemble macro-level processes of behavior and experience into putative small-scale units of computation. Second, I infer, characterize, and differentiate these cognitive processes in a somewhat fine-grained fashion, typically based on direct measures of neural activity that are obtained with millisecond precision. Third, by using this word, I acknowledge the constraints of my neurophysiological assessments, since many measures of brain activity, such as in the present thesis, are suitable proxies for cognitive processes, while they do not allow us to elucidate intricate details of human neurobiology itself. This approach is by no means my very own, personal take on cognition, but a customary perspective in cognitive neuroscience (Amodio, 2010a). In fact, it might epitomize a commonly advocated understanding of psychology based upon and enriched by, yet not completely absorbed in, biology (see Jancke & Petermann, 2010). In any case, I wish to clarify that the use of the word neurocognition in the present thesis is not necessarily as mindless and blemished as Lilienfeld and colleagues (2015) would have us think it is.

Abstract

Human behavior and experience is far from independent; social norms, conveying what is typical and appropriate in a given social group, thoroughly mold what we do, think, and feel. While much is known about the motivational mechanisms of social influence based on group norms, researchers have begun to decipher its neurocognitive principles only recently. In this thesis, I provide a brief introduction into psychological investigations of social influence and recently emerging neuroscientific research in this field. Moreover, I extend previous findings and preliminary theories on the basis of eight original publications, which include one review article, one behavioral study, and six studies employing event-related brain potentials (ERPs). Upon reviewing the existent literature, I have proposed that norm-based social influence involves two core processes: (i) the differentiation of social proof and deviance in neural circuits coding behavioral outcomes, and (ii) the direct modulation of the neural representations of to-be-judged features. Moreover, my colleagues and I have tested and clarified the validity of a common behavioral measure of conformity as well as the functional significance of ERP components relevant to the investigation of outcome processing and sustained attention. Most importantly, we have built upon our synthesis of the literature and our methodological studies and provide conclusive findings that confirm and extend the proposed accounts of social influence. We demonstrate that the differentiation of norms that confirm or contradict the individual's own views—thus constituting social proof or conveying social deviance, respectively—engages both early and late processes of attentional prioritization. Episodes of social proof, as well as stimuli associated with it, take precedence over those related to deviance. This finding is replicated across different samples, tasks, and paradigms. We propose that social norms feed into basic cognitive processes, similar to the value-informed deployment of attentional resources. As a whole, my series of research corroborates the notion that social influence is based upon some of our general cognitive faculties, such as outcome valuation and reinforcement learning, and can entail modulations of even the most fundamental functions, such as perception and attention. This could explain the pervasiveness of social influence on human behavior and experience, and it speaks to the timeliness and relevance of exploring its neurocognitive foundations.

Zusammenfassung

Menschliches Verhalten und Erleben ist alles andere als unabhängig; soziale Normen, welche vermitteln, was in einer sozialen Gruppe üblich und angebracht ist, formen, was wir tun, denken und fühlen. Während viel über die motivationalen Mechanismen von sozialem Einfluss, der auf Normen basiert, bekannt ist, haben Forscher erst in letzter Zeit damit begonnen, die zugrundeliegenden neurokognitiven Prinzipien zu entschlüsseln. In dieser Dissertation liefere ich eine kurze Einführung in psychologische Untersuchungen des sozialen Einflusses sowie in die neue neurowissenschaftliche Forschung in diesem Feld. Außerdem erweitere ich vorhandene Befunde und vorläufige Theorien zur Neurokognition des sozialen Einflusses auf der Grundlage von acht Originalarbeiten, welche einen Übersichtsartikel, eine Verhaltensstudie sowie sechs Studien mit ereigniskorrelierten Hirnpotenzialen (ERPs) umfassen. Nach Durchsicht der Literatur schlage ich vor, dass Norm-basierter sozialer Einfluss zwei zentrale Prozesse beinhaltet: (i) die Unterscheidung sozialer Bewährtheit und Abweichung in neuronalen Strukturen, die Verhaltenserträge kodieren, und (ii) die unmittelbare Modifikation der neuronalen Repräsentationen zu beurteilender Reizmerkmale. Darüber hinaus haben meine Kollegen und ich die Validität eines gängigen Verhaltensmaßes der Konformität und die funktionelle Signifikanz von EKP-Komponenten, die für die Untersuchung von Ergebnisverarbeitung und anhaltender Aufmerksamkeit von Bedeutung sind, getestet und klargestellt. Vor allem aber haben wir auf unserer Synthese der Literatur und auf unseren methodologischen Studien aufgebaut und präsentieren schlüssige Befunde, die die vorgeschlagenen Mechanismen des sozialen Einflusses bestätigen und erweitern. So zeigen wir, dass die Differenzierung von Normen, die die eigenen Ansichten unterstützen oder ihnen widersprechen und somit soziale Bewährtheit oder soziale Devianz vermitteln, sowohl frühe als auch späte Prozesse attentionaler Priorisierung beinhaltet. Episoden sozialer Bewährtheit und Reize, die damit assoziiert sind, haben Vorrang gegenüber solchen, die mit sozialer Devianz assoziiert sind. Diesen Befund konnten wir in verschiedenen Stichproben, Aufgaben und Paradigmen replizieren. Wir schlagen vor, dass soziale Normen basale kognitive Prozesse beeinflussen, ähnlich der wertabhängigen Ausrichtung von Aufmerksamkeitsressourcen. Insgesamt unterstreicht meine Forschungsreihe die Idee, dass sozialer Einfluss auf einigen unserer allgemeinen kognitiven Fähigkeiten wie der Bewertung von Erträgen und Verstärkungslernen basiert und Veränderungen selbst der grundlegendsten Funktionen wie Wahrnehmung und Aufmerksamkeit zur Folge haben kann. Das könnte die Allgegenwärtigkeit von sozial beeinflusstem Verhalten und Erleben erklären und unterstreicht die Aktualität und Bedeutung der Untersuchung der neurokognitiven Grundlagen des sozialen Einflusses.

The only way we know it's true is that we both dreamed it.
That's what reality is. It's a dream everyone has together.

Jeffrey Eugenides in *Middlesex* (2002)

1 Introduction

What do you do when you sit in the movie theater and get the faint impression of fumes in the air? Most likely, you will assume that there is a fire nearby. Sniffing the unsettling smell of smoke, your first instinct is to save yourself and potentially others as well by yelling “Fire!” and storming out of the theater. On second thought, however, this seems both foolish and, in a way, risky. You might simply be wrong and miss the rest of a promising movie. Moreover, you can easily imagine what the others would think of you if there was no actual fire. If your response was so aberrant, because what you smell is actually a bucket of slightly overcooked popcorn, the other people in the theater would certainly think you are a dramatic person who gets overly excited and imaginative when watching a movie in the dark. Pondering these potential costs of falsely raising an alarm, you just remain in your seat. And what you would most certainly do then, instinctively and naturally, would be to look around yourself to find out whether anybody else in the theater seems as discomposed as you. If no one did, you would most likely call your first impression into question, trying to find out if what you smell *really* is the smoke of a fire. But if you did evidence enough people turning their heads and sniffing dubiously, you would decide that you were right: There is a fire and you should get out of the theater.

The above example foreshadows several components of the psychological phenomena and processes that will be introduced, investigated, and discussed in the present thesis. First, we are social beings, constantly surrounding ourselves with other people. Therefore, much of what is intriguing about our species is how we behave within or in relation to groups. Second, we are aware of other people’s expectations regarding our own behavior and fear the consequences of our deviance, even if it is only the impression we make that is at stake. Third, and most importantly, we closely monitor the behavior of those around us not only—perhaps not even primarily—to discern their expectations, but to compare our own take on reality with theirs. We try to find proof for what we assume to be accurate. As Deutsch and Gerard (1955, p. 635) put it: “It is not surprising that the judgments of others [...] should be taken as evidence to be weighed in coming to one’s own judgment. From birth on, we learn that the perceptions and judgments of others are frequently reliable sources of evidence about reality.”

The process of using other people's judgments to validate one's own views and facilitate decision-making is widely "underappreciated" (Cialdini, 2007). Its influence on a person's behavior is, however, well-documented. For example, learning through the media that others have chosen to kill themselves as a means to deal with their problems encourages some people to commit suicide as well, often in the specific way in which the initial suicide conveyed through the media was committed (see, e.g., Niederkrotenthaler et al., 2010; Phillips, 1980; Phillips & Carstensen, 1986; but see Wasserman, 1984). This phenomenon—dubbed *Werther effect* in reference to Goethe's 1774 novel *Die Leiden des jungen Werther*, which prompted a series of self-killings upon its original publication (Phillips, 1974)—is a dramatic example of the impact of a person's perception of viable behavior on his or her choices (Cialdini & Trost, 1998). Such influences are, however, not necessarily as devastating as the Werther effect. For example, awareness of how much energy other people save or money they give to charities greatly affects our own energy-saving and charitable donations (Allcott, 2011; Croson, Handy, & Shang, 2009; Schultz, Nolan, Cialdini, Goldstein, & Griskevicius, 2007). In the movie-theater example mentioned above, confirming one's vague impression of danger by examining other people's behavior could help the individual survive by eventually following through on an intuitive response. In essence, relying on others when validating one's assumptions of reality is not reprehensible per se; instead, it mostly represents a well-adapted strategy to evaluate the appropriateness of one's view of the complexities of the world (Claidière & Whiten, 2012).

1.1 Experimental Demonstrations of Social Influence

Both scientific assessment and individual experience tell us that a person's behavior is hardly the plain effigy of an independent processor's covert workings. Rather, our attitudes, judgments, and actions are thoroughly molded by the intricate context in which they occur. At the beginning of the 20th century, Wilhelm Wundt, regarded by many as the founder of academic psychology (Blumenthal, 1975; Fahrenberg, 2012; Jüttemann, 2007; Titchener, 1921), was among the first to acknowledge the need for a scientific exploration of the complex social forces that act upon the individual (Farr, 1983; Greenwood, 2003). And to this day, it is precisely the premise of social psychology that the interpersonal context, real or imagined, largely determines our actions and cognitions (Allport, 1959; Gergen, 1985; Mischel, 2004).

Although social influence often occurs in plain sight and to some might have an air of triviality, dexterous experimentation is still needed to formally describe and, eventually, fully understand it. In fact, some portrayals of this seemingly mundane phenomenon have reverberated strongly throughout society, well beyond the realms of academia. One of these studies was Stanley Milgram's (1963) demonstration of how individuals obeyed the orders of a blunt authority and ostensibly punished innocent others. Another one was the rousing depiction of questionable norms and practices in a simulated prison at Stanford University (Haney, Banks, & Zimbardo, 1973). The fact that both studies have repeatedly been the subject of feature films—with the latest movie on each study released as recently as 2015—is testament to the stir these studies have caused.

Given their medial appeal and the luridness of their secondary dissemination, Milgram and Zimbardo's studies (Haney et al., 1973; Milgram, 1963) have left an undeniable imprint on laypersons' understanding of social influence. Although the reception and interpretation of both studies remain controversial, they are widely regarded as documenting that the alignment of one's actions or beliefs to those of others is equivalent to the subjugation of an individual's will and reason (Haslam & Reicher, 2012). What tends to be overlooked, however, is that social influence is by no means as simple as that; it is neither as distinctly condemnable, nor as easily explained as one might assume.

In fact, two research programs that were far less dazzling formed the foundation of the scientific study of social influence. Decades before the abovementioned research on obedience, Muzafer Sherif investigated in a series of laboratory studies the formation of group norms and how they shaped what the members of the group appeared to perceive (Sherif, 1935). Twenty years later, Solomon E. Asch likewise studied the influence of a group during a perceptual, albeit considerably easier task (Asch, 1951, 1955, 1956). Together with their subsequent consolidation in a unified framework on the motivational basis of conformity (Deutsch & Gerard, 1955; see also Campbell & Fairey, 1989), these two experimental approaches marked the true beginning of the science of social influence. Ever since, many questions have been asked and a myriad of findings as well as a handful of theoretical accounts have been put forth to answer them (Cialdini & Goldstein, 2004). By that, we have gained a detailed and refined understanding of how others affect ourselves.

Just as academic psychology as a whole has undergone a series of transitions in the last century, moving from early psychophysics to behaviorism to cognitivism to neuroscience (Shiraev, 2014), the study of social influence has been dominated by varying currents. Early research focused largely on demonstrating the sheer existence of conformal behavior (Asch, 1955; Sherif, 1935) and explaining its motivational underpinnings (Deutsch & Gerard, 1955). In essence, scholars sought to clarify *why* people conformed to others: What were their motives, their goals in doing so? Slightly later, research began to shift from motivation to mechanisms, as conformity was explained in comprehensive frame-works on dual modes of cognitive processing (Axsom, Yates, & Chaiken, 1987; Chaiken, 1980; Chen, Shechter, & Chaiken, 1996). The question that was now being asked was whether a certain cognitive style triggered the heuristic reliance on other people's behavior. And finally, only little more than a decade ago, researchers began studying the neural principles of social influence (Berns et al., 2005). Since then, several investigations have tried to elucidate which brain networks and neural computations are involved in this type of behavior (see Stallen & Sanfey, 2015).

Over the decades, the science of social influence has evolved into a broad and diverse field of research. Obviously, several different phenomena can be subsumed under this term, such as compliance with requests, obedience to authorities, and conformity to norms (Cialdini & Goldstein, 2004). They all share the idea that a relevant factor in individual behavior is the surrounding social forces. What they differ in, essentially, is the specific type of force: While some forces are clearly purposeful (orders and requests), others accrue somewhat incidentally within groups (norms). It is the latter form of social influence that follows the tradition of Asch (1956) and Sherif (1935), who already focused on the assimilation to others in the absence of an explicit demand for such adjustment. And it is precisely this type of influence that I have studied in the context of the present thesis.

1.2 Aims of the Present Thesis

The aims of the present thesis are to give an overview of how previous work on the neurocognition of social influence can be synthesized and to introduce a series of novel contributions to this field. My colleagues and I have used psychophysiological and behavioral experiments to study how basic mental operations are involved in this phenomenon. After introducing the general issues at hand and providing a common ground regarding both contents and relevant methods, I will summarize pertinent studies in this field, integrate them, and directly expand their interpretation on the basis of my own investigations. Although my own experimental work is primarily based on an electrophysiological approach, this thesis complies with recent demands for a psychologically grounded social neuroscience (Amodio, 2010a; Bartholow, 2010) and is intended to inform both classical psychological and contemporary neurocognitive theory.

At the heart of the present thesis is a straightforward scenario: an individual making judgments shortly after or before learning about the judgment of others. As will be explained in more detail later, this captures the basic configuration of every-day decision-making in a social context (Rilling & Sanfey, 2011) and follows an old tradition of laboratory research on group influence (Asch, 1955; Crutchfield, 1955; Deutsch & Gerard, 1955; Sherif, 1935). Moreover, the simplicity and the tight structure of this scenario allow us to track not only the behavioral consequences of social influence, but also the underlying mental processes. In several experimental studies, we recorded electroencephalography (EEG), from which we derived event-related brain potentials (ERPs). As certain deflections of the ERP are well-established signatures of specific cognitive functions (Amodio, Bartholow, & Ito, 2013), we thus gained inside into the processes that are involved in situations of social influence.

1.3 Outline of the Synopsis

The core of the present cumulative thesis is a series of eight journal articles, written and published over the course of the last three years. This includes one review article, one paper based on a methodological behavioral study, and six papers in which we report novel ERP studies. The publications will be introduced, discussed, and integrated into the existing literature in the present synopsis. It should be noted that this text is not intended to fully substitute all of the articles that it connects. Each article comprises plenty of experimental details, intricate analyses, explorations, as well as both skeptical and complaisant evaluations of the research presented therein. This level of detail and depth will not be achieved in the synopsis. Rather, I aim to carve out the interrelation as well as the cumulative and incremental contributions of my research. Therefore, the central ideas and key findings of all studies will be laid out, yet some details will have to be left out. Also, I will not divide this synopsis into two entirely separate sections, one dedicated to previous studies and one covering my own work. Instead, I will derive general conclusions based on the synthesis of my own and other scholars' findings both during and after the presentation of my research. Even though this synopsis will suffice to understand all of the studies upon which the present thesis is based, the interested reader is referred to the eight original publications to understand and evaluate the research therein even more profoundly (for an overview of the papers, see Table 1).

In *Publication I* (Schnuerch & Gibbons, 2014), we summarized and integrated several publications on the cognitive and neurophysiological foundations of social conformity. As neuroscientific research on social influence has only emerged over the last decade, much of it is only vaguely connected and still represents a somewhat heuristic exploration of possible neural correlates and principles. It was our aim to structure the work published to that date and distill the mechanisms that are recognized across different studies.

As our synthesis of the literature showed that several studies point toward the possibility that the representation of to-be-judged information plays a key role in social conformity, we employed ERPs in *Publication II* (Schnuerch, Koppehele-Gossel, & Gibbons, 2015) to measure the depth of stimulus processing. By this, we were able to support the claim that superficial, as compared to elaborate, processing of what is currently being judged is associated with greater conformity.

In the majority of my studies, I investigated how we evaluate the appropriateness of our own behavior in light of social norms, focusing on the role of attentional processes in norm-based social influence. In addition to exploring the neurocognition of social influence, I specifically used novel experimental data to address the validity and functional significance of behavioral and neurophysiological indices of how we process norms in comparison to our own judgments. In *Publication III* (Schnuerch, Schnuerch, & Gibbons, 2015), we demonstrated how an intuitive assessment of the corrective effects of social norms on individual behavior is strongly inflated by the unwanted contribution of regression toward the mean. Also, we proposed a straightforward mathematical procedure based upon a hierarchical linear model to correct for this inflation. In *Publication IV* (Gibbons, Schnuerch, & Stahl, 2016), we investigated the nature and mediators of the feedback negativity (FN), a well-established brain potential related to outcome evaluation (Miltner, Braun, & Coles, 1997) that is relevant to reinforcement learning in social contexts as well. Similarly, *Publication V* (Furley, Schnuerch, & Gibbons, 2016) highlights the role of the late positive complex (LPC) as an index of sustained, motivated attention toward relevant stimuli (Liu, Huang, McGinnis-Deweese, Keil, & Ding, 2012; Sutton & Ruchkin, 1984), particularly in regard to basic social and affective features.

To specifically test the role of outcome processing and attentional selection in social influence, three studies with varying paradigms were run that yielded converging and complementary findings. In *Publication VI* (Schnuerch, Trautmann-Lengsfeld, Bertram, & Gibbons, 2014), we studied neural signatures of social proof and deviance, derived from the comparison of one's own and other people's judgment. Specifically, we focused on the FN, probing and extending Festinger's (1954) prediction that the experience of social deviance can lead to the evasion of the respective reference group. In *Publication VII* (Schnuerch & Gibbons, 2015), we replicated our own and other scholars' previous findings on the neurocognitive foundations of social proof in a novel experimental paradigm. Also, we found evidence of the involvement of attentional-selection processes and of sustained effects of social (dis)agreement on stimulus processing. In *Publication VIII* (Schnuerch, Richter, Koppehele-Gossel, & Gibbons, 2016), we show that similar neurocognitive processes are involved under more ecologically valid conditions.

Table 1: Publications included in the present thesis

Publication	Full Reference
I	**Schnuerch**, R., & Gibbons, H. (2014). A review of neurocognitive mechanisms of social conformity. *Social Psychology, 6*, 466–478.
II	**Schnuerch**, R., Koppehele-Gossel, J., & Gibbons, H. (2015). Weak encoding of faces predicts socially influenced judgments of facial attractiveness. *Social Neuroscience, 10*, 624–634.
III	**Schnuerch**, R., Schnuerch, M., & Gibbons, H. (2015). Assessing and correcting for regression toward the mean in deviance-induced social conformity. *Frontiers in Psychology, 6*, 669.
IV	Gibbons, H., **Schnuerch**, R., & Stahl, J. (2016). From positivity to negativity bias: Ambiguity affects the neurophysiological signatures of feedback processing. *Journal of Cognitive Neuroscience, 28*, 542–557.
V	Furley*, P., **Schnuerch***, R., & Gibbons, H. (2016). The winner takes it all: Event-related brain potentials reveal enhanced motivated attention toward athletes' nonverbal signals of leading. *Social Neuroscience.* Advance online publication. [* joint first authorship]
VI	**Schnuerch**, R., Trautmann-Lengsfeld, S. A., Bertram, M., & Gibbons, H. (2014). Neural sensitivity to social deviance predicts attentive processing of peer-group judgment. *Social Neuroscience, 9*, 650–660.
VII	**Schnuerch**, R., & Gibbons, H. (2015). Social proof in the human brain: Electrophysiological signatures of agreement and disagreement with the majority. *Psychophysiology, 52*, 1328–1342.
VIII	**Schnuerch**, R., Richter, J., Koppehele-Gossel, J., & Gibbons, H. (2016). Multiple neural signatures of social proof and deviance during the observation of other people's preferences. *Psychophysiology, 53*, 823–836.

Note. Publications are sorted based on their interrelation and the order of studies as presented in the main text, rather than by the date of publication.

2 Social Influence

Social influence is a rich and complex aspect of human behavior and, accordingly, an equally widespread field of research. As mentioned above, there are several phenomena and processes that can be subsumed under this term. However, neither the term itself nor the labels for the variety of phenomena it comprises are used consistently throughout the literature. There is considerable variation in what scholars understand by various forms of social influence (Cialdini & Trost, 1998). It is therefore necessary to provide a rough overview and establish a certain terminological and definitional basis for the work laid out in the present thesis.

A frequently cited taxonomy is the notion that social influence comprises two types of behavior: conformity and compliance (Cialdini & Goldstein, 2004). This choice of terms is slightly confusing, as one subtype of conformity is also called compliance (Maass & Clark, 1983). Moreover, there is some inconsistency in the literature as to how conformity and compliance relate to the concept of social norms (see, e.g., Cialdini & Trost, 1998; Fehr & Fischbacher, 2004; Kallgren, Reno, & Cialdini, 2000). Perhaps the most comprehensive and parsimonious overview of the functional and terminological relation between norms and socially adjusted behavior was proposed only very recently (McDonald & Crandall, 2015). Mainly—although not exclusively—based upon this approach, I will define the relevant concepts in the following.

2.1 Compliance

Undoubtedly, compliance is one of the most relevant and most commonly discussed manifestations of social influence. A typical definition is that compliance is an acquiescent adjustment of individual behavior in response to a request (Cialdini & Trost, 1998). Requests need not include an explicit appeal, such as when adverts emphasize the good qualities of a product without literally asking the observer to buy it (see Cialdini & Goldstein, 2004). However, a defining feature of compliance is that there is an *explicit intent* behind the process of social influence; that is, the individual's adjustment is the intention, usually of a single person or entity, such as a company or a team. In this regard, it clearly differs from the much more incidental influence of social norms, which neither requires, nor typically involves, someone's overt intention to influence the individual (see, e.g., Asch, 1955).

Although the conceptualization of compliance as a separate facet of social influence is well-established (Cialdini & Goldstein, 2004), it hardly works as distinctly as common modular models suggest. For example, although compliance with a request can clearly be differentiated from conformity to social norms from a procedural view (Cialdini & Goldstein, 2004), norms are actually frequently used as part of strategies aimed at inducing compliance (Jacobson, Mortensen, Jacobson, & Cialdini, 2015). Activating a concept of how most people act (a descriptive norm) is an effective tool in persuading us to follow an appeal (Goldstein, Cialdini, & Griskevicius, 2008; Schultz et al., 2007). Typical claims such as "this year's best-selling brand" or "up to two thirds of our customers opt for energy-efficient heating" are based upon this very concept (Cialdini, 2007).

A particularly striking example of compliance is individual adjustment induced by authoritative orders. This notion relies on a striking demonstration of obedience in Milgram's highly publicized Yale studies (Milgram, 1963, 1965a, 1965b). Milgram expressly aimed to go beyond the exploration of conformity, on which Asch, one of his mentors and supervisors, focused and which Milgram himself had studied as well (Milgram, 1961). As documented by Russel (2011), Milgram was adamant about finding out whether the force of others would also be measurable once the individual was asked to make *relevant* decisions under social pressure—rather than completely inconsequential perceptual judgments, as typically investigated in Asch's experiments (Asch, 1951, 1955, 1956). Moreover, Milgram aimed to clarify, in a random American sample, whether individuals would indeed fully succumb to authoritative orders. As we know today, this proved to be the case, which inspired a vivid empirical and theoretical discourse (see, e.g., Blass, 1991, 1999, 2009; Burger, 2009; Reicher & Haslam, 2011; Slater et al., 2006).

Research on compliance is not only perplexing, but it also holds evident political relevance. Among other things, it might help us appreciate the role of authoritarian hierarchies in molding individual behavior (Milgram, 1965b; Russell, 2011). Albeit inspired by and related to research on conformity (Haslam & Reicher, 2012), compliant behavior, such as obedience to authorities, represents a separate, perhaps even more complex facet of the social nature of human conduct. When investigating conformity, as in the present thesis, it thus seems ill-advised to draw seemingly parallel conclusions regarding compliance. Also, most studies on the neurocognition of social influence have focused on conformity to group norms (see Stallen & Sanfey, 2015). Therefore, only little reference to the clearly related, yet essentially different aspect of compliance with requests will be made throughout this thesis.

2.2 Conformity

Although *conformity* is a ubiquitous, commonly-used term, precise definitions are scarce. As noted by Cialdini and Trost (1998), it is often used to describe the act of adopting other people's behavior by moving away from one's own stance. Defined this way, conformity only occurs when an individual adopts group behavior even though it contradicts his or her initial view. "Thus, we conform to others when perceived or real pressure from them causes us to act differently from how we would act if alone [...]" (Cialdini & Trost, 1998, p. 162). This definition demarcates a firm border between conformity and normative behavior, as norms do not necessarily cause a shift in the individual's behavior and need not contradict a person's own view.

There is, however, good reason to forego such a strict definition of conformity. First, the narrow definition might simply be a convention based upon the early use of the word, particularly in Asch's well-received publications (1951, 1955, 1956). Second, there are several authors who define the concept more broadly, describing even the adjustment to group behavior *under uncertainty*—when there is no apparent contradiction between a person's stance and a group's aberrant view—as conformity (Insko, Smith, Alicke, Wade, & Taylor, 1985; Walther et al., 2002; see also Cialdini & Trost, 1998). Only recently, the term has been used to simply describe that an individual chooses a certain behavior because it is the most frequent in others (Claidière & Whiten, 2012). Third, and most importantly, it seems parsimonious to apply an inclusive definition that allows us to link, rather than demarcate, our conceptions of norms and conformity. Conformity might thus be defined simply as a person's adjustment to a social norm (McDonald & Crandall, 2015). Throughout the present thesis, I adhere to this definition.

One central question regarding the nature of conformity is what motives lead an individual to adopt the group's behavior. Based on the premise that our behavior in a social context is goal-oriented and serves individual purposes (Cialdini & Trost, 1998), two central motives for conformity have been identified (Asch, 1955; Campbell & Fairey, 1989; Deutsch & Gerard, 1955), both of which serve a superordinate, somewhat general quest (Cialdini & Trost, 1998; Wood, 2000). The first basic goal is to be accurate in our judgments and in our take on reality; the second basic goal is to be affiliated and socially accepted; the superordinate goal is to maintain a positive self-concept by establishing and confirming that we are accurate and affiliated (Cialdini & Goldstein, 2004). The processes of social impact on individual behavior based on the two basic motives, accuracy and affiliation, have been labeled informational and normative influence, respectively.

2.2.1 Informational Influence

We often conform to others because we use them as a reliable source of information that allows us to enhance the quality of our own judgments or even our own representation of reality (Campbell & Fairey, 1989; Deutsch & Gerard, 1955; Insko et al., 1985). In Sherif's (1935) seminal experiments, participants sat in the dark and were asked to indicate how much a small dot that was projected onto the wall moved around. In fact, the dot did not move at all, yet one automatically gets the impression of movement after some time (the so-called *autokinetic effect*). Not much surprisingly, this task is extremely difficult, as it is virtually impossible to assess how wide the corridor of movement is. In this uncertain situation, it is helpful to either copy other people's judgment or at least to verify whether one's vague impression of the size of the apparent movement is supported by others. To this end, we compare our own and other people's judgments. If we experience a strong deviation from the most prevalent response, we will be tempted to adjust our own judgment in the group's direction or even to adopt its judgment—not just in our public response, but in our own private understanding as well. Therefore, informational influence can entail the private acceptance of the majority's judgment, implying an updating (conversion) of the individual's own actual views (Kelman, 1961).

2.2.2 Normative Influence

Perhaps the more striking demonstrations of group influence—which originally coined the term conformity in social psychology—involved an adjustment of individual behavior to an obviously incorrect group against better judgment, such as in Asch's seminal line-judgment studies (Asch, 1951, 1955, 1956). In these experiments, participants were asked to indicate which of three comparison lines was shown on a card displaying one single line. Importantly, the lines were noticeably different, and control participants who completed this task alone performed near perfection, with a mean error rate of 1% (Friend, Rafferty, & Bramel, 1990). Participants who performed the task in a group, which actually consisted of confederates of the experimenter whose answers had been scripted beforehand, gave 33% incorrect responses on average, as they followed the majority's openly erroneous responses. As revealed in subsequent interviews, and later supported by independent studies and targeted manipulations, a typical motive for adjustment in the "Asch task" was to avoid standing out by deviating from the majority's expectations (Deutsch &

Gerard, 1955). Usually, such normative influence results only in public compliance with the group, without a true adoption of majority responses as correct assessments of reality (Cialdini & Goldstein, 2004; Cialdini & Trost, 1998).

2.3 Social Norms

The concept of social norms has long been controversial and its power in explaining human behavior has been challenged (Cialdini et al., 1990). Even today, some researchers argue that our understanding of norms is vague at best (Fehr & Fischbacher, 2004; Krupka & Weber, 2013). Although the latter view—potentially triggered by an eclectic focus on norms in *economic* behavior—is disputable, there is no denying the need for novel research to further what we know in this field (see, e.g., Hogg & Reid, 2006).

McDonald and Crandall (2015) recently defined social norms as “expectation[s] about appropriate behavior that occurs in a group context” (p. 147). It should be noted that the key concept here is “appropriate behavior”, which is somewhat inclusive and covers different forms of normative standards. According to the widely adopted Focus Theory of Normative Conduct (Cialdini et al., 1990; Kallgren et al., 2000), two types of norms can be differentiated. *Descriptive norms* simply denote what is most common within a social group (Cialdini & Trost, 1998; McDonald & Crandall, 2015). That is, they convey the most prevalent (i.e., the majority’s) behavior (Cialdini et al., 1990). *Injunctive norms* are not merely descriptive, but rather prescriptive—in the sense of conveying what ought to be done or what is approved of in a given group (McDonald & Crandall, 2015). The differentiation of the two types is not purely an effort to arrange our concept of norms. Rather, it reflects profound variations in their nature.

2.3.1 Descriptive Norms

Descriptive norms provide helpful guidance as to which type of behavior is likely correct in a certain situation. Conversely, the process of following descriptive norms is conformity based on informational social influence (McDonald & Crandall, 2015). Whenever we are insecure about how we should act, the behavior that is most common within a group can be interpreted as the best guess at what is currently appropriate (Cialdini & Goldstein, 2004). Such reliance on the majority under uncertainty is particularly pronounced when correct decisions are rewarded or urgently needed (**coleman1958**; Baron, Vandello, & Brunsman, 1996). Therefore, abiding by descriptive norms can be described as a heuristic recourse to the simple rule of thumb that “consensus implies correctness” (Axsom et al., 1987). As the informational value of descriptive norms is often more than a simple instruction as to how one should behave to act appropriately, such norms often help us validate and adjust our own view of the world (Deutsch & Gerard, 1955).

2.3.2 Injunctive Norms

The basis of injunctive norms is the expectation of the group (Cialdini & Trost, 1998). Often, there are more or less explicitly voiced, sometimes completely arbitrary expectations regarding the behavior of individuals within a group. To gain or maintain the approval of others, it is thus necessary to abide by these rules. This usually happens without conversion of the individual’s beliefs; rather, he or she merely acquiesces to the majority’s judgment or behavior (Kelman, 1961). Thus, following injunctive norms is what is usually described as conformity due to the normative influence of the group (McDonald & Crandall, 2015). Such abidance by norms can likewise be described as a heuristic strategy; this one, however, is based on the—presumably reasonable—guideline “go along to get along” (Chen et al., 1996).

Some theorists argue for a third type of norm: subjective norms (Fishbein & Ajzen, 1975). They represent a person’s assumption about how closely related parties, such as family and friends, expect him or her to behave (Cialdini & Trost, 1998). Therefore, subjective norms could be regarded as an extension or qualification of the concept of injunctive norms, reflecting not the *objective* expectations of others, but a person’s subjective representation of what a group approves of. This refinement of the typology of norms is, however, not commonly adopted in contemporary theories (see, e.g., McDonald & Crandall, 2015).

2.4 Social Proof and Deviance

An essential part of the process by which we get influenced by social groups is the comparison between our own and other people’s behavior. In a groundbreaking theoretical paper, Leon Festinger (1954) proposed that social comparison is a core process of human cognition. This is consistent with the assertion that—well beyond the social domain—all stimulation is processed in relation to a normative standard (Kahneman & Miller, 1986). The general reliance on comparative processing appears to be a well-adapted strategy that does not only afford us the opportunity to gain valuable knowledge of ourselves

(Mussweiler & Rüter, 2003), but also makes us more certain (Mussweiler & Posten, 2012) and more efficient (Mussweiler & Epstude, 2009) in our judgments.

The central tenet of Festinger's (1954) Theory of Social Comparison Processes is that we constantly try to evaluate our own "opinions and abilities" (Festinger, 1954, p. 117). Whenever no objective reference is available, we consult other people's behavior to assess our own accuracy, thus trying to find *social proof* for our views (Cialdini, Wosinska, Barrett, Butner, & Gornik-Durose, 1999). In case we find proof, we maintain our previously held opinion. If, however, we register our own *social deviance*, we usually correct our judgments, such that they match the group's or, at least, lie closer to them. The farther we deviate from others, the stronger will be our tendency to correct ourselves (Festinger, 1954). Alternatively, we might attempt to change the group's views rather than shift our own, which does, however, depend on our means to actually affect the course of the group (Moscovici, Lage, & Naffrechoux, 1969). Interestingly, a different approach to deal with the ongoing perception of social deviance is to simply terminate or avoid the comparison of our own and certain other people's behavior (Festinger, 1954).

The human tendency to assert social proof or deviance should be a particularly pivotal mechanism of conformity driven by the *informational* value of *descriptive* social norms. In uncertain situations, comparative thinking allows us to reduce the level of uncertainty by using the judgments of the majority (i.e., the descriptive norm) as a point of reference and incorporating them into our own. A straightforward example of this process is the formation of judgments in Sherif's dot-motion experiments (Sherif, 1935). The individual only has a vague impression of the degree of motion, which he or she then obviously compares with other people's judgments, leading to a steady, gradual approximation of the individual's responses to the group's or—in case of two respondents—to each other (Sherif, 1935).

Most of the theories that specifically aim to explain social influence (see, e.g., Campbell & Fairey, 1989; Cialdini & Goldstein, 2004; Deutsch & Gerard, 1955; Latané, 1981; Tanford & Penrod, 1984) focus on what might be called macro-level processes, primarily the motivational underpinnings of socially adjusted behavior (for a similar discussion, see Germar, Schlemmer, Krug, Voss, & Mojzisch, 2014). To give an example: The notion that our general goal of maintaining a favorable self-concept drives our pursuit of accuracy and affiliation in social decision-making (Cialdini & Trost, 1998) does imply the contribution of certain cognitions (Cialdini & Goldstein, 2004). What it does not encompass, though, is a proposal of specific components of cognitive processing that are involved in social influence. Social-comparison theory (Festinger, 1954), however, basically premises the existence of a somewhat universal component of social cognition that is *essential* for social influence. One might say that it thus provides the groundwork for our scientific assessment and subsequent understanding of the neurocognitive principles of social influence. Not surprisingly, much of the recent research on the physiological, cognitive, and neurocomputational foundations of social influence has focused on the role of comparison processes (see, e.g., Chen, Zhong, et al., 2012; Izuma & Adolphs, 2013; Kim, Liss, Rao, Singer, & Compton, 2012; Klucharev, Hytönen, Rijpkema, Smidts, & Fernández, 2009; Klucharev, Munneke, Smidts, & Fernández, 2011; Shestakova et al., 2013)

Although the assumed comparison mechanism is strikingly simple, it seems expedient to explicitly deconstruct from a modern-day cognitive perspective what Festinger's (1954) theory—mostly devised in broader, socioscientific terms—actually implies. The corrective impact of the descriptive norm first requires the mental availability of one's own view (or judgment) as well as the norm. Based upon those pieces of information, the cognitive system must afford an internal juxtaposition of both views. Finally, this comparison must be translated into an evaluation, that is, into a score representing the quality (e.g., direction) and quantity (i.e., the degree) of one's deviation from or agreement with the group. This, then, ought to feed both into subsequent cognitive processing and into the individual's judgment (Festinger, 1954).

3 Methodology

Before turning to the central part of the synopsis, covering the neurocognitive principles of social influence, it is necessary to establish a firm methodological ground for the upcoming sections. As noted earlier, most of my own research is based on the assessment and analysis of ERPs, as are several other recent studies on social influence (Chen, Zhong, et al., 2012; Kim et al., 2012; Kimura, Murayama, Miura, & Katayama, 2013; Shestakova et al., 2013; Trautmann-Lengsfeld & Herrmann, 2013, 2014; Yu & Sun, 2013). Conversely, I will provide a brief introduction into the neurophysiological rationale and technique of EEG and ERPs. However, as in cognitive neuroscience in general, functional magnetic resonance imaging (fMRI) has become the dominant method in social neuroscience (see Amodio et al., 2013; Singer, 2012). Not much surprisingly, a significant portion of studies into the neurocognition of social influence are based upon fMRI (for an overview, see Izuma, 2013; Stallen & Sanfey, 2015). Therefore, I shall additionally provide a very brief introduction into this technique.

3.1 Event-Related Brain Potentials (ERPs)

ERPs represent a well-established and widely-used method in cognitive neuroscience (see, e.g., Coles, 1989; Fabiani, Gratton, & Coles, 2000; Luck, 2005; Shah et al., 2004). One of the defining features of this technique is that brain activity is measured *directly* (Amodio et al., 2013), distinguishing it from fMRI, which is based upon an indirect inference of neural activity (Bartholow & Amodio, 2009; Logothetis, 2002). ERPs have a remarkable temporal resolution, allowing us to track with millisecond precision how specific neural processes evolve over a certain period of time (Fabiani et al., 2000). Crucially, a vast body of previous research has established the functional relation between cognitive processes and ERP signatures, enabling us to infer otherwise concealed mental processes based on the shape and timing of ERPs (Amodio, 2010a). They are therefore deemed an invaluable tool in monitoring cognitive processing in the human brain in real-time (Amodio et al., 2013; Coles, 1989).

A notable shortcoming of ERPs is their low spatial resolution. As voltage fluctuations registered at the scalp are potentially explained by an infinite number of different combinations of neural generators (Luck, 2005), the source of brain activity measured by ERPs remains uncertain (Fabiani et al., 2000). This is typically referred to as the *inverse problem* (see Darvas, Pantazis, Kucukaltun-Yildirim, & Leahy, 2004). According to Luck (2005), it is currently impossible to assess the margin of error of an ERP-source estimation, such that “the spatial resolution of the ERP technique is fundamentally undefined” (p. 25). This stands in stark contrast to fMRI, whose spatial resolution is not only known, but often as high as in the range of a few millimeters (Norman, Polyn, Detre, & Haxby, 2006; Yoo, Talos, Golby, Black, & Panych, 2004). It should be noted, however, that several techniques, particularly mathematical routines to estimate potential electrical dipole sources, have been devised to address the problem of generator localization in ERPs (see, e.g., Luck, 2005; Pascual-Marqui, 2002; Wagner, Fuchs, & Kastner, 2004).

3.1.1 Neurophysiological Background

The basis of ERPs is the EEG, a continuous recording of voltages, emanating from the brain, on the surface of the scalp. Building upon previous work in animals, Hans Berger famously conducted the first human EEG at the University of Jena in 1929 (Berger, 1929; Haas, 2003). Whereas psychiatric and psychological research using EEG originally focused on spontaneously occurring oscillations of voltage, scholars later began to assess variations in brain activity induced by specific events: ERPs (see Fabiani et al., 2000). To extract ERPs from EEG data, a large number of certain epochs taken from the continuous recording is averaged. Each epoch represents the overall neural response recorded during and directly after the occurrence of the event of interest (e.g., a certain stimulus class). Averaging several repetitions of the same event *cancels out* the random noise generated by the unsystematic, continuous brain activity measured in the EEG. The systematic pattern of activity specifically induced by the crucial event, however, is selectively *preserved* during this procedure (Luck, 2005). The electrophysiological source of an ERP is activity in a certain area (or several areas) of the brain. However, the relation between the brain’s most fundamental currency—the firing of neurons—and the appearance of the ERP is somewhat indirect, as Luck (2005) asserts in his seminal volume, upon which the following simplified description is primarily based.

When a neuron receives input from another, usually via a neurochemical transmitter, ion channels in the postsynaptic membrane close or open. This often results in a postsynaptic potential, that is, a change of the flow and balance of ions along the membrane. Several postsynaptic potentials usually occur simultaneously at several points of synaptic contact;

for example, at various apical dendrites of a neuron. The summation of these potentials then leads to a change of voltage in the broader area of the postsynaptic neuron where the neurotransmitter made contact with it. In the case of excitatory potentials, the charge around the apical synapses, where the postsynaptic potential develops, is negative. At the opposing end of the neuron, a compensatory flow of ions occurs, leading to a positive charge. Consequently, the synaptically stimulated, apical and the non-activated, proximal part of the neuron have opposing electrical charges. This relative difference gives rise to an electrical dipole along the neuron (i.e., an arrangement of two neighboring, yet spatially separated positive and negative charges). If thousands or millions of roughly parallel neurons receive the same input simultaneously, an array of dipoles, each occurring based on the distribution of charges along a neuron, is created. This results in what might be termed an average or macro-dipole, a so-called *equivalent current dipole* (ECD). Such dipoles and the resultant ECDs are particularly likely to occur in pyramidal cells, aligned almost parallel and placed in a perpendicular fashion to the surface of the cortex. On the scalp, located parallel to the majority of the gyri of the cortex, ECDs are measured as voltage fluctuations in the EEG and, based thereupon, the ERP. As the conduction of voltages through brain, skull, and scalp occurs at only slightly less than the speed of light, the ERP tracks ECD-based neural signals almost in real-time.

Whether the *polarity* of an EEG or ERP signal is indicative of a certain class of neural operations is subject to debate. One reason why negativities and positivities in the EEG might indeed reflect two distinct neurophysiological processes is that the polarity of EEG recordings depends on *where* the synaptic transmission responsible for the dipoles and ECD occurred (Elbert & Rockstroh, 1987). Depolarization of the postsynaptic membrane close to the skull elicits negativity in skull-surface measurements, as the negative pole of the dipole is above the positive pole. Depolarization farther from the surface (i.e., deeper within the cerebral tissue) leads to positivity in the EEG, as the area where the postsynaptic potentials occur, leading to a regional negativity, is below the opposing inactive end of the neuron where the relative positivity of the dipole develops. Crucially, the location of subcortical afferents might be related to the functional role of their signals. Synaptic input at the distant, apical dendrites of cortical pyramidal neurons presumably increases the excitability of these neurons, whereas proximal synapses close to the somae of these neurons usually facilitate a decrease in their overall excitability (Elbert & Rockstroh, 1987). Thus, negative EEG signals, based on postsynaptic potentials close to the skull, could indicate excitatory processes (i.e., an increase in neural activity in the generator regions), whereas positive EEG signals, based on activity near the somae of cortical neurons, should indicate the inhibition of activity in pyramidal neuron assemblies. It should be noted, however, that this view has become secondary to a more detailed understanding of ERP components that is based on an extant literature allowing us to assign specific psychological processes to a certain ERP signature, almost irrespective of the nature of the precise cellular mechanisms of its emergence (see Amodio et al., 2013; Fabiani et al., 2000; Luck, 2005).

3.1.2 ERP Components

3.1.2.1 General Features

An ERP is usually assessed and depicted relative to the onset of the event that induces the fluctuations of cortical dipoles. Its appearance is characterized by a series of systematic deflections (components) with a certain polarity (negative vs. positive) and latency (i.e., the point in time, relative to the event, at which the deflection peaks). All key parameters of the ERP and its components are relative, not absolute values. First, an ERP generally depicts the *difference* between activity measured at a certain site and a reference site; a typical point of reference is the averaged signal at both mastoid processes, where the same degree of unsystematic noise as at scalp electrodes, yet only very little actual neural activity is measured (Luck, 2005). Second, the voltage of the ERP (e.g., the amplitude of a given component) always refers to the change in voltage relative to a predefined baseline phase. Often, a short period preceding the event is used to assess the baseline level of activity (e.g., 100 ms before the onset of a stimulus). Third, the polarity of a component refers to its voltage relative to preceding components: If the overall ERP shifts and all voltages are negative, a deflection in the positive direction is usually interpreted as a positivity, even if its voltage is not actually above zero.

The labels assigned to ERP components are often based on their polarity and latency. As summarized by Fabiani and colleagues (2000), though, there are various other habits of labelling, such as applying an ordinal number for a deflection of a certain polarity (e.g., N1, P2, P3), describing the functional origin of a component (e.g., mismatch negativity, novelty-P3) or naming its approximate location (such as P2a or frontal P3). Notably, the labels of many components are not used consistently, such that minor (e.g., P3 vs. P300) or even major (e.g., P2a vs. P3f vs. FP vs. FSP) discrepancies occur (for examples, see Bloom, Potts, Evans, & Drobes, 2013; Potts, 2004; Potts, Martin, Burton, & Montague, 2006; Sutton & Ruchkin, 1984).

In many depictions and discussions of ERPs, researchers also take the spatial distribution of components across the scalp into consideration. This topography often has a bearing on the interpretation of ERP findings, since many components have a typical or standard focus. For example, the face-related N170 (Bentin, Allison, Puce, Perez, & McCarthy, 1996) is usually observed to be more strongly pronounced at right-hemispheric lateral posterior sites than at the analogous left-side electrodes (see, e.g., Eimer, 2000; Itier & Taylor, 2004; Rossion & Jacques, 2008). Moreover, some components that appear very similar according to their main features (polarity and latency) can be distinguished as separate, potentially independent signatures on the basis of their topographies (Luck, 2005). For example, the well-known P3 refers to at least two functionally different components, one of which has a posterior maximum, while the other one peaks at frontocentral electrodes (Polich, 2007; Polich & Criado, 2006; see also Sutton & Ruchkin, 1984). Finally, topography is often reported for ERP *effects*, rather than or in addition to the components themselves. This way, one can determine the scalp position at which an experimental manipulation, or whichever independent variable is currently considered, exerted its maximal influence (for an example, see Gibbons, Bachmann, & Stahl, 2014). It should be noted, however, that the assessment of scalp topography does not provide sufficient information as to the actual neural sources of surface discharges (Luck, 2005). Due to the inverse problem, the generators of a component need not necessarily lie in proximity to its scalp-topographical peak. For some components, such as for the N170, the source is indeed close to the location of the surface peak (Deffke et al., 2007; Latinus & Taylor, 2006). For other components, such as the auditory N1 component, there is a notable distance between cerebral sources and ERP peak location (Giard et al., 1994; Jääskeläinen et al., 2004).

3.1.2.2 Quantification

ERP components are typically quantified for all recording sites, for pre-selected electrodes that are known to depict a certain component most clearly and directly, or for clusters of electrodes. Of course, quantification is done separately for different experimental conditions, such that ERP components can be compared between these conditions. To quantify a component, three different major strategies are most frequently used. First, one can compute the mean amplitude in an appropriate time range in which the component of interest is observed (Amodio et al., 2013). This time range is sometimes determined a priori based upon previous research or, perhaps more often, chosen to fit the characteristics of a component, as assessed by visual inspection of the ERP averaged across participants (Amodio et al., 2013). For example, if a component peaks early and features a brief ascent and steep descent in amplitude, a somewhat narrow frame would be chosen (e.g., 30 ms around the early visual P1). Later, slower, and broader components of the ERP are usually assessed by means of averaging amplitudes in wider time windows (e.g., 200 ms around a late, parietal P3). This general approach of time-window amplitude averaging is used most frequently (Amodio et al., 2013) and preferred by some authors on theoretical grounds (Luck, 2005).

Second, one can assess the peak amplitude of a component, which refers to the maximum voltage in case of positivities and the minimum voltage for negativities. This is usually done automatically by means of a computer algorithm that extracts the peak voltage of a certain polarity in a certain time range, separately for different channels in each participant (Amodio et al., 2013). For example, to quantify the peak of the previously mentioned N170, one might determine the most negative amplitude within a range of 150-210 ms after the onset of a face (Ofan, Rubin, & Amodio, 2011). In the case of the N170, one might additionally restrict the search for negative peaks to lateral posterior electrodes (Ratner & Amodio, 2013). Sometimes, researchers use not only the amplitude value at the specific sampling point where the component peaks, but rather the mean amplitude at a certain number of sampling points around the peak (e.g., ±5 points). Note that a component's mean amplitude around its peak is based on the specific latency of this component at a given electrode in an individual participant. In the standard amplitude-averaging approach (as described above), an appropriate time window is determined and fixed for all participants, without consideration of interindividual variations in component latency.

Third, one can quantify a component by focusing on its latency (Amodio et al., 2013). To this end, the abovementioned peak-detection procedure is used, yet one assesses the interval between the onset of the event and the peak of a component. A typical application of this method is when the focus of an analysis is on the speed, rather than on the intensity, of neural operations (see, e.g., Eimer, 1998). Often, analyses of latency are conducted to ascertain that variations of (peak) amplitude are indeed modulations of amplitude, rather than side effects of latency differences between conditions.

3.1.2.3 Classification and Application

Many researchers discriminate between different classes of components. A common approach is to make a distinction between early (sensory or perceptual), mid-latency (cognitive, often classification-related), and late (higher cognitive) components of the ERP (Fabiani et al., 2000). According to Luck (2005), the most relevant components include C1, P1, N1, N170, P2, N2, P3, N400, error-related negativity (ERN), and the lateralized readiness potential (LRP). Several of these components, as well as additional ones not listed here, will be essential in understanding the work on the neurocognition of social influence that has been published so far, including my own contributions. However, I will not provide an extensive and comprehensive overview of all these ERP components since this would go well beyond the scope of the present thesis. Instead, each relevant component will briefly be introduced in the context of social-influence research in the upcoming main chapter. For additional details, I refer the reader to eminent reviews and volumes in this field (e.g., Amodio et al., 2013; Bartholow & Amodio, 2009; Fabiani et al., 2000; Kok, 1997; Luck, 2005; Luck & Kappenman, 2011; Rugg & Coles, 1995).

As noted earlier, there is a solid body of research on the functional significance of many ERP components, such that specific hypotheses about psychologically induced variations of their amplitude, latency, and topography can be derived a priori. Much of the previous research on common components was based on general experimental investigations that have isolated the driving forces behind a component fairly well (for various examples, see Luck & Kappenman, 2011). This is why ERPs are no longer restricted to investigations into fundamental cognitive functions, but can also be used to study more complex psychological issues. For many components, their basic determinants and their potential "meaning" under various circumstances have been established by now (Amodio et al., 2013). Therefore, one can apply this knowledge to gain valuable insight into the cognitive processes that are at the heart of sociopsychological phenomena (Amodio, 2010a; Bartholow & Amodio, 2009).

3.2 Functional Magnetic Resonance Imaging (fMRI)

A relevant portion of previous studies on the neurocognition of social influence have used fMRI (e.g., Berns, Capra, Moore, & Noussair, 2010; Berns et al., 2005; Deuker et al., 2013; Edelson, Sharot, Dolan, & Dudai, 2011; Izuma, 2013; Klucharev et al., 2009; Klucharev et al., 2011; Stallen, Smidts, & Sanfey, 2013). It is one of the most frequently employed techniques in cognitive neuroscience and at the center of a heated debate regarding the promises and pitfalls of neuroimaging in psychological research (Diener, 2010; Vul, Harris, Winkielman, & Pashler, 2009). The method is based on MRI, which is the assessment of body structure based on different magnetic responses in different types of tissue (Logothetis, 2002).

As roughly summarized by Ashby (2015), MR machines produce a strong, static magnetic field, and a special coil emits radiofrequency pulses to elicit an in-phase precession of protons, usually in hydrogen molecules, in the tissue that is being scanned. Upon the offset of the pulses, the protons return to their previous alignment induced by the static magnetic field, which triggers the release of energy that is measured by the coil as the "raw MR signal" (Ashby, 2015, p. 94). Crucially, different types of tissue (e.g., bone, gray matter, or tumors) have quite different densities, which differentially affects the proton relaxation. Therefore, the speed of the relaxation is somewhat characteristic of a certain tissue. By additionally applying three varying gradients of magnetic radiation along the three dimensions, one can attribute a specific relaxation signal registered outside the body to a specific location within (Ashby, 2015).

While MRI is a static assessment of tissue, similar to a photography or an x-ray image, albeit in 3D, the time dimension is added in fMRI. Specifically, MR images of the brain are repeatedly taken to estimate the flow of blood as a proxy for cerebral activation (Harris, Reynell, & Attwell, 2011). The crucial parameter is the blood-oxygenation level dependent (BOLD) signal (Logothetis, 2002). The oxygen carrier in our blood, hemoglobin, has different magnetic properties depending on whether it does or does not carry oxygen molecules (Pauling & Coryell, 1936), which affects the abovementioned MR signal. If an area in the brain is activated, the consumption of oxygen increases, which is met by an enhanced flow of oxygenated blood toward this region. Therefore, after an initial dip in the level of available oxygen due to the elevated regional demand, the proportion of oxygenated relative to deoxygenated blood increases close to an activated area of the brain, which triggers a positive BOLD signal (Harris et al., 2011; Logothetis, 2002). Although the association between what is supposed to be measured (brain activity) and what actually accounts for the measurement (blood-oxygenation level) appears somewhat oblique, the BOLD signal closely and reliably reflects regional fluctuations of local field potentials in nervous tissue (Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001).

The spatial resolution of fMRI is very high, especially compared to other techniques that aim to image functional properties of the brain (Amodio et al., 2013). A notable shortcoming of this method, however, is its temporal precision. The speed at which repeated images of the brain are taken is severely limited. Moreover, changes in the BOLD signal are inert, taking several seconds to peak and decaying even more slowly (Ashby, 2015). As this does not match the actual speed of many electrophysiological signals in the brain, fMRI offers only a poor account of the unfolding of neural processes over time (Luck, 2005).

4 The Neurocognition of Social Influence

What is the neural foundation of social influence? I can forestall at this point that this question will not be answered in the present thesis—just as it is not answered in the current literature on this matter. Such a question is as tempting as it is presumptuous. In fact, it is an example of the false hopes or scientific hubris that might arise from recent advances in our technological means (Diener, 2010; Vul et al., 2009). Cognitive neuroscience is neither equipped nor inclined to uncover *the* neural correlate of a complex, macro-level process such as social influence. Rather, an appropriate neurocognitive perspective on psychological processes—especially the broader, more complex ones—is to deal in smaller, more circumscribed units. Particularly in social neuroscience, the quintessential approach is to disassemble complex social behavior into its cognitive ingredients (for similar discussions, see Abe, 2011; Amodio, 2010a; Bartholow, 2010). Accordingly, we have begun to unravel small and confined, yet essential mechanisms of the human cognitive system that play crucial roles in the overall process of other people's influence on ourselves (for an overview of these studies, see *Publication I*; Falk, Way, & Jasinska, 2012; Izuma, 2013; Mojzisch & Krug, 2008; Stallen & Sanfey, 2015).

In many recent studies, the involvement of certain fundamental cognitive functions in social influence has been substantiated by studying their neurophysiological signatures. Scholars have gathered convincing evidence that the higher-level process of social influence involves modulations of lower-level processes such as perception, attention, value coding, and reward expectation. It should be noted early on, however, that the contribution of these mechanisms largely depends on the psychological scenario—as emulated by choosing certain parameters and experimental constraints—that is being studied. For example, while some of the research has dealt with the actual processing of group norms, other studies have focused on the consequences of knowing a group norm during one's subsequent assessment of the current object of judgment. Such choices of focus are strongly affected by the neurocognitive perspective on the larger issue at hand, and the divergent results should be considered complementary rather than contradictory, with their prevalence depending mostly on the specific situational demands and constraints.

Before I elaborate on previous and very recent investigations into the neurocognition of social influence, common conceptual and experimental approaches should be clarified. First, all studies have used *mock* social sources of influence. That is, whenever I mention that averaged group responses, norms, or other people's judgments were presented in other scholars' or my own experimental investigations, these social cues are fully experimentally manipulated. Consequently, they do not represent the actual judgments of true participants. Second, two different settings have repeatedly been used to induce norm-based influence and measure conformal adjustment in naïve participants (see also Toelch & Dolan, 2015). In what might be termed the *one-shot paradigm* of social influence, participants learned about a social norm and were then confronted with the to-be-judged stimulus to which the previously presented norm pertained (Berns et al., 2005; Germar, Albrecht, Voss, & Mojzisch, 2016; Germar et al., 2014; Trautmann-Lengsfeld & Herrmann, 2013, 2014). Thus, only one single judgment, presumably varying as a function of the preceding norm, is given. In what we could call the *two-shot paradigm*, participants gave their judgment on a certain item and received, usually immediately after their own response, information regarding the norm. In an unannounced follow-up-rating session, all items were rated once more, this time without any feedback as to the norm (Chen, Wu, Tong, Guan, & Zhou, 2012; Kim et al., 2012; Klucharev et al., 2009; Klucharev et al., 2011; Nook & Zaki, 2015; Shestakova et al., 2013; Zaki, Schirmer, & Mitchell, 2011). Thus, each item was rated exactly twice, with the norm being conveyed between the ratings. The actual task that participants performed varied across studies; however, in most studies employing the one-shot paradigm, objective perceptual decisions were made, while most experiments based on the two-shot paradigm involved subjective, evaluative decisions (for an extensive discussion and an overview of previous studies, see *Publication I*). Interestingly, the distinction between different tasks might be less relevant than the differentiation of the one-shot and two-shot designs. As noted above and as will be explained in detail in the sections to come, the choice of paradigm largely determines one's focus of analysis, which in turn limits one's range of potential discovery.

As noted earlier, our current knowledge on the neurocognitive mechanisms of social influence is based almost entirely on research into the effects of group norms on individual behavior (i.e., conformity as per the definition presented earlier), although a few additional studies have investigated related social phenomana, such as persuasion (see, e.g., Falk, Berkman, Mann, Harrison, & Lieberman, 2010; Falk et al., 2015; Klucharev, Smidts, & Fernández, 2008). Given the differentiability of varying forms of social influence, as advocated by renowned scholars in social psychology (Cialdini & Goldstein, 2004; Wood, 2000), it seems prudent to focus on the type of scenario and psychological process explicitly addressed in the vast majority of studies and in my own line of research (for a related discussion of this distinction, see *Publication I*). Therefore, most of the following will emphasize social influence induced by group norms.

4.1 Modified Representations of To-Be-Judged Features

In *Publication I*, a comprehensive review of the literature published before 2014, we have carved out a fascinating similarity of several seemingly unrelated studies: There is evidence that social influence involves the modulation of the individual's own representation of the currently to-be-judged feature. This idea derives from experiments in which the influence of a previously presented norm on subsequent processing was assessed (one-shot paradigm). Specifically, several researchers asked participants to judge a stimulus as to a certain quality and presented the majority's response directly before the to-be-judged item was processed (Berns et al., 2005; Germar et al., 2016; Germar et al., 2014; Koban & Wager, 2016; Trautmann-Lengsfeld & Herrmann, 2013, 2014). First and foremost, all of these studies demonstrated the pervasive influence of social norms by showing that individual responses were aligned to the preceding majority judgment. Additionally, the studies neatly converge in finding evidence of systematic changes in the basic mental representation of the to-be-judged feature, which is particularly noteworthy given the fact that vastly different scenarios, tasks, and measurements were used.

4.1.1 Previous Findings

Berns and colleagues (2005) asked participants to perform a mental rotation task while being scanned in an MRI machine. Participants saw two images of geometric shapes and had to indicate by button press whether the two images depicted the same shape, with one having been rotated differently than the other, or two different shapes. In two thirds of all trials, four small pictures on the side of the screen indicated—before the pair of to-be-judged images appeared—whether four ostensible other participants had unanimously judged the upcoming shapes as identical or different. In a control condition, which was administered as a separate block before or after the social condition (with the order counterbalanced across participants), the judgments of a *computer* were ostensibly presented instead of the mock group's responses. Participants were significantly more likely to follow the incorrect judgments of the group (41%) than those of the computer (32%). The main focus of the fMRI analysis was on the processing of the to-be-judged shapes, during which BOLD responses were recorded as a proxy for regional brain activity. Berns and colleagues assessed neural activity during the processing of shapes to which participants subsequently responded incorrectly by following the group or the computer. Crucially, they compared trials in which participants incorrectly conformed to the group with trials in which they incorrectly aligned their response to a computer. The main finding was that activity in the baseline network associated with the mental-rotation task, particularly in the visual cortex, was larger during adjustment to the group as compared to adjustment to a computer. According to the authors, this indicates that the group norm affected participants' actual perceptual analysis of the to-be-judged objects.

Trautmann-Lengsfeld and Herrmann (2013, 2014) later challenged the interpretation favored by Berns and colleagues (2005). They argued that the poor temporal resolution of fMRI does not allow us to ascertain that activity modulations in cortical sensory networks indeed reflect changes during early stages of visual perception. Instead, it might indicate a modulation of later, reentrant processes in visual areas (Trautmann-Lengsfeld & Herrmann, 2013). To address this issue, they ran a similar experiment, but analyzed ERPs instead (Trautmann-Lengsfeld & Herrmann, 2013, 2014). Participants viewed two neighboring random-dot patterns and indicated which of the two contained more dots. Again, the stimuli were preceded by ostensible group judgments. The most striking finding was that the early visual brain potential P1, an index of perceptual processing in visual cortices (Martínez et al., 1999), was attenuated when participants agreed with an incorrect group opinion as compared to when they agreed with a correct group response (Trautmann-Lengsfeld & Herrmann, 2013). The authors interpreted this finding as more compelling evidence that social influence affects *early* perceptual processing, rather than late reentry processes based on feedback signals from higher cortical structures. Although these findings are not entirely conclusive on their own, including a possible confound with spatial cueing (for an extended discussion, see *Publication I*), Trautmann-Lengsfeld and Herrmann's (2013) claims fit well with Berns et al.'s (2005) proposition. Moreover, the findings were partially replicated in a subsequent study (Trautmann-Lengsfeld & Herrmann, 2014), in which both the tendency to conform and the abovementioned attenuation of the posterior P1 were stronger for participants scoring low (as compared to high) on a survey measuring trait autonomy.

The notion that perceptual processing can be influenced by social norms recently received support from a study employing methods of computational modeling. Germar and colleagues (2014) used a diffusion-model analysis to estimate the contribution of different cognitive processes to overt binary decisions, which is done by taking a person's responses and the distribution of their timing into account. The basic rationale is that a decision process can be understood as a stochastic diffusion of accumulated evidence between two boundaries that represent the two possible responses (Ratcliff, 1978). A decision is reached whenever the threshold of a specific response has been overcome (Voss, Rothermund, &

Brandtstädter, 2008). When a stimulus is processed and evidence in favor of either one of the two responses is gathered, a diffusion process determined both by a systematic and a random component is initiated (Germar et al., 2014). Two of the parameters that determine the speed and output of this diffusion are the relative starting point and the drift rate. The starting point refers to the potential level of bias in favor of one of the two responses, with the diffusion process already beginning closer to one threshold (i.e., response) than the other. The drift rate indicates the speed at which sensory information in support of a certain response is gathered. A large drift rate signifies a strong and systematic diffusion toward a certain response, which is typically observed when the to-be-judged feature of a stimulus is particularly obvious (Voss, Nagler, & Lerche, 2013).

In two experiments, Germar and colleagues (2014) asked participants to indicate whether a dichromatic pattern contained more blue or orange dots, learning about the judgments of a mock group of fellow participants immediately before each pattern. Participants were much more likely to give a certain response if the group had previously given this specific response. Crucially, this was not a mere priming or anchor effect caused by the previous occurrence of a certain response, as the adjustment effect was significantly smaller in a control condition in which the group's responses were seemingly irrelevant. Importantly, Germar and colleagues estimated diffusion-model parameters for each participant and condition, and they found that the overt behavioral social-influence effect was best explained by a modulation of the drift rate, while the relative starting point appeared unaffected by the group's judgments. Having learned that the group had given a certain response, participants subsequently accumulated evidence in favor of this specific response more strongly and quickly than when the group had given the opposing response. This indicates that the uptake of sensory information was directly modulated by the previously observed group norm. Norms thus appear to profoundly bias perceptual processes rather than inducing only a generalized bias to produce a given response (Germar et al., 2014). This notion was replicated only recently in a similar paradigm (Germar et al., 2016).

Interestingly, a very recent study on pain perception provided additional support for the assumption that social influence goes beyond the level of overt responding. Koban and Wager (2016) demonstrated that previously presented information on how others ostensibly rated the intensity of a painful stimulus affected not only participants' ratings as to how painful the stimulus was: Autonomic responses to pain stimuli, measured in an objective way by recording skin conductance responses, were systematically modulated by what a prior norm indicated regarding the painfulness of these stimuli. The effect of pain-related norms on pain evaluation and on autonomic responses was mediated by participants' explicitly stated expectations that developed based on the purported norms.

Zaki and colleagues (Zaki et al., 2011) demonstrated effects of social norms on fundamental internal representations during highly subjective, evaluative judgments. Note that they used the previously mentioned two-shot paradigm of social influence. However, they did not focus on the processing of the norms themselves, but rather on how the to-be-rated items were processed when they were shown for a second time. Participants rated the attractiveness of faces and learned, after each rating, how an ostensible group had purportedly rated this face on average. Slightly later, participants unexpectedly rated all faces once more, yet brain activity was now monitored using fMRI and no group judgments were shown. Faces that had been rated more favorably by the group than by the individual elicited stronger BOLD responses in the ventral striatum and in orbitofrontal cortex than faces that had been rated less favorably by the group than by the individual. Crucially, both areas are known to code the subjective value that one assigns to an item (see, e.g., Berns, McClure, Pagnoni, & Montague, 2001; Schultz, 2000, 2002). This finding was replicated within the same sample that underwent the social-influence paradigm by running an additional control task that allowed the authors to localize brain activity induced by non-social, monetary reward (Zaki et al., 2011). These results indicate that social influence directly shapes the neural representation of value, obviously causing a systematic alignment of decision-relevant features to match a group's valuation (Zaki et al., 2011). This finding is at odds with a study in which no modulation of activity in brain networks associated with subjective valuation was observed when adolescents learned about their peers' preferences (Berns et al., 2010). Other studies do, however, support the notion proposed by Zaki and colleagues (2011) by showing that norms conveying other people's attitudes indeed modify activity in one of the central value-related neural hubs, the ventral striatum, during one's assessment of to-be-judged items (Campbell-Meiklejohn, Bach, Roepstorff, Dolan, & Frith, 2010; Mason, Dyer, & Norton, 2009; Sherman, Payton, Hernandez, Greenfield, & Dapretto, 2016; see also Campbell-Meiklejohn et al., 2012).

Finally, there is evidence that norms pertaining to our memories can actually alter our own memories. In a study by Edelson and colleagues (Edelson et al., 2011), participants watched a video and were asked several questions regarding the video a few days later. Crucially, they could observe how ostensible other participants responded to these questions. The real participants strongly aligned their responses to those of others, making substantially more errors when others incorrectly remembered details from the video than when they did not. The authors assessed BOLD responses during the perception of group norms (i.e., the purported memories of the unanimous majority). Crucially, they compared neural activity for norms leading to *persistent* conformal memories (i.e, false memories that were still in line with the incorrect group norm a week after it was perceived) with neural activity underlying the processing of norms that entailed only *temporary* conformal memories (i.e., those that were correctly reported a week after the incorrect norm was perceived). Amygdala activity and amygdala-hippocampus connectivity during the processing of group norms was notably stronger for norms that resulted in persistent memory conformity as compared to those leading only to temporary false memories. According to the authors, the involvement of hippocampal regions and the amygdala, which projects heavily to the hippocampus, indicates that social influence can truly alter the neural representation of memory (but see Deuker et al., 2013).

4.1.2 Novel Contributions

My own work in this line of research comprises one theoretical and one empirical contribution. First, I have integrated several findings on the neurophysiological and cognitive principles of social influence (*Publication I*), from which I have derived the proposition that social norms can systematically shift the mental representation of to-be-judged features. Second, in a straightforward ERP study, I have extended this basic notion and combined it with a different theoretical perspective that focuses on the general mode of cognitive processing (*Publication II*).

4.1.2.1 A Comprehensive Account

Clearly, the studies recapped above are quite diverse regarding their methods (fMRI, EEG, computational modeling, and peripheral physiological recordings) and paradigms (objective perceptual judgments, pain judgments, subjective evaluative ratings, memory task). Nevertheless, they share, at a somewhat abstract level, one basic notion: Social influence appears to be capable of directly shaping fundamental internal representations (*Publication I*). Obviously, the specific correlate of internal processes that is affected is vastly different between studies. This apparent discrepancy is easily resolved by taking a closer look at the varying objectives of participants in the different studies. It seems highly likely that the specific type of information (or feature) that is at the heart of one's current judgment is affected by social norms. During perceptual decisions, the representation of perceptual information is modulated by social norms (Berns et al., 2005; Germar et al., 2016; Germar et al., 2014; Trautmann-Lengsfeld & Herrmann, 2013, 2014); one's physiological response to pain is altered by currently available social norms on pain (Koban & Wager, 2016); norms regarding an item's valuation affect how we represent its value (Campbell-Meiklejohn et al., 2010; Mason et al., 2009; Zaki et al., 2011); norms on what others purportedly remember from the past shape our own formation of memory (Edelson et al., 2011).

The applicability of this mechanism is somewhat limited. It is restricted to situations in which we have gained knowledge of a social norm before we process a to-be-judged stimulus and make our own judgment. Often, however, we are already very familiar with an item when we encounter a norm regarding this item, such that we might be more inclined to compare the norm and our own assessment, instead of assessing the item in light of the norm. Moreover, the conditions and moderators of this type of mechanism have yet to be clarified. For example, based on previous studies, we can only assert that norm-induced modulations of relevant representations can occur when norm and judgment relate to the same issue: During perceptual judgments, a norm regarding perceptual features can alter perceptual processing in the individual. What is unclear, though, is whether norms can exert the same fundamental influence if they refer to a feature that is currently *not* at the heart of the individual's judgment. What influence does a norm regarding perceptual features of an item exert if one currently does not engage in a perceptual task, but rather subjectively evaluates the item? More refined manipulations are needed to disentangle whether the type of feature referred to by the norm (e.g., perceptual, evaluative, mnemonic) fully determines the type of modulation of subsequent representations, even if the norm does not pertain to the same specific feature upon which the individual rests his or her current judgment. Although this is highly speculative, it is conceivable that for a norm to exert a profound influence on mental representations, the subject of the norm has to be the subject of the individual's judgments as well, thus putting the norm at the center of the individual's attention.

The notion that social norms have such a fundamental influence calls for a more in-depth understanding of how this works at the neural level. In *Publication I*, we have proposed that knowledge of a norm creates a specific expectation, which alters and directs basic encoding processes in a top-down fashion (Pessoa, Kastner, & Ungerleider, 2003; Shuler & Bear, 2006; Summerfield & Egner, 2009). Expectations could elicit a neural template, such as a certain pattern of preparatory activation of the very neuron populations that code for the feature in question. When the actual to-be-judged stimulus is processed, the response of these neurons is a function of the actual information generated by the external stimulus itself *and* the expectation-induced activation template previously induced by the norm. One might also describe this process as a norm-based bias in the representation of the critical, norm-related stimulus feature, which is in keeping with recent interpretations of socially influenced perceptual processing (Germar et al., 2016; Germar et al., 2014). This notion is compatible with contemporary models of the general computational principles of the human brain which encompass the idea that the encoding of basic external information involves a constant, expectation-based backflow of information from higher-level neuron assemblies (Friston, 2005; Rao & Ballard, 1999). Thus, the representation of an actual sensory event is a reconciliation of external (bottom-up) and internal (top-down) signals (Summerfield & Egner, 2009).

The general idea that expectations mediate the influence of social norms on the representation of the currently to-be-judged feature (as briefly laid out in *Publication I*) has recently received empirical support (Koban & Wager, 2016): Norm-induced modulations of pain perception and peripheral physiological responses to pain were fully mediated by the explicit expectations triggered by the norms. It should be noted that such expectations are clearly not always as conscious and nameable as in the pain-perception study by Koban and Wager (2016). Models of expectation-based top-down comparisons in perceptual processes do not necessarily assume that people actually become aware of these expectations (Summerfield & Egner, 2009). Rather, these are highly automatized predictions of the neural apparatus that need not be reflexible. This is not so say, however, that they cannot become aware to the individual, such as when participants are explicitly asked which expectations social norms elicit (Koban & Wager, 2016). Nevertheless, the specifics of the general mechanism, as roughly proposed above, are still vague and highly speculative. Targeted manipulations and analyses are needed to substantiate this suggestion. What seems safe to state, however, is that social influence can indeed shape basic mental operations. Note that this assertion echoes some of the classical theories on social influence. As explained in detail earlier, various theorists have proposed that socially adjusted behavior can be the consequence of conversion, which denotes a true adoption of other people's judgments (Cialdini & Goldstein, 2004; Kelman, 1961; Moscovici et al., 1969). The mechanism of descriptive social norms directly tampering with the individual's mental representation of what is currently being judged might be the neurocognitive correspondent of the assumed process of conversion (Zaki et al., 2011).

4.1.2.2 The Role of the Depth of Processing

The notion that one crucial locus of the influence of social norms is the representation of the to-be-judged information itself is not entirely novel. The neurocognitive perspective involves a precise assessment of separate stages of processing (see, e.g., Germar et al., 2014) and the intensity of neural responses (see, e.g., Trautmann-Lengsfeld & Herrmann, 2013). Established sociopsychological theories have not resolved social influence to this level of cognitive detail. Nevertheless, some of them already consider the representation of judgment-relevant information, including the to-be-judged stimulus, quintessential in the process. Specifically, it has been proposed that one's current cognitive mode and, consequently, the *quality* of one's representation of judgment-relevant information is predictive of the tendency to adjust one's views or judgments (Axsom et al., 1987; Chaiken, 1980; Chen et al., 1996). According to the heuristic-systematic model of information processing (Chaiken, 1980), there are two fundamentally different modes of processing: a quick, heuristic and a systematic, elaborate mode. We tend to rely on heuristics—simple short-cut rules—whenever we choose or need to apply little cognitive effort during our decisions (Chen, Duckworth, & Chaiken, 1999). Rather than dissecting information intently and intensively, we stick to a simple, over-learned rule that represents our best guess at behaving appropriately without investing too much time and energy (see also Petty & Cacioppo, 1996). Crucially, social norms are the basis of two commonly used heuristics (Chen et al., 1996; see also Cialdini & Trost, 1998; Wood, 2000). When our primary motive is to be accepted by the group, we know that we can achieve this somewhat effortlessly by openly complying ("go along to get along"). When our primary goal is to be accurate and precise in our judgments, yet our resources are currently limited, we know that the majority's judgments tend to be correct, such that the adoption of majority views seems promising ("consensus implies correctness"). Therefore, heuristic processing should involve an increased tendency to adjust one's own judgment to the majority's, almost irrespective of one's current primary motive.

Cognitive styles are widely assumed to be directly linked to a person's current mood, with negative mood inducing a tendency to process information systematically, while positive mood fosters a heuristic mode (see, e.g., Bless et al., 1996; Bodenhausen, Kramer, & Süsser, 1994; Mackie & Worth, 1989). Fittingly, the tendency to conform to social norms varies as a function of mood: People in a positive mood rely more heavily on the judgments of others as compared to those in a negative mood (Tong, Tan, Latheef, Selamat, & Tan, 2008). Although participants' mood was experimentally manipulated in the study by Tong and colleagues (2008), their finding provides only indirect evidence of a link between cognitive style and conformity. There was no reliable assessment of the actual style of processing, which was only assumed to be in a certain state due to participants' mood.

In *Publication II*, we aimed to complement the findings of Tong and colleagues (2008) by providing the missing link between processing style and conformity. Our simple and straightforward idea was that the depth of one's analysis of a to-be-judged stimulus should be related to the degree of one's adjustment to group norms. We reasoned that individuals currently or dispositionally employing an effortless, heuristic strategy should be more inclined to adopt the group's responses than individuals who process judgment-relevant information more elaborately. To assess the depth of encoding of to-be-judged information, we used ERPs and chose faces as the target stimuli. The perceptual analysis of faces in the human brain is characterized by a highly specialized, distinct component: the N170, a negativity peaking around 170 ms at lateral posterior electrodes. Its neural generators are assumed to include the posterior fusiform gyrus and lateral occipitotemporal cortex (Deffke et al., 2007; Schweinberger, Pickering, Jentzsch, Burton, & Kaufmann, 2002). There is widespread consensus that it is a precise neural signature of the structural and configural encoding of faces (e.g., Bentin et al., 1996; Eimer, 1998, 2011). That is, it reflects an advanced encoding of faces as a whole, rather than a stepwise detection of salient features or subcomponents (Eimer, 2000). Previous work has shown that N170 amplitude is sensitive to motivational variations and, most importantly, is a reliable indicator of the general strength of face encoding (Amodio et al., 2013; Eimer, 2011; Ofan et al., 2011; Ratner & Amodio, 2013; Schmid & Amodio, 2016).

We asked participants to rate the attractiveness of a series of somewhat ambiguous virtual faces on a 6-point scale. Crucially, participants saw the ostensible average judgment of previous participants before each face. First, we demonstrated that participants significantly aligned their judgments to the purported norms. We were able to ascertain that participants did not simply adjust their ratings to the numerals indicating the group's judgments (in the sense of a simple priming effect), as no evidence of adjustment whatsoever was found when the numerals were not labeled as averaged group judgments. As part of our main analysis, we then assessed, for each participant, their individual propensity to conform and their average intensity of face-specific encoding. To this end, we first performed a linear regression analysis for each individual, predicting attractiveness ratings based on preceding group norms. For each participant, the slope of this regression (i.e., the increase of individual ratings with increasing group judgment) was used as a conformity score. Subsequently, we measured the mean amplitude of individual N170 peaks at parieto-occipital electrodes PO7 and PO8 during the observation of the to-be-evaluated faces, averaged across a total of 200 trials. In line with previous research (see, e.g., Ofan, Rubin, & Amodio, 2014), the facial N170 was more strongly pronounced at right-hemispheric sensors, such that we subsequently focused on electrode site PO8. However, one might argue that this need not be the case for all participants, which might become an issue when performing across-participant correlations. We therefore repeated all main analyses reported below after determining the laterality of the N170 individually for all participants, which yielded the exact same pattern of results.

Confirming our expectations, N170 amplitude was robustly related to conformity scores across participants (Spearman's ρ = .48). Note that the N170 is a negativity, such that the positive correlation indicates that a high level of conformity is associated with a low level of N170 amplitude. One might argue that this relation merely indicates that both face processing and conformity are related to a common third variable, such as a central personality feature. To test this idea, we analyzed whether the correlation would hold even if we partialled out the contribution of major personality characteristics (such as extraversion, neuroticism, openness, conscientiousness, agreeableness, or trait anxiety).

Confirming our initial findings, we found that the relation was still reliably observed even when these personality scores were controlled for.

The findings of *Publication II* support the claim of Tong and colleagues (2008) that conformal adjustment to a group is related to a person's mode of processing. The more superficially a person encodes the to-be-judged stimuli, the more likely he or she is to adopt group norms regarding these items. Clearly, our study is correlational, prohibiting causal inference. Yet our findings can and should be interpreted in the context of previous studies in which the style of processing was indirectly manipulated and no actual measure of the strength of stimulus encoding was available. The use of ERPs allowed us to fill this empirical gap and support previous behavioral data and theories regarding the role of cognitive processing in socially adjusted behavior (Axsom et al., 1987; Chen et al., 1999; Chen et al., 1996; Tong et al., 2008). Moreover, I propose that this association is equally likely a part of both commonly reported motives that drive conformal behavior. First, if participants had only an inadequate representation of the to-be-judged stimuli or could not bring themselves to attend to them closely, they still might have wanted to achieve at least a certain level of accuracy in their judgments. Based on their knowledge that the group is often a valid source of information, they could have resorted to its norms, which would constitute informational social influence driven by an *accuracy* motive. Second, participants who were inclined to adjust their judgments to the group anyway might not have been motivated to process the to-be-judged faces at all or might have even tried to dampen the input evoked by the faces. The latter would have allowed them to avoid "cognitive dissonance" (Festinger, 1962) that would arise from following others even if their judgments were sometimes openly erroneous. Assuming that the motive of a generalized tendency to conform to the group, irrespective of its accuracy, is to avoid negative social sanctions, this would suggest that normative social influence, driven by an *affiliation* goal, was the motivational foundation of the observed relation. To summarize both accounts, it seems possible that superficial encoding of the virtual faces might have necessitated conformity (informational influence) or a superordinate inclination to conform (normative influence) might have provoked or even necessitated superficial processing. Either way, our data corroborate the notion that social influence is robustly associated with the style of processing. Specifically, we emphasize the role of the depth of encoding of the object of one's judgment.

In light of our proposal that the mental representation of to-be-judged information is a pivotal locus of social influence (*Publication I*), the findings of *Publication II*, in combination with complementary behavioral research, can be seen as an affirmation and extension of our theorizing. First, it underlines the idea that social influence does not only work via a late, decision-related correction of one's response, but rather relies on modifications of early, basic representations during the decision-making process. Second, it adds the notion that it is not only specific norms (such as "this face is popular" or "this pattern is mostly orange") that directly modulate basic representations. The overall thoroughness of the formation of representations likewise contributes to our conforming to social norms.

Interestingly, Germar and colleagues (2014) reported that participants processed to-be-judged visual patterns more carefully when they were under the impression that they saw relevant, as compared to irrelevant, group judgments. Only recently, they replicated this finding and provided additional support using ERPs (Germar et al., 2016). At first glance, this might appear to be at odds with the above claims that social influence is associated with superficial processing of to-be-judged information. Both findings are, however, indeed compatible. Importantly, the results presented by Germar and colleagues (Germar et al., 2016; Germar et al., 2014) are not entirely surprising, as they could constitute an instance of the well-known phenomenon of *social facilitation* (see, e.g., Bond & Titus, 1983; Harkins, 1987). Situations in which potentially valuable social information is available might elicit a general enhancement of the individual's vigilance or motivation to excel at a task, perhaps triggered by an increased "generalized drive" (Zajonc, 1965). Therefore, comparing a decision-making scenario in an expressly social situation with a less socially relevant situation (as done in Germar et al., 2016; Germar et al., 2014) should reveal enhanced stimulus processing in the former as compared to the latter. What we did in *Publication II*, however, was to specifically focus only on participants who were all uniformly exposed to social influence. Even if being exposed to others deepens one's analysis, one might still find that—within a population *uniformly* exposed to social influence—those who process the to-be-judged items more superficially tend to rely on group norms particularly heavily (Tong et al., 2008). As proposed by Germar and colleagues (2016), the overall depth of one's assessment of the to-be-judged item could have a moderating effect on other mechanisms, such as biased perceptual analysis, that actually cause social conformity.

4.2 Social Proof and Deviance as Behavioral Outcomes

To understand the neural and cognitive foundations of social influence, it seems prudent to explore how knowledge of a norm alters the way in which we process the objects of our judgments. Another perspective, however, is to put special emphasis on the moment of one's perception of the norm itself. Crucially, as noted earlier, it stands to reason that we often develop an attitude toward a to-be-judged issue *before* we are confronted with the norm. Perceiving a social norm, then, should usually instigate a process of comparison. In fact, it has been proposed that our assessment of other people's judgments against the backdrop of our own is a central aspect of social interaction and a major determinant of social influence (Cialdini et al., 1999; Festinger, 1954). This is in keeping with the general assertion that humans tend to process their environment relatively or comparatively (Kahneman & Miller, 1986).

At this point, it might be helpful to establish a certain terminology that will recur in most of the studies reported hereafter. When a person makes a judgment, or already holds an attitude toward a given issue, and this person is then confronted with how others have judged the same issue, the assumed process of comparison can yield two results: The group's judgment can be consistent with or different from the individual's. Various denominations of the two outcomes of this comparison are conceivable and, fittingly, they are not labeled consistently throughout the literature. Some refer to them simply as social conflict versus no social conflict (Klucharev et al., 2009), but one might also call them social (in)congruency, (dis)agreement, or (mis)match. It should be noted that all of these terms are used interchangeably in the present context. Mostly, however, I prefer to label the two potential comparison results as social proof and deviance.

Not much surprisingly, a substantial share of the literature regarding the neurocognition of social influence has investigated how individuals assess and evaluate group norms in relation to their own judgments. Although they have used different labels, these studies have basically examined how social proof and deviance are represented in the human brain. Initiated by an innovative fMRI study (Klucharev et al., 2009), various scholars have tried to characterize the neural processes engaged in determining these two relevant social outcomes and to find out how such processes relate to subsequent behavior and cognitions. This line of research has led to the proposition that neural circuits generally involved in reinforcement learning play a key role in social influence (see Izuma, 2013): Norms appear to trigger the differentiation of social proof and deviance, which is afforded by neural structures that typically code *behavioral outcomes*.

Our behavior in response to a given stimulus is largely determined by the outcome of this behavior, that is, how pleasurable or painful the consequences of this specific response to this specific stimulus are (Thorndike, 1911). Based on previous reward or punishment, many animals learn how to behave upon the occurrence of a certain stimulus or event (Thorndike, 1927). Although this idea might seem self-evident from a present-day perspective, it was fiercely contested upon its original publication (see Thorndike, 1933). The concept of learning based upon reinforcement has evolved to a significant degree over time, extending its scope and application both within psychology and well beyond (see, e.g., Frank, Seeberger, & O'Reilly, 2004; Kober & Peters, 2012; Moore & Atkeson, 1993; Sandholm & Crites, 1996). One core aspect of previous and contemporary research has been to disentangle the neural networks and mechanisms underlying reinforcement learning (for a recent review, see Lee, Seo, & Jung, 2012). Midbrain dopaminergic neurons have been identified as the key, albeit not the only, structures that represent the value of reinforcers (i.e., stimuli perceived as rewarding or punishing) and cues promising the later receipt of reinforcers (for an overview, see Schultz, 2010). Perhaps the most relevant contribution comes from neurons in the substantia nigra and the ventral tegmental area as well as from neurons in forebrain areas, some of which—such as the basal ganglia and frontal cortical structures—receive projections from the midbrain dopaminergic neurons (Schultz, 2000). In the below summary of previous findings, some of the reward-sensitive forebrain areas will be of particular interest, as they appear to code not only typical reinforcers, such as food or money, but even more complex ones, such as (dis)agreement with social norms.

The neural representation of behavioral outcomes is usually not absolute, but rather relative (Schultz, 2010). Specifically, we process reward and punishment relative to our current expectation of what the outcome will be—although it should be noted that these are not necessarily conscious expectations on which the individual actively reflects, but often implicit, computationally derived predictions based on the system's previous experience. The notion that outcomes are processed in relation to one's predictions is quite logical or even necessary, seeing as outcomes have a remarkable potency to induce substantial changes in behavior (Thorndike, 1911). That is, they represent an effective source of *learning*, in which the comparison of actual and predicted events generally plays a crucial role: Learning, the change in the strength of an association between two units (e.g., between two events or between an action and an outcome), advances as a function of the degree of *unpredictability* of one unit from the other (Rescorla & Wagner, 1972; Schultz, 2006). In instrumental learning (i.e., learning driven by reinforcement or the lack thereof), this means that our association between a certain action and an outcome grows more quickly if the actual outcome differs *much* from the predicted outcome than if predicted and actual outcome are hardly different or even identical. Therefore, a crucial parameter that affects the strengthening of an action-outcome association is the *reward prediction error*: the degree of deviation of an outcome expectation from the actual outcome. Therefore, it seems consistent and, in a way, economical that the human brain generally processes reward relative to a prediction based on previous reward and punishment (Schultz & Dickinson, 2000). The midbrain dopaminergic system—with its connections to basal and cortical structures of the forebrain, including the ventral striatum and the cingulate cortex—is assumed to provide the neural basis of the representation of prediction errors during the assessment of behavioral outcomes (Schultz, 1997, 2002). Dopamine neurons bidirectionally code both *positive prediction errors* (i.e., an outcome is better than predicted) and *negative prediction errors* (i.e., an outcome is worse than predicted) via activation and depression, respectively (Schultz, 2007).

4.2.1 Previous Findings

Klucharev and colleagues (Klucharev et al., 2009) asked participants to rate a series of faces as to their attractiveness. After each response, participants saw the ostensible average rating of participants who had purportedly completed the same task previously. A small number additionally indicated the degree of deviation of the participant's judgment from the ostensible group's (0, ±2, ±3). Participants did not encounter the group at any point or were made to feel pressured to belong. Rather, the norm was merely a depiction of the typical behavior of the group; that is, participants were confronted with a descriptive social norm (Klucharev et al., 2009). In an unannounced reassessment session, participants were later asked to evaluate all faces for a second time. As expected, a substantial adjustment of individual ratings was observed. Faces initially rated less positively by the individual than by the group were evaluated significantly more favorably during the second as compared to the first assessment. Likewise, participants lowered their ratings of faces that they had previously rated more favorably than the group. Only faces on which the individual agreed with the norm were rated similarly during the first and second assessment. The authors concluded that a strong tendency to conform was evident. Whenever social deviance was detected, ratings were adjusted in the direction of the norm. The main focus of analysis was, however, on the moment when participants were confronted with the group norm, which should trigger perceptions of social proof or deviance.

When participants observed a mismatch, as compared to a match, between their own and the group's judgment, activity in the anterior cingulate cortex increased, while activity in the nucleus accumbens, located in the ventral striatum within the basal forebrain, decreased (Klucharev et al., 2009). This effect was robust and much stronger when ostensible group norms were presented than when computer-generated judgments were purportedly shown. Moreover, activity increases in the anterior cingulate in response to mismatches positively predicted participants' subsequent tendency to adjust their ratings toward the norm. The pattern of activity modulations in the anterior cingulate cortex and nucleus accumbens suggests that social proof and deviance are represented similarly (or, at least, by the same neural networks) as reinforcement in general (Klucharev et al., 2009). A vast body of research has implicated the anterior cingulate in the processing of negative feedback and punishment (Gehring & Willoughby, 2002; Holroyd & Coles, 2002). Likewise, there is ample evidence of the involvement of the nucleus accumbens in reward coding (Berns et al., 2001; McClure, Berns, & Montague, 2003).

In the study of Klucharev and colleagues (2009), the pattern of activity indicated the involvement of typical neural responses to reinforcers, from which the authors concluded that behaving in line with social norms is perceived as rewarding, while deviating from it is treated as an error or punishment. Although it is well established that such behavioral outcomes are represented according to their deviation from an internal reward prediction (Schultz & Dickinson, 2000), which Klucharev et al. (2009) explicitly acknowledged, they did not clarify the exact nature of prediction-error coding in their experiments. There was no suitable control condition, without feedback or with a neutral outcome, and situations of agreement between norm and individual judgment were directly compared with those of disagreement. Therefore, it remains unclear whether (a) social proof elicited a positive prediction-error signal, or (b) social deviance led to a negative prediction-error signal, or (c) both prediction errors occurred. Nevertheless, the findings of Klucharev and colleagues (2009) conclusively indicate the involvement of neural hubs that are known to code behavioral outcomes and contribute to reinforcement learning (Izuma, 2013). This is supported by data showing (a) that a gene assumed to affect dopaminergic neurotransmission is significantly associated with the likelihood of conforming to the group, and (b) that perceiving a majority's judgment deviate from one's own correct assessment is associated with increased activity in the anterior cingulate (Deuker et al., 2013). By downregulating activity in the posterior medial frontal cortex, which includes the anterior cingulate, via transcranial magnetic stimulation (TMS), the causal involvement of this area was demonstrated (Klucharev et al., 2011): Participants whose posterior medial frontal cortex was suppressed were significantly less likely to adjust their judgments toward the group's than participants with sham or no stimulation.

Moreover, ERPs were used in order to test whether social proof and deviance differentially affected a well-known correlate of outcome processing. As mentioned earlier, the FN is deemed a reliable neurophysiological signature of the assessment of behavioral outcomes (see, e.g., Gehring & Willoughby, 2002; Miltner et al., 1997). It should be noted that various different conventions regarding the terminology of the FN and related components exist. Loosely based on a recent study (*Publication IV*), the following differentiation and nomenclature seem appropriate: When participants observe the outcome of an action, a relative negativity, often superimposed upon a broader drift of the ERP in the positive direction, is often observed for both negative and positive feedback around 250 ms (see, e.g., Gehring & Willoughby, 2002; Kreussel et al., 2012; Pfabigan, Alexopoulos, Bauer, Lamm, & Sailer, 2011). There is some variation as to the shape of the deflections, as a peak in the negative direction is not consistently observed for positive feedback (see, e.g., Baker & Holroyd, 2011; Holroyd, Hajcak, & Larsen, 2006; Miltner et al., 1997). The component elicited by negative feedback is often referred to as the FN (or FRN), while the component induced by positive feedback is the reward positivity (Rew-P; Baker & Holroyd, 2011). When the ERP for positive feedback is subtracted from that for negative, the resultant difference curve reveals a negativity around 250 ms that denotes the FND, the difference between the raw FN and the raw Rew-P (for a similar terminology, see *Publication IV*).

Several recent studies using various different tasks consistently report an FND during the processing of social norms following individual judgment, based on a larger relative negativity during the detection of one's deviance from the norm compared to one's agreement with it (Chen, Wu, et al., 2012; Kim et al., 2012; Kimura & Katayama, 2013; Kimura et al., 2013; Shestakova et al., 2013; Yu & Sun, 2013). As the FN is associated with the processing of negative feedback, loss, and punishment (Gehring & Willoughby, 2002; Holroyd & Coles, 2002), while the Rew-P is triggered by positive feedback, reward, and gains (Baker & Holroyd, 2011; Holroyd, Krigolson, & Lee, 2011), this strongly supports the findings from fMRI and TMS studies (Klucharev et al., 2009; Klucharev et al., 2011). As explained above, it appears that group norms are instantaneously evaluated as to their rewarding or punishing character, based on whether or not they are in accordance with one's own judgment. This suggests that reinforcement learning, fed by outcome-triggered evaluative signals, is at the heart of our adjustment to norms (for an extended discussion, see *Publication I*). Additional research has further extended this view by demonstrating that the source of a descriptive norm is a crucial moderator of the neurocognitive processes underlying our evaluation of social norms (Izuma & Adolphs, 2013). As this exceeds the scope of the present synopsis, the interested reader is referred to recent reviews (Falk et al., 2012; Izuma, 2013; Stallen & Sanfey, 2015).

4.2.2 Novel Contributions

Previous work on the contribution of outcome assessment to social influence is pervasive and conclusive. Nevertheless, it merely represents a starting point from which to explore the specifics of this mechanism. In fact, even today, seven years after Klucharev and colleagues (2009) first suggested the involvement of brain functions associated with the processing of behavioral outcomes and, thus, with reinforcement learning in social influence, there is much to be clarified about the assumed process. My colleagues and I have filled empirical gaps regarding fundamental methodological issues (*Publications III, IV, V*) and extended our knowledge of the process itself (*Publications VI, VII, VIII*). Importantly, the key presupposition in this series of studies is the idea that our capability to determine the value of our actions, which is the quintessential premise of reinforcement learning, largely determines our mental representation of social norms.

4.2.2.1 Correctly Assessing Deviance-Induced Conformity

In recent years, a frequently used measure of conformity was the degree of the correction of one's judgment as a function of the level and direction of (dis)agreement with the social norm, as assessed in the typical reinforcement-learning paradigm of social influence (Klucharev et al., 2009). Participants typically rate an item, receive information about the group norm, and later rate all items once more. The change of each item's evaluation (second minus initial assessment) is used as dependent variable, while the deviation of the norm from individual judgment is the critical independent variable (norm is below the individual's judgment; norm is equivalent to the individual's judgment; norm is above the individual's judgment). Downward changes for images initially rated more favorably by the individual than by the group (norm-lower condition) and upward corrections for images initially rated less favorably by the individual than by the group (norm-higher condition) are consistently reported in this paradigm. Unfortunately, this setup entails an unwanted confound (Yu & Chen, 2015): The norm can only be lower than the participant's own judgment whenever the latter was sufficiently high. Likewise, the norm can only exceed a participant's judgment if the latter was sufficiently low. To put it differently, initial judgments that are already somewhat low cannot be met by lower norms; also, high initial judgments cannot be met by higher norms. Thus, certain deviations are tied to a certain extremity of the initial response. At first glance, this might appear irrelevant, as it is a natural consequence of our use of finite scales in experiments and, presumably, of our actual mental representation of evaluations. It is, however, a problem if we consider a well-known natural consequence of the extremity of judgments. If we assess a scalar measurement twice, initially extreme judgments tend to lie closer—regress—to the mean of the distribution during their second assessment (Blomqvist, 1987; Cutter, 1976; Pezdek & Eddy, 2001; Stigler, 1997; Tversky & Kahneman, 1974). If we look at judgments followed by lower group norms, we observe a downward correction of the second assessment, which we interpret as the direct consequence of the deviation from the intervening norm. However, since the initial judgments followed by lower norms were actually somewhat high themselves, it is likely that their decrease was not, or not solely, an effect of the intervening norm's deviation, but an instance of natural regression toward the mean (RTM). Likewise, we consider upward corrections an effect of the intervening norms that lie above participants' initial judgments. Again, however, it is possible that the sheer lowness of those specific judgments that were followed by higher norms actually triggered the correction in the upward direction, which is both the direction of the norm and of the mean of the rating distribution. Therefore, behavioral measures of deviance-induced conformity are bound to be thoroughly imprecise. They could reflect actual conformity and RTM, or they might simply represent RTM effects that were completely misinterpreted as effects of norms and their deviation from individual judgment.

Yu and Chen (2015) replicated the procedure of Klucharev and colleagues (2009), yet controlled for the influence of RTM by removing the contribution of the level of initial ratings. Note that, as explained above, the level of an initial rating is associated both with a certain tendency to regress and with a certain norm deviation that can follow the initial rating. This unwanted confound of deviation and RTM prevents us from knowing whether behavioral adjustment is caused by deviation, the confounded RTM, or both. Yu and Chen (2015) tested two different strategies to control for the level of initial ratings: (a) selecting only a subset of items for analysis to match the mean level of initial ratings followed by norms deviating in different directions; and (b) including the level of initial judgments as a covariate in a regression analysis. Both strategies caused the previously observed effect of deviation between norm and individual judgment on rating changes to vanish completely. It should be noted, however, that it would be premature to dismiss previous work on the grounds of this study and to conclude that the deviance-based paradigm does not induce conformity. Most importantly, Yu and Chen (2015) ran only 13 participants, which forbids us to interpret null effects at all. Moreover, completely purging the influence of initial ratings on rating change seems overly drastic. Conformity might vary as a function of the extremity of one's own initial judgment; for example, if one attends more closely to social norms when they concern items on which one has taken a firm stance and chosen a somewhat extreme judgment. Therefore, the contribution of initial ratings should not be eliminated, but rather corrected for only to the degree to which they actually inflate the assessment of conformity.

In *Publication III*, we addressed the issue of RTM in deviance-induced conformity for two reasons. First, we wanted to reassess whether the typical paradigm indeed yields no true conformity effect whatsoever. Second, we aimed to devise a novel, more precise correction procedure. To tackle both enterprises, we used what is typically assumed to be the best approach to control for RTM (Barnett, 2004; Yu & Chen, 2015): a straightforward control-group design. In addition to an experimental group that performed the basic task previously described, we ran a control group of participants who did the exact same experiment except for one important detail: They did not see any group norm after each initial judgment. We did, however, program the experimental protocol in a way that allowed us to register for each trial which group norm would have been presented in this moment—if this were the experimental group. That is, after each participant judgment, a norm that could theoretically be presented following this judgment was drawn, yet not shown. Subsequently, we analyzed for both groups whether the level of deviation between individual judgment and group norm affected participants' rating changes. As expected, we observed a very strong effect in the experimental group: Items in the norm-lower condition had a negative change score (indicating downward corrections), while items in the norm-higher condition had a positive change score (indicating upward corrections). For items on which norm and individual agreed, change scores close to zero were observed. This finding neatly replicates various previous findings (e.g., Kim et al., 2012; Klucharev et al., 2009; Klucharev et al., 2011; Shestakova et al., 2013).

Interestingly, however, we observed a very similar effect of norm deviation on rating changes in the control group. This is particularly striking given that the control group did not receive any information regarding these group norms whatsoever. Nevertheless, we found a significant effect of this invisible pseudo-factor when analyzing rating change as a function of the norms that would have been, yet were not actually, presented. This clearly demonstrates that rating changes, which would have been attributed to the intervening norms if we had not implemented the control group ourselves, occur even in the complete absence of social norms. This reflects the influence of RTM (see also Yu & Chen, 2015). Crucially, the pseudo-effect in the control group was significantly smaller than the effect in the experimental group, which indicates that more than RTM contributed to behavioral adjustment when participants actually observed group judgments. The observed rating changes in the experimental group were most likely compounds, comprising both true conformity effects and RTM.

Finally, we used a hierarchical linear model to extract from the control group an algorithm allowing us to predict for each participant in the experimental group the level of RTM that was to be expected for a given item. Specifically, we analyzed for all items (level 1) in all participants (level 2) of the control group how rating change, which is only the consequence of RTM, can be predicted based on the level of an initial rating. This gave us a simple algorithm that allowed us to estimate for a given initial rating how much RTM would have to be expected in a follow-up rating. We applied this algorithm to each item in each participant of the experimental group and subtracted the resultant estimate of to-be-expected RTM from the observed rating change. The result of this subtraction should be an appropriate estimate of the true conformal adjustment behind the observed compound rating change for each item. Finally, we used these corrected rating changes, the putative estimates of true conformity, as dependent variable and repeated the analyses described earlier. As expected, a significant effect of norm deviation on corrected rating changes was observed, supporting our notion that a true conformity effect exists in this paradigm (see also Huang, Kendrick, & Yu, 2014; Nook & Zaki, 2015; Zaki et al., 2011). In the control group, our correction for RTM completely eliminated the previously observed pseudo-effect of the nonexistent variable norm deviation, supporting our idea that this effect was completely carried by RTM.

In *Publication III*, we thus strongly corroborated the previous claim that behavioral effects observed in paradigms assessing deviance-induced conformity contained an unwanted contribution of regression toward the mean. We did, however, demonstrate that RTM causes an inflation of the true conformity effect, yet does not cause the observed effect in its entirety. Finally, we proposed a novel correction procedure to effectively, yet carefully and selectively remove the contribution of RTM from one's measure of behavioral adjustment. One crucial benefit of our approach is that the correction is performed on the lowest level of analysis (the to-be-judged items), thus allowing an early cleaning of measurements. This enables us to use cleaned values not only during group-level analyses, but already during the determination of individual behavioral scores. This is particularly important as across-participant correlations of conformity scores and external variables are often used in this line of research, for example, when neuroimaging or genetic data are available (Falk et al., 2012). Therefore, clean and valid *individual* measures of conformity are necessary.

4.2.2.2 The Functional Significance of the FND

As expressed earlier, ERP studies on social influence have mostly focused on the representation of social proof and deviance as behavioral outcomes. Specifically, the FND has been assessed and discussed in considerable detail (Chen, Wu, et al., 2012; Kim et al., 2012; Shestakova et al., 2013). Based on the finding of an FND for social deviance relative to social proof, it has been concluded that group norms that are at odds with one's own previous assessment are represented very much like negative feedback, loss, or punishment (Kim et al., 2012). What has mostly been neglected in this line of research, however, is that the neurophysiological mechanisms underlying the FND in general have been subject to debate in recent years. It is prudent to carefully examine and experimentally test the functional significance of this component to better understand its principles. This, then, should feed into our interpretation of the FND caused by social norms, which can then further inform our understanding of the reinforcement-learning mechanisms of social influence.

The seminal finding that subtracting the electrophysiological response to positive feedback from that to negative feedback yields an enhanced negativity around 250 ms (i.e., the FND) has typically been interpreted as a sign of enhanced activity in the anterior cingulate cortex during the assessment of negative outcomes (Miltner et al., 1997). Triggered by input from midbrain dopaminergic neurons, the anterior cingulate cortex might be disinhibited and signalize that outcomes are worse than expected (i.e., negative prediction errors), which is critical information in reinforcement learning (Holroyd & Coles, 2002). The FND would thus rely on signals of the need for adjustment from the anterior cingulate (Ridderinkhof, Ullsperger, Crone, & Nieuwenhuis, 2004). A slightly different theoretical account likewise suggests that the anterior cingulate is at the heart of this component, but it assumes that it signals prediction errors in general, irrespective of whether the outcome is better or worse than anticipated (Alexander & Brown, 2011). More recently, the somewhat advanced discussion of the neural generators of the FND has given way to a slightly more basic concern with how the FND actually comes about, which is crucial in interpreting its functional significance. This is a matter of debate as the FND is derived from a difference wave, rather than being an actual component itself (Miltner et al., 1997). Therefore, multiple explanations of the occurrence of a relative negativity are conceivable. Moreover, as explained earlier, ERP deflections often reflect the summation of various overlapping neural responses, sometimes even differing in polarity. Consequently, it is occasionally hard to discern the origins of variations of actual components, let alone difference waves (Luck, 2005).

Recent studies have questioned the basic idea that an enhanced response to *negative* feedback elicits the FND. Several findings indicate that a stronger response to positive than to negative feedback causes the observed difference (e.g., Baker & Holroyd, 2011; Eppinger, Kray, Mock, & Mecklinger, 2008; Foti, Weinberg, Dien, & Hajcak, 2011; Kreussel et al., 2012; Kujawa, Smith, Luhmann, & Hajcak, 2013). Positive, but not negative, outcomes might trigger a positivity (Rew-P), presumably originating in the basal ganglia (Foti et al., 2011). This has, however, long gone unnoticed because researchers often assessed *difference waves* rather than raw components: Subtracting the response to positive feedback from that to negative feedback, one observes a negativity in the FN time range, which is erroneously interpreted as a sign of an increased negativity caused by negative feedback (Miltner et al., 1997). However, even if one does look at the raw components elicited by positive and negative feedback separately, one might be inclined to assume that the effect is driven by the negative condition. Sometimes, albeit not always, during feedback processing, a strong negativity (N2) that is *not* directly associated with the valence of the feedback occurs almost simultaneously with the feedback-driven components (Baker & Holroyd, 2011). As a consequence of this concurrence of components, even the raw Rew-P triggered by positive feedback is not always visible as a positive component, but only attenuates the N2 (Holroyd et al., 2011). What is actually an increased positivity to positive feedback (Rew-P) can thus easily be perceived as a weaker negativity. Crucially, no reward response in the basal ganglia is triggered by negative outcomes. Therefore, there should be no Rew-P and, thus, no notable attenuation of the simultaneous N2. If one visually compares the components during positive and negative feedback in the FND time range, one gets the impression that negative outcomes elicit a stronger negativity than positive outcomes—even though this presumably actually reflects a reduced negativity for positive feedback, caused by the underlying Rew-P (Baker & Holroyd, 2011). This alternative electrophysiological explanation of the occurrence of the FND would have implications for our understanding of the psychological mechanisms of outcome assessment: The key driver of the well-known electrophysiological differentiation of negative and positive feedback might be our response to positive outcomes.

Another limitation of our knowledge on the FND concerns potential situational moderators. A close look at the literature reveals that an FND is usually reported when unambiguous, easily discriminable feedback was presented (e.g., plus and minus signs). Clearly, learning under every-day conditions does not rely on only one type of feedback per outcome valence, but rather comprises a wide array of options. For example, we can affirm someone's behavior with an almost endless number of phrases of praise (e.g., yes, good, perfect, correct, right, that's it, absolutely) and reprehension (e.g., no, wrong, incorrect, fault, not at all). Likewise, social feedback can arise from various different signals such as several potential facial expressions of affirmation and disapproval (Pfabigan, Alexopoulos, Bauer, & Sailer, 2011; Zhang, Li, Qian, & Zhou, 2012). Interestingly, when slightly more variable, rather than binary, feedback was presented in previous studies, the FND was found to be severely diminished (Liu & Gehring, 2009; Liu, Nelson, Bernat, & Gehring, 2014; Mars, de Bruijn, Hulstijn, Miltner, & Coles, 2004; Pfabigan, Zeiler, Lamm, & Sailer, 2014).

In *Publication IV*, we directly addressed both of the above concerns. Most importantly, we tested whether the ambiguity of feedback signals affected ERP responses. We argued that the neural encoding of feedback should be fundamentally different if the exemplars representing the feedback are completely unambiguous than if the determination of their meaning requires more effort. If the same two words repeatedly occur (e.g., yes vs. no), minimal perceptual cues, such as the first letter or the rough shape of the word, should suffice to access the outcome value of the word, that is, whether it represents positive or negative feedback. The same type of quick extraction of feedback valence should not occur for feedback delivered by a more variable array of exemplars. To test this idea, we asked participants in Experiment 1 to deduce the ostensible, but actually nonexistent rule behind the visual feedback they received for pressing either of two buttons in response to colored digits or letters. As we were only interested in neural signatures of feedback processing, not of successful learning, there was no actual rule and, hence, no contingency between responses and feedback. Crucially, feedback was either presented in a blocked fashion (with one word for each feedback valence being used in several consecutive trials) or in a randomized order (such that feedback was drawn at random from five different feedback words per valence). We found no FND whatsoever for ambiguous feedback, that is, for feedback for which the specific exemplar of the valence category was drawn randomly and thus less predictable. This suggests that an easy discrimination of feedback stimuli is necessary to elicit an FND.

Perhaps most importantly for the present thesis, the design of Experiment 1 in *Publication IV* additionally allowed us to assess whether the selective occurrence of an FND in the blocked, but not in the random, condition was due to an increased negativity for negative feedback (Miltner et al., 1997) or an enhanced positivity for positive feedback (Eppinger et al., 2008). We found clear evidence that the latter was the case, with a marked Rew-P for blocked positive feedback standing out from all other conditions, obviously causing the FND. We thus strongly support the assumption that the differentiation of feedback in the FND time range relies on an increased reward response rather than on an error signal (Foti et al., 2011), which might tentatively be called a positivity bias in feedback processing.

Two additional results in *Publication IV* should be discussed. First, we found that the frontocentral P2, often interpreted as an index of top-down attentional selection (Anllo-Vento, Luck, & Hillyard, 1998; Carretié, Hinojosa, Martín-Loeches, Mercado, & Tapia, 2004; Potts, Patel, & Azzam, 2004), was markedly enhanced for negative feedback in the random, but not the blocked condition. We propose that this reflects a bias to detect negative feedback under conditions of ambiguous outcomes. Previous evidence suggests that uncertainty triggers the anticipation of negative events (Sarinopoulos et al., 2010) and preferential processing of negative information (Xu et al., 2011). This is in keeping with the notion that the familiarity and safety of our environment determine whether we exhibit general tendencies toward approach or avoidance behavior (Peeters & Czapinski, 1990). Fittingly, we conceptually replicated this idea in Experiment 2, in which we monitored neural activity during the processing of valence in general, that is, not restricted to feedback processing. When participants performed a speeded classification of affective words, we again observed an increased frontocentral P2 for negative as compared to positive stimuli, but only in a highly variable, ambiguous context, not in an unambiguous setting.

Another remarkable finding in *Publication IV* was that the LPC, a distinct, sustained positivity presumably reflecting controlled, motivated attention (Hajcak, Dunning, & Foti, 2009; Weinberg & Hajcak, 2011; Zhang et al., 2013), was not affected by feedback ambiguity. It thus appears to be a robust and consistent index of outcome evaluation, irrespective of variations in basic encoding processes of the exemplars representing the feedback. Interestingly, the LPC was larger for negative than for positive feedback, which is in line with some reports on feedback-induced LPC modulations (Frank, Woroch, & Curran, 2005; West, Bailey, Anderson, & Kieffaber, 2014; West, Bailey, Tiernan, Boonsuk, & Gilbert, 2012), yet at odds with others (e.g., Pfabigan, Alexopoulos, Bauer, Lamm, & Sailer, 2011; Pfabigan et al., 2014). Although this is tentative and clearly speculative, one might explain this divergence with different foci during the assignments in different studies. Whenever participants are—inadvertently or on purpose—prompted to use their *mistakes* as the primary source of information during learning, the LPC might be enhanced for negative feedback. Whenever the detection of *correct* responses is emphasized, this should lead to an enhanced LPC for positive feedback.

Our main conclusion based on *Publication IV* is that specific context factors, such as the ambiguity of feedback, can strongly affect how we process outcomes. Our general focus or attention toward signals of a certain valence shifts as a function of the variability of current feedback information. Moreover, the most well-established ERP signature of feedback processing, the FND, appears to primarily reflect how neural structures coding for reward, but not those that represent punishment, differentially respond to positive and negative feedback. All of these conclusions have serious implications not only for our understanding of outcome processing in dedicated and controlled learning tasks such as the one used in *Publication IV*, but also for the interpretation of reinforcement learning in a social context. Norms appear to be appraised much like simpler types of outcome information, which evidently serves as a neural representation of the social correctness of our own behavior. To fully grasp and recognize this process, however, an advanced understanding of the basic principles of feedback processing underlying reinforcement learning is necessary. *Publication IV* serves this exact purpose as it allows us to apprehend other scholars' and our own studies on reinforcement mechanisms of social influence more adequately.

4.2.2.3 The Functional Significance of the LPC

As explained above, the most robust ERP signature of feedback processing in *Publication IV* turned out to be the LPC, which was sensitive to the valence of feedback irrespective of the current context. Notably, both positive and negative feedback have been reported to lead to an enhancement of the LPC in previous studies (see, e.g., Pfabigan et al., 2014; West et al., 2014), suggesting that it does not code outcome *valence* (Yeung & Sanfey, 2004). To understand its functional significance, it has turned out to be particularly instructive to consider studies investigating affective and social processes. For example, both positive and negative words, as compared to neutral ones, elicit an enhanced LPC

(Schacht & Sommer, 2009). Likewise, the LPC is very similar for pleasant and unpleasant images, while it is larger for both than for neutral ones (Cuthbert, Schupp, Bradley, Birbaumer, & Lang, 2000; Schupp et al., 2000). Attractive faces lead to larger LPC amplitudes than unattractive ones (Johnston & Oliver-Rodriguez, 1997; Zhang & Deng, 2012). Stimuli providing information regarding another person's behavior in a social-dilemma game have been found to elicit a larger LPC than socially non-informative cues (Bell et al., 2015). Also, explicit prejudice against out-group members is associated with an increased LPC during the processing of out-group, as compared to in-group, faces (Ito, 2004). All of the above findings can be reconciled by assuming that the LPC is sensitive to the motivational significance of stimuli (Weinberg & Hajcak, 2011). It appears to reflect sustained attentional prioritization and an enhanced perceptual analysis of relevant sensory input (Schacht & Sommer, 2009). This is in keeping with the neural source of the LPC in the context of affective processing, which comprises a network of cortical areas related to attentional processes, such as the ventrolateral prefrontal and cingulate cortex, as well as subcortical structures associated with affective processing, such as the amygdala (see Liu et al., 2012).

Publication V supports the above notion regarding the functional significance of the LPC based on a study on social perception. By presenting images of nonverbal displays whose social relevance to the observer should vary and measuring LPC amplitude as the dependent variable, we provided novel data that might elucidate whether the component indeed serves as a reliable indicator of attentional orienting reflecting the social motivational significance of a stimulus (Weinberg & Hajcak, 2011). The starting point of our investigation was the assumption that our capacity to process and interpret nonverbal expressions has evolved to favor the perception and analysis of ecologically relevant displays over less significant ones (see, e.g., Zebrowitz & Collins, 1997). In keeping with this idea, it has been theorized that displays of supremacy and dominance are particularly relevant cues that are readily detected as part of an adaptive strategy of person perception (Furley & Schweizer, 2015; Rule, Adams, Ambady, & Freeman, 2012; Tracy, Shariff, & Cheng, 2010). The potency of this essential human faculty can be demonstrated quite strikingly in the context of sports. Even naïve observers are able to accurately estimate sports scores on the basis of thin slices of behavior (Ambady & Rosenthal, 1992), that is, using minimal perceptual cues such as brief video clips of intermissions during sports competitions (Furley & Schweizer, 2014). This capability relies on observers' quick and accurate perception of dominance during the examination of athletes (Furley & Schweizer, 2015). Following this logic, we applied ERPs—for the first time in this specific context—and asked participants in the experimental study reported in *Publication V* to estimate the current score based on black-and-white images of athletes presented for 500 ms. These images depicted athletes during game intermissions and did not involve any explicit signs of victory or loss, such as raised arms, covered faces, or tears. The athletes shown in the images were currently either in the lead, in a draw, or trailing. Based on ecological theories of person perception, it stands to reason that the most motivationally relevant images should be those associated with dominance and supremacy (Furley & Schweizer, 2015), that is, the images of athletes currently leading. Assuming that the LPC tracks the degree of attentional prioritization of relevant information (Weinberg & Hajcak, 2011), its amplitude should be largest for images showing athletes currently leading, especially as compared to those who were currently trailing.

First, we replicated previous findings that amateur observers are indeed able to roughly estimate the score of a sports competition based on thin-slice images of athletes (Furley & Schweizer, 2014). Although participants in our study did not reliably differentiate the score for athletes in a draw and for leading ones, they were clearly able to discriminate the score in images from these two categories from the score in images displaying athletes who were currently trailing. Most importantly, however, our ERP hypothesis was neatly confirmed, as we did not only observe the typical posterior LPC during the observation of the to-be-evaluated images, but also a linear increase with increasing score in favor of the displayed athlete: LPC amplitude at medial and right parietal recording sites was largest for images of leading athletes, smaller for athletes currently in a draw, and smallest for images of trailing athletes. Not only does this reflect the assumed subtle differences in motivational relevance of the nonverbal cues, but its topography also closely matches previous findings of a central-to-posterior peak of the LPC (Kuhr, Schomberg, Gruber, & Quirin, 2013; Weinberg & Hajcak, 2011) with a slight right-lateralization during evaluative judgments (Cacioppo, Crites, & Gardner, 1996). Interestingly, the LPC increase for leading as compared to trailing athletes was not only substantial and significant, but also related to individual score-estimation performance across participants. A strong LPC-based differentiation of leading and trailing athletes was associated with more accurate estimations of whether an athlete was currently in the lead. This indicates that the LPC modulation was not a byproduct or side effect, but rather an index of a relevant neural process that fed into overt behavior.

One might object that we cannot fully rule out that something other than the motivational relevance of the nonverbal cues in the images elicited the observed LPC modulation. It cannot be a trivial aftereffect of physical differences between the images of the different categories, as they were matched for relevant physical attributes, such as luminance, contrast, and spatial distribution. However, one might argue that the richness or expressiveness of nonverbal cues in leading athletes might have necessitated an increased effort to encode the images. However, two arguments against this alternative explanation can be brought forward. First, even if one were to assume that nonverbal behavior displayed by leading athletes is more expressive or complex than that of trailing athletes, it still seems highly unlikely that the richness of nonverbal expressions declines gradually with decreasing score. If anything, it is conceivable that athletes currently in a draw, who are supposably in a less emotional state than during leading or trailing, should display the least expressive cues, which should result in the smallest LPC amplitude for draw images if expressiveness was indeed the driver of the LPC modulation. What we did find, however, was a stepwise reduction in LPC amplitude from lead to draw to trail images. Thus, it would be hard to reconcile the expressiveness account with our exact pattern of results. Second, we found a significant relation between score-estimation performance and LPC modulation, indicating that behavioral responses in our social-perception task were functionally related to processes indexed by the LPC. Crucially, dominance is readily perceived from sports images (Furley & Schweizer, 2015) and is considered particularly prominent and relevant in person perception (Rule et al., 2012), making it parsimonious to assume that it was the dominance-based relevance of images of leading athletes that led to a stronger LPC.

One might further object that we cannot rule out the possibility that images of leading athletes led to increased motivated attention simply because participants' internal representation of the score-estimation task made them understand these images as the target images. Although participants rated the images on a bipolar scale, they might have reframed their assignment as the search for signs of victory. Therefore, it might not have been the leading athletes' general significance in social perception and interaction, but rather their momentary significance to solve the task that made them stand out. While this is indeed a valid objection and further research should address this notion, it does *not* question the functional significance of the LPC, which is of paramount interest to the present thesis. This alternative account likewise implies that the particular relevance of certain images is what drives attentional orienting, thus eliciting an increased LPC.

Publication V further corroborates the notion that the LPC is a sensitive index of sustained, motivated attention, especially in the context of elementary social and affective processes (see, e.g., Bell et al., 2015; Schupp et al., 2000). Higher-order neural hubs that code affective information or govern the distribution of attentional resources most likely provide top-down signals that feed, in a reentrant fashion, into structures coding basic perceptual features, thus enhancing the visual representation of motivationally relevant stimuli (see Liu et al., 2012; Schacht & Sommer, 2009). Admittedly, one might criticize that our conclusion is somewhat circular: We confirm the assumed motivational relevance of images of leading athletes by use of the LPC and, in turn, interpret this as confirmation of the LPC's involvement in indicating motivational relevance. However, our study should by no means be regarded in isolation. Our predictions were directly and firmly based on previous work in the field of nonverbal behavior, as was our specific expectation regarding LPC amplitude enhancement. Thus, negating our interpretation of the LPC would essentially put our findings at odds with both lines of research, while an adherence to the proposition regarding the functional significance of the LPC as an index of motivated attention toward relevant sensory input provides a straightforward, parsimonious, and comprehensive explanation.

4.2.2.4 Social Deviance and Subsequent Attention to Norms

In addition to enhancing our understanding of the behavioral and electrophysiological signatures analyzed in studies on social proof and deviance (*Publications III, IV, V*), we also sought to advance the theoretical account itself. In a series of three ERP studies, we first conceptually replicated the notion that norms initiate the differentiation of social proof and deviance in neural structures coding behavioral outcomes. Moreover, we valuably extended this idea by considering potential downstream consequences. We used the advantages of ERPs to explore this neurocognitive mechanism of social influence and to further our understanding of its psychological foundation.

To validate our own capacities and opinions, we compare our own to other people's behavior (Festinger, 1954). There are, however, crucial boundaries to the extent and stability of this process, as laid out in Festinger's groundbreaking theory on social comparison. He proposed that, if we have the choice with whom to compare ourselves, we prefer to use those whose views are already similar to our own as standards of judgment. Moreover, if we apprehend that we hardly find social proof for our views, yet register a surplus of our deviance from a certain group, we will cease to compare ourselves with this group. In *Publication VI*, we considerably extended this notion, as we demonstrated that it is not only the objective degree of social proof that determines our inclination to compare ourselves with others. Instead, we found that a person's mere responsiveness to the perception of social deviance from a group was robustly associated with participants' propensity to shy away from this very group's subsequent judgments.

In the first part of the experiment reported in *Publication VI*, participants rated a series of images and were informed after each judgment about how a group of four previous participants, ostensibly drawn randomly from a prior study, had rated the same image. In the second part, participants were asked to rate a completely novel set of images, but they now saw the ostensible averaged ratings of the same group of participants *before* each to-be-rated image. Notice that participants did not reassess the same images, but made unrelated judgments in the first and second part of the experimental session. To make the impression of the group and its norms particularly vivid and memorable, we arranged four images ostensibly depicting the participants whose judgments we had aggregated around the averaged judgment that was presented in the center of the screen. The main focus of our ERP analysis was on the two moments when participants processed the group norms: after their own judgments in the first part and before their own judgments in the second part. In the first part, we hypothesized that participants would appraise the outcome value of the norm, that is, they would evaluate whether it constituted social proof of their own judgments or tagged it as socially deviant. In the second part, we were interest in how attentively participants processed the group's norms. Given that the ostensible group whose norms were shown in the first part was the source of the norms in the second part as well, we assumed that the perception of group norms in the second part would be biased based on the preceding experience of social proof and deviance in the first part.

First of all, we found that participants strongly aligned their judgments to those of the group in the second task, in which they knew the group's judgments before their own. As in very similar investigations (e.g., *Publication II*), this effect was most certainly of a truly social nature. Specifically, the slow timing of the stimuli, with approximately 2 s passing on average between the onset of the norm and participants' responses, makes it unlikely that the display of group norms merely caused a priming effect. However, the behavioral results were not our primary focus in this study. As hypothesized, in the first task, we observed a typical FND during the analysis of deviance from the norm relative to the processing of norms conveying social proof of the individual's view. At frontocentral electrodes, a larger negativity was found for norms disagreeing with the individual's judgment than for agreeing ones. Note that this need not necessarily reflect an increase in negativity in the deviance condition, but might actually reflect an enhancement of an underlying positivity (Baker & Holroyd, 2011; Holroyd et al., 2011). The actual source of this effect was, however, not of primary interest in this study.

We observed a strong parietal P3 during norm processing in the second task. The posterior P3 reflects the allocation of attentional resources (Polich & Criado, 2006), and its amplitude is often found to be enhanced for stimuli of high relevance to the observer (Gray, Ambady, Lowenthal, & Deldin, 2004) and for those requiring decisions or actions (Polich, 2007). It has been proposed that the P3 is fed by signals from the prefrontal, the anterior cingulate, and the orbitofrontal cortex, all of which are involved in coding the relevance of current sensations (Nieuwenhuis, Aston-Jones, & Cohen, 2005). Fittingly, the personal significance of a stimulus to an observer is reliably reflected in the amplitude of the P3 (see, e.g., Gray et al., 2004; Littel, Euser, Munafò, & Franken, 2012). Crucially, we found a significant negative relation between participants' neural response to social deviance in the first task (FN amplitude) and their attention toward the group norms in the second task (P3 amplitude): The stronger participants registered their deviance from the group according to the FN increase, the less intensely they processed this group's norms during a subsequent encounter, as reflected in a smaller P3. One might be skeptical of such a correlation, given that it reflects the association between two measures of brain activity induced by similar events during the same experimental session. To rule out several trivial explanations of our finding, we conducted a series of control analyses. First, we demonstrated that attention to group norms in the second task was not associated with neural feedback responses in general, but specifically to the response to norms indicating social *deviance*: Subsequent attention to the group was not at all related to the previous neural response to norms conveying social *proof*. Second, we demonstrated that the deviance-induced FN was not associated with the P3 in general, but specifically with P3 amplitude during subsequent norm processing: There was no correlation between first-task FN and the P3 that was registered during the processing of the to-be-evaluated images. Third, we still found the association when using individual FND scores by subtracting proof-induced amplitude in the FN range from deviance-induced amplitude in this time range, instead of using raw FN amplitudes triggered by norms indicating deviance. Fourth, we still replicated the correlation when we controlled for the amplitude of the P3 during norm processing in the first task, thus ruling out the possibility that the association between first-task FN and second-task P3 was only due to a trivial association between the P3 in the second task and a similar P3-like positivity covertly underlying the FN in the first task.

Note that we explicitly aimed to monitor participants' *subjective* perception of their own deviance from social norms. Therefore, we necessarily focused on the naturally occurring variability of FN amplitudes across participants. A downside of this approach is, however, that the findings in *Publication VI* were correlational, preventing us from drawing firm conclusions regarding causal relations between our variables. Due to the temporal structure and organization of the experiment, we can safely preclude that attention to the norm (second part) affected the processing of social deviance (first part). It is, however, possible that a third variable is linked to both parameters of interest and mediates the observed association. For example, rather trivial characteristics of our participants, such as their general motivation, interest, or neural excitability might let us find an association between two different measures of event-related brain activity. Alternatively, personality features, such as (social) anxiety or neuroticism, might mediate the observed correlation between two neural responses to social information. However, the specific direction of our correlation makes it extremely unlikely that any of the above was the latent mediator of the association. If anything, the concomitant measurement of a person's overall responsiveness to norms—in terms of general physiological excitability or perhaps due to a personal interest triggered by anxiety or similar traits— should let us find a *positive* relation between two responses to social norms. What we did find, however, was that strong responses to social deviance were associated with weak attention to later group norms. Therefore, in the absence of a striking and convincing alternative explanation of our findings, we tentatively posit that strong responses during the perception of one's own deviance from a norm entail a direct change in how intensely the respective group's norms are subsequently processed. This is in line with Festinger's (1954) assertion that the processing of others as a source of comparison is contingent upon the sufficient detection of agreement between our own and their judgments. It seems that this principle can be monitored at the level of basic cognitive processes, such as the deployment of attentional resources. In fact, our own findings go beyond Festinger's idea by suggesting that it is not only the actual degree of disagreement, but also the subjective experience, one's current or general sensitivity to cues of social deviance, that shapes one's later focus on a group's norms.

Interestingly, we found that the correlation between the neural response to a norm's outcome and the subsequent deployment of attentional resources was carried by the response to social deviance, not to proof. First, this indicates that the detection of dissent does play an important role in the representation of social norms. If the FND was exclusively carried by a Rew-P induced by consent with others, the raw FN would mainly reflect an N2-like negativity, presumably unrelated to outcome evaluation, that is not attenuated by a simultaneous positivity (Holroyd et al., 2011). Seeing as the FN (not only the FND, and not the Rew-P at all) was strongly associated with subsequent processing in a functionally meaningful way, the FN most likely indeed reflected an outcome-related process. This supports recent claims that neural responses to social deviance are highly relevant signals in the assessment of norms (Kim et al., 2012; Klucharev et al., 2009; Shestakova et al., 2013). However, this does not imply that this is always the case; rather, the circumstances of one's confrontation with a norm might determine whether deviance-induced, proof-induced, or both types of signals are the key drivers of norm evaluation.

4.2.2.5 Attentional Signatures of Social Proof and Deviance

When we observe a social norm, we obviously assess whether it signals social proof or deviance, which we represent in the same way as behavioral outcomes, such as typical reinforcers. Norms thus allow us to determine the appropriateness or incorrectness of our own behavior in relation to the social standard. It seems highly likely, however, that our immediate translation of norms into an outcome value similar to performance feedback, reward, or punishment triggers additional processes beyond this assessment. Downstream consequences of the early recognition of value need to be considered. There is consensus that the reward we associate with a stimulus feeds directly into attentional processes, exerts a profound influence on our perceptions, and impacts even higher-order cognitive functions (Anderson, 2013; Anderson, Laurent, & Yantis, 2011; Della Libera & Chelazzi, 2006; Serences, 2008). The reinforcing character of social norms might instigate a similarly constructive deployment of attentional resources in order to favor information that promises an improvement of one's resources or chances at success (see Anderson, 2013). In line with this idea, we noticed remarkable, yet previously untested and unreported modulations of the LPC in several ERP studies on social proof and deviance (*Publication VI*; Huang et al., 2014; Kim et al., 2012). As noted earlier, there are several reports of LPC variations during the processing of feedback (e.g., Frank et al., 2005; Pfabigan, Alexopoulos, Bauer, Lamm, & Sailer, 2011; Pfabigan et al., 2014; West et al., 2014; West et al., 2012). Yet, while it is unclear whether positive, negative, or both types of feedback enhance LPC amplitudes (see *Publication IV*), the variations during norm processing that we noticed in our own and other scholars' previous work consistently involved—at least at the descriptive level available to us upon visual inspection—increased amplitudes during the perception of social proof as compared to deviance.

In *Publication VII*, we followed up on this idea and investigated potential downstream effects of norm evaluation on attentional-selection processes. We used an adaptation of the two-shot paradigm (Klucharev et al., 2009), yet avoided the problem of RTM altogether by providing social-norm information in a way that is not tied to the level of participants' initial ratings (for an extensive discussion, see *Publication III*). Participants rated faces, depicted as portrait photographs, as to their attractiveness on a 3-point scale. After each judgment, they saw the ostensible percentage of previous participants who had rated the current image in the same way as the actual participant. We manipulated the ostensible percentage of agreeing others to be either below or above 50%, thus signaling that participants' ratings deviated from or were in line with the absolute majority. In an unannounced reassessment phase after the initial rating session, they rated all images again, yet did not receive feedback anymore. As expected, participants were more likely to adjust their rating of images for whose assessment they purportedly did not receive support from the majority as compared to those images for whose assessment they did receive confirmation from the majority of previous raters. In a paradigm whose measurements cannot be carried or inflated by RTM, we thus corroborate the claim that the perception of social proof determines the likelihood (and extent) of subsequent adjustments of individual judgment.

Again, however, our main focus was on the analysis of ERPs. First of all, we were able to replicate the finding of an FND during the perception of social deviance versus proof, thus substantiating previous results showing a feedback-like assessment of social norms. What makes this replication particularly valuable is that we found the effect in a different paradigm than previous studies (see, e.g., Shestakova et al., 2013). We did not oppose the individual's response to the group norm, both displayed in the same format (i.e., as two ratings on the same scale). Nevertheless, we still found that perceiving that the majority of other people does not agree with one's own assessment elicits a markedly different neural response than perceiving that they do. Obviously, cues as to how other people behave are reliably used as points of reference that are compared to one's own behavior. More importantly, we demonstrated for the first time that the LPC was significantly larger during the perception of social proof than during the perception of one's deviance from the norm. As explained in considerable detail earlier and supported by our findings in *Publication V*, the LPC serves as a reliable index of sustained attention toward motivationally relevant stimuli. Confirming our results in *Publication V*, we found that the peak of differential brain activity in the LPC range elicited by proof as compared to deviance was distributed broadly across the scalp, yet maximal at posterior sensors with a slight tendency toward the right hemisphere (see also Cacioppo et al., 1996). Our finding of an increased LPC for norms conveying social proof as compared to those conveying deviance shows that the representation of norms in terms of their outcome value obviously entails substantial modulations of essential cognitive functions.

In *Publication VII*, we performed another analysis that goes beyond what was previously done in research on social norms: We analyzed neural responses to the faces during the reassessment phase, that is, when participants saw the faces once more and gave their follow-up ratings at the end of the experiment. We hypothesized that faces on whose judgment one had previously agreed with the norm (proof-associated faces) should be processed differently during the reassessment than faces on whose judgment one had disagreed with the norm (deviance-associated faces). Confirming our expectations, we found that the posterior medial P2, an index of early perceptual attention that is particularly sensitive to the affective and motivational value of faces (Amodio, 2010b; Ito & Urland, 2005), was larger in amplitude for proof-associated than deviance-associated faces. We thus show for the first time that downstream consequences of the outcome deduced from norms and their (dis)agreement with the individual's judgments include not only modulations of immediate attentional processes triggered by the appearance of the norms, but also attentional effects on the much later processing of norm-related stimuli. This is in line with the general idea that previously perceived norms inform subsequent representations of the stimuli to which they pertain (Zaki et al., 2011).

The most relevant and provoking findings in *Publication VII* were modulations of late attentional selection during norm processing (LPC effect) and early attentional selection (P2 effect) during a later encounter with the items. Interestingly, both findings can be explained by a common cognitive principle: In the absence of any notable losses, rewarding stimuli are attended to particularly intensely (Anderson, 2013). This mechanism of value-guided attention has recently been demonstrated with social forms of reward, such as smiling faces (Anderson, 2015). Fittingly, perceiving one's agreement with a norm appears to entail an enhancement of sustained attention, presumably to thoroughly encode the stimulus which one obviously judges in a socially rewarding way (see Schacht & Sommer, 2009). Moreover, during a subsequent evaluation of the faces, we still seem to put particular emphasis on stimuli that are associated with social reward (i.e., those for whose assessment we perceived social proof). The latter finding is particularly striking, as it represents the first demonstration of continuing effects of norm-evaluations on basic cognitive functions.

We thus valuably extend earlier reports on the process of comparing social norms with our own judgments to determine the appropriateness of our own views (Kim et al., 2012; Klucharev et al., 2009; Klucharev et al., 2011; Shestakova et al., 2013) by showing that the output of this comparison, the evaluation of our own behavior, feeds into essential attentional-selection processes even much later. It seems as if our perception of a stimulus is guided by how well our own behavior toward this stimulus matches a social norm. It has been proposed that the P2 reflects the automatic prioritization of meaningful content (Carretié et al., 2004; van Hooff, Crawford, & van Vugt, 2011). This would suggest that even in the absence of control or intent, we might be driven to attend to stimuli promising consent with others more strongly than to those that are associated with social conflict. Clearly, we cannot fully determine on the basis of our experiment, in which only proof and deviance were compared without a neutral baseline, whether the P2 effect actually reflects increased attention toward proof-associated stimuli, decreased attention toward deviance-associated stimuli (similar to an evasion of objects promising dissonance and conflict), or both. Nevertheless, both ERP findings (LPC and P2) consistently indicate that attentional selection is sensitive to the reinforcing character of social norms. This is in keeping with the idea that the deployment of attentional resources follows value-based information and guides our system to encode potentially advantageous information (Anderson et al., 2011).

4.2.2.6 Social Proof Based on Observed Preferences

The neurocognitive basics of the representation of social norms relative to our own judgments seem somewhat robust given the fact that multiple studies using variable setups, stimuli, and group scenarios provided converging evidence of the involvement of processes related to outcome assessment (see *Publication I* for an overview). Notably, however, they all share a rather basic conception: The individual makes an overt judgment and is then confronted with the group's average or majority judgment. While this certainly captures and epitomizes the general idea of how we assess norms, it does not cover all aspects of how social influence actually works. Norms in this paradigm directly follow the individual's own judgment. Moreover, this often includes a placative juxtaposition of individual and group judgments, usually placing them next to each other on the same display (Huang et al., 2014; Kim et al., 2012; Klucharev et al., 2009; Shestakova et al., 2013). The temporal and spatial proximity of individual decisions and group norms certainly triggers a particularly—perhaps unrepresentatively—vigorous comparison of both pieces of information. When we learn about social norms in real life, though, this often happens in regard to stimuli that we did not evaluate only a second ago. Rather, we develop our own attitudes over time and are confronted with other people's judgments of a certain stimulus at a more or less unanticipated point in time. Also, we often do not learn about norms as perfected as those used in typical experiments, where they were conveyed in the shape of averaged group ratings, majority judgments, or the percentage of agreeing others. Instead, we often acquire our knowledge regarding social norms based on repeated observations of other individuals' judgments (Sherif, 1935).

In *Publication VIII*, we directly addressed the above concerns regarding the specificity of previous paradigms used to study the neurocognition of social influence. We adapted a paradigm from a previous fMRI study (Mason et al., 2009) for use in an ERP setting. The whole session was divided into four parts: initial assessment, observation, delay, and surprise reassessment. First, participants rated a set of 20 abstract symbols, similar to logos or icons, on a six-point scale ranging from *ugly* to *beautiful.* This allowed us to register participants' attitudes in advance, thus providing us with a baseline against which we could measure later judgments potentially affected by social influence. After this brief introductory rating phase, participants repeatedly observed how each of twelve other individuals had ostensibly rated the same set of abstract symbols. All 20 symbols were presented twelve times, each time in a newly randomized order. Each of these twelve cycles began with a visual notification that the judgments of a new participant from a previous sample would now be shown. The purported judgments were delivered in a straightforward dichotomous fashion and indicated by a colored frame surrounding each object. Given their intuitive connotations, we presented a green frame to indicate an ostensibly favorable attitude and a red frame to show dislike. By observing the judgments of these mock previous participants, our actual participants could learn that some objects were mostly liked (11 out of 12 times), while others were mostly disliked (11 out or 12 times). After a short delay, in which an irrelevant reaction-time task was administered to keep participants engaged, they rated all items again. Crucially, during the observation phase, we assessed ERPs to test whether participants processed items differently depending on whether the currently displayed individual judgment did or did not match their own attitude that had been reported earlier in the experiment in the first assessment phase. When participants initially liked a symbol and observed a mock participant's favorable rating of this symbol or when they initially disliked a symbol and observed someone else's unfavorable rating, this constituted a match. Mismatches occurred when other people's unfavorable judgments were observed for initially liked symbols or when others' favorable judgments were observed for initially disliked symbols. Thus, we again focused on the evaluation of other people's judgments relative to one's own stance (Klucharev et al., 2009), yet in a more natural context in which no artificially aggregated group judgments were presented and no overly suggestive temporal or spatial juxtaposition of participants' and other people's judgments was displayed.

We first conducted a behavioral pilot experiment to ascertain that the colors themselves did not affect participants' ratings. One might argue that green and red convey a certain valence even without any social connotation (Gil & Le Bigot, 2014). If this were the case, one should observe an influence of the repeatedly presented colored frames on follow-up ratings, as would be expected based on the well-established evaluative-learning effect: Simultaneously or consecutively pairing a neutral and a valenced stimulus multiple times is known to lead to substantial changes in the evaluation of the originally neutral stimulus to match the valence of the originally valenced stimulus (for an overview, see Gast, Gawronski, & De Houwer, 2012). Thus, if the colors of the frames communicated a valence by themselves, repeatedly presenting objects paired with either green or red frames should lead to systematic upward and downward changes, respectively, of the objects. In the pilot experiment, however, we found no trace of a significant influence of the frame colors when they were presented without a social meaning (non-social group). When the colors *did* represent the ostensible judgments of previous participants (social group), we indeed observed a substantial influence of the colored frames: Objects assumed to be liked by many others were rated significantly more favorably than objects assumed to be liked only by a single observer. Thus, the colors per se were no effective carriers of meaning and valence, conveying an evaluative information only when the observer assumed that they indicated the preferences of others.

In the main EEG experiment, we replicated the effect observed in the social group of the behavioral pilot experiment. In the previous paragraph, we have described this crucial behavioral effect as a main effect of the purported popularity on participants' post-learning ratings (baseline-corrected for their initial assessments of all objects). However, one can also capture the effect of the observed preferences as the *adjustment* of ratings varying as a function of whether or not other people's ratings ostensibly *matched* the individual's judgments. In line with previous research, we found that participants changed their judgments to match more closely those of the observed others. First of all, we noticed a general tendency of initially low ratings to be higher in the second phase (upward shift) and of initially high ratings to be lower in the second phase (downward shift). This pattern would be predicted based on RTM. Importantly, however, we found that the upward rating shift of initially disliked objects was significantly stronger for objects that were popular among others than for those that were unpopular with others. Likewise, the downward rating shift of initially liked objects was stronger for objects unpopular among others than for those that were popular among others. As RTM is independent of the ostensible popularity that is randomly assigned to a given object, this difference of shifts reveals a corrective influence of observed preferences on individual judgments.

Most importantly, at the level of ERPs, we also observed a substantial differentiation of ostensible other people's judgments that matched the real participants' initially reported attitudes (social proof) and those that did not match (social deviance). Observing other people's judgments that did not match participants' own was accompanied by an increased N400, broadly distributed across the scalp. Observing preferences that matched participants' own elicited a larger medial posterior P2 and an increased LPC. Notably, all three effects were based on a thorough differentiation of proof and deviance during the processing of initially liked items (i.e., the symbols that participants had initially rated favorably), while no significant ERP effect was observed for initially disliked items.

The N400 is usually studied in the context of language processing (Kutas & Federmeier, 2011), typically to monitor neural responses to breaches of semantic expectation (Kutas & Hillyard, 1980, 1984). However, increased N400 amplitudes have also been reported for more complex breaches of expectation, such as stereotype-inconsistent social information (**hehman2014**; Proverbio & Riva, 2009; van den Brink et al., 2012). Generally, N400 increases appear to reflect the increased effort of encoding discrepant pieces of information (Arzouan, Goldstein, & Faust, 2007; Obleser & Kotz, 2011). Only recently, it has been shown that social deviance in the classical paradigm (Klucharev et al., 2009) can elicit an N400 enhancement as well (Huang et al., 2014). Huang and colleagues (2014) proposed that the norm-based N400 modulation represents the same process of outcome evaluation based on the prediction-error principle that is usually indicated by the FND (e.g., Kim et al., 2012). This explanation is not entirely consistent with previous studies and theoretical accounts relating N400 enhancements to the representation of breaches of expectation (Kutas & Federmeier, 2011) and increased encoding effort (Obleser & Kotz, 2011), but not to reward prediction errors. We thus propose that our findings in *Publication VIII* support the notion that mismatches between one's own attitudes and social standards, such as norms or other people's preferences, are processed very much like violations of expectation (Chen et al., 2010). It seems likely that it is particularly demanding or cognitively costly to reconcile other people's views and one's own impression when they clearly contradict each other. One presumably attempts to update one's representation of the stimulus, or the value associated with it, in order to avoid the dissonance of contradicting the judgments of a relevant source. Interestingly, the increase in integration effort occurs both somewhat quickly during one's processing of a stimulus accompanied by a social reference and in the absence of one's own overt judgment.

As explained above, P2 and LPC are indices of early and late attentional selection, and both of them have previously been shown to be sensitive to the affective and social value of a stimulus (Amodio, 2010b; Carretié et al., 2004; Ito, Larsen, Smith, & Cacioppo, 1998; Schupp et al., 2000). Moreover, both components are known to track motivationally relevant information such as during the evaluation of feedback (see *Publication IV*). Fittingly, we found a larger LPC during the observation of social proof as compared to deviance and a larger P2 for faces previously associated with proof than for those associated with deviance in *Publication VII*. In *Publication VIII*, we confirm our own data and interpretation regarding the role of attentional processes in norm assessment. It seems as if we are inclined to prioritize stimuli that promise social reward in terms of consent with others on these stimuli's evaluation. Both rather early, perception-related attentional mechanisms (P2) and a late, sustained enhancement of encoding (LPC) reflect this bias toward items associated with social proof. This is in keeping with the general idea of value-driven attention in both social (Anderson, 2015) and non-social domains (Anderson, 2013).

We observed a significant P2 effect of social (mis)match in *Publication VIII*, in which no overly obvious confrontation of participants' and other people's judgments was presented. In previous studies, in which we used somewhat blunt juxtapositions of individual judgments and norms, we did not observe such an early P2 effect, and the earliest differentiation of proof and deviance was the FND around 300 ms (*Publications VI, VII*). One might object that it is utterly counterintuitive that a less obvious confrontation triggers an *earlier* differentiation of social proof and deviance than more obvious juxtapositions. Interestingly, however, we found a somewhat similar shift toward earlier effects in outcome processing when using ambiguous instead of unambiguous feedback in a simple learning task (*Publication IV*). Admittedly, the specifics of the early effects of outcome representation in the non-social learning task (*Publication IV*) were quite different from the effects induced by observing other people's preferences (*Publication VIII*): In the learning task, a frontocentral P2 was found to be increased for negative feedback, while a larger posterior P2 for positive outcomes was observed during the observation of other people's preferences. Nevertheless, both studies converge in demonstrating earlier, attention-related effects in situations in which the assessment of behavioral outcomes is not quite so simple. This might indicate a general principle of processing outcome information, such as norms or feedback words, more vigorously, attentively, and quickly under uncertainty or higher cognitive demand (see *Publication IV*). Whether positive or negative outcomes are prioritized during such early steps of analysis might depend on situational constraints. The idea of an early negativity bias in demanding feedback processing, yet an early positivity bias in demanding social-proof processing is in keeping with the observation that *negative* information generally appeared to attract more attention during non-social feedback processing, while *positive* outcomes seemed to be favored during norm processing: Larger LPC amplitudes indicating enhanced late attentional selection were found for negative as compared to positive feedback words (*Publication IV*), yet for positive as compared to negative social outcomes (*Publications VI, VII, VIII*). This might reflect a true difference in how we assess reward or punishment for our actions in a non-social versus social context. Alternatively, this might simply be a side effect of different outcome ratios. In our learning task with simple feedback words (*Publication IV*), negative feedback occurred slightly less frequently than positive feedback, presumably making the former a more informative event. In contrast, in our studies on social norms (*Publications VI, VII, VIII*), negative outcomes (norms conveying social deviance) were as frequent as positive outcomes (norms conveying social proof) or even more frequent. This might have reduced or even overturned a bias toward negative feedback (see Conclusion and Outlook for further discussion of this issue).

Finally, I would like to return to our finding that social proof and deviance were thoroughly differentiated for items toward which the individual held a positive attitude, while no such effects were observed for stimuli that participants initially expressed to dislike (*Publication VIII*). Four potential explanations of this asymmetry are conceivable. First, we might generally be inclined to find social proof for our selective preferences, not for our rejections. Therefore, we might attend to social standards regarding our own favorable opinions, but not our less relevant unfavorable views. Second, the comparison of another person's judgments and one's own might not primarily rely on one's stable attitude toward the currently observed item, but on the retrieval of one's earlier explicit assessment of it. Positive judgments might simply be memorized or retrieved better than negative evaluations. Third, previous studies suggest that social deviance is particularly salient whenever a norm indicates that an individual judgment is lower—that is, more skeptical, less indulging—than a social reference (Huang et al., 2014; Kim et al., 2012). Crucially, if one compares one's own judgment with a single other individual's (as in *Publication VIII*), one might use *one's own judgment* as the fixed reference against which the other person's responses are compared. Thus, the typical roles of to-be-reviewed judgment and social reference, usually assigned to the participant and the ostensible group, respectively, might be reversed. Hence, social deviance for items initially liked by the participant (i.e., items rated less favorably by the ostensible other individual than by the actual participant) would represent the case of judgments being lower than the point of reference. Given the potential reprehensibility of downward deviations in evaluative judgments (Kim et al., 2012), this might be particularly salient, leading to a careful differentiation of social proof and deviance. Fourth, due to the intuitive and otherwise unproblematic choice of colors, deviation from initially favorable views were always indicated by a red frame surrounding an object. Red might simply have signaled social deviance more distinctly, leading to strong effects of deviance versus proof only in these trials, while observed upward deviations, conversely indicated by a green frame, might have been less expressive. All of these possibilities do not question the overall interpretation of our findings in *Publication VIII*, yet they should be carefully tested in future studies.

5 Conclusion and Outlook

Social influence clearly is one of the classical issues of (social) psychology, and to this very day, its importance both in basic and applied research remains eminent (Cialdini & Goldstein, 2004; McDonald & Crandall, 2015). The contemporary approach of studying the neural functions and, relatedly, the steps of cognitive processing involved in it provides a novel perspective to our understanding of the elementary principles of social influence (Izuma, 2013) as well as possible applications thereof (Stallen & Sanfey, 2015). Not only do we know more about general micro-level processes and mental operations that contribute to our susceptibility to social influence (Falk et al., 2012), but we also begin to develop a refined picture of who is likely to be influenced (Falk et al., 2014), with which consequences (Cruwys, Bevelander, & Hermans, 2015), and to which potential benefit (Falk et al., 2010). One of the achievements of this line of research might be that the pervasiveness and effectuality of social influence seem slightly more comprehensible given our insight into which profound, well-trained, and highly adapted neurocognitive processes mediate the impact of other people on the individual (Campbell-Meiklejohn et al., 2010). Western societies in particular tend to view conformal behavior as a sign of weakness (Bond & Smith, 1996; Markus & Kitayama, 1994) and do not tire of emphasizing its potential harm (Albert, Chein, & Steinberg, 2013; Milgram, 1963; Rose, Bearden, & Teel, 1992; Steinberg, 2008). But even though it is neither among its goals nor within its rights for psychology to blindly exculpate individuals who abide by questionable norms, a deeper understanding of the intricate principles of social influence might allow us to adopt an open, less judgmental approach to what is sometimes, yet not always, a condemnable facet of behavior. Also, in order to deal with those excrescences of social influence that are indisputably dangerous, an advanced understanding of its mechanisms would clearly be beneficial. Beyond that, we might even be able to use our knowledge to actively apply social influence in service of objectively desirable outcomes. Interestingly, the transfer of fundamental social-influence knowledge to real-world issues is neither far-fetched, nor a sales pitch of basic researchers. Recent studies have convincingly demonstrated that social influence shapes even relevant aspects of individual behavior in real-life situations (Bond et al., 2012; Nook & Zaki, 2015). Moreover, we are currently beginning to fathom the impressive potential of social-influence campaigns that are directly informed by neurophysiological data (Cascio, O'Donnell, Bayer, Tinney, & Falk, 2015; Falk et al., 2016; O'Donnell & Falk, 2015). It thus seems that expanding our knowledge of the neural and cognitive principles of social influence is as topical as it is relevant and promising.

5.1 Extending our Understanding of Social Influence

In the present thesis, I have largely focused on the influence of social norms on individual behavior. Typically, two basic types of norms and their respective influences are distinguished (Cialdini et al., 1990). However, it should be noted that the distinction of informational influence generated by descriptive norms and normative influence generated by injunctive norms, although longstanding and established, is difficult or even artificial at times (Cialdini & Goldstein, 2004) and was not of paramount interest at this early point in our line of research (see also Klucharev et al., 2009). Therefore, we did not specifically aim to induce only one specific type of influence while fully controlling for any contribution of the other type. Nevertheless, the obvious characteristics of our tasks and experimental setups suggest that we have covered almost exclusively *descriptive* social norms, seeing as we (a) used simple depictions of majority behavior without applying normative pressure such as having participants and confidents interact, and (b) asked participants to perform rather special tasks—for example, evaluating ambiguous images—that might have boosted the informational value of norms (*Publications II, VIII*). The same arguments can be made for many previous studies (e.g., Chen, Wu, et al., 2012; Germar et al., 2014; Huang et al., 2014; Kim et al., 2012; Klucharev et al., 2009; Klucharev et al., 2011; Trautmann-Lengsfeld & Herrmann, 2013), which is why our knowledge on the neurocognition of social norms is presumably based mostly on descriptive norms. This should be noted early-on and kept in mind to appreciate the scope and limits of the following discussion of other scholars' and my own research.

To date, we know that the neurocognition of social influence based on descriptive norms essentially comprises (at least) two different mechanisms: modified representations of to-be-judged features and the representation of social proof and deviance as behavioral outcomes. While the former is particularly relevant in situations resembling the frequently used one-shot paradigm with norms preceding one's analysis of the to-be-judged stimulus, the latter applies to scenarios modeled by the two-shot paradigm. Therefore, our focus of analysis and the specific configuration of a social interaction largely determine which process applies. As mentioned in the introduction, the two accounts are not mutually exclusive, but rather complementary. Zaki and colleagues (2011) used a two-shot paradigm, yet focused on the neural representation of the to-be-evaluated images during the reassessment phase, in which a direct modulation of the essential computation of value could be observed. Similarly, we combined a two-shot paradigm including the typical analysis of reinforcement-related responses during norm processing with the idea that basic attentional functions during stimulus processing in the reassessment phase might be informed by preceding norms' outcome value (*Publication VII*). These two studies thus neatly exemplify the coexistence and compatibility of both approaches to social influence (see also *Publication I*).

5.1.1 The Fusion of Actual and Norm-Conveyed Input

When norms precede our judgments, they obviously have the potential to modify our internal representation of what is to be judged, thus exerting a profound influence on how we perceive the world (*Publication I*; Germar et al., 2016; Germar et al., 2014; Koban & Wager, 2016; Zaki et al., 2011). Moreover, the overall depth of our representation of to-be-judged information is a key determinant of how norms affect our decisions (*Publication II*), presumably moderating the direct influence of norms on our mental representations of to-be-judged items (Germar et al., 2016). However, not only norms themselves (i.e., their actual content) alter essential neural computations, but also the reinforcing value of a norm (i.e., its status as rewarding or punishing). Stimuli associated with social proof appear to be processed more intensely not only during the actual moment in which we learn about the norm in association with the item to which it pertains (*Publication VIII*), but also when we subsequently encounter this item again (*Publication VII*). Already at the level of early perceptual attention, we display a bias toward consent-related stimuli. This might be an instance of how value drives our attentional and perceptual system (Anderson et al., 2011; Serences, 2008). Several studies thus concordantly demonstrate how the most basic cognitive processes are already informed by other people's behavior, which fiercely reiterates the assertion that social forces substantially shape our take on reality (Deutsch & Gerard, 1955). One might have assumed that there is a long-lasting coexistence of our very personal representation of the world around us and other people's impressions, and that both sources of information are weighed and reconciled at a late stage, such as shortly before one opts for a certain judgment. This does, however, seem utterly unlikely, given the abovementioned findings. Our own perceptions and those of others appear to fuse at an early point, such that even the most basic neural operations are functions of both the input provided by our sensory system in regard to the to-be-judged object and what we learn from relevant social sources. This is well in line with the old concept of human suggestibility (Hull, 1933), which actually inspired some of the earliest research into social influence (Asch, 1955). But while classical sociopsychological research in the tradition of Asch and Sherif has ventured more into the motives and antecedents of conformity (Cialdini & Trost, 1998; Deutsch & Gerard, 1955), it is only the contemporary, neuroscientific approach to social influence that revisits and expands this concept.

Interestingly, the direct impact of social norms on how we represent to-be-judged features might be a neurocognitive explanation of one of the types of conformity explained in the introduction, namely conversion (Zaki et al., 2011). As noted earlier, while we sometimes only comply publicly to maintain a group's grace, we sometimes truly and internally adopt social norms. Descriptive norms in particular trigger these revisions of our initial stances (McDonald & Crandall, 2015). Fittingly, all studies reporting actual modifications of the representation of to-be-judged features employed scenarios that included simple descriptive norms most likely exerting informational influence. But the fact that social influence entails a consolidation of actual (sensory) and purported (norm-based) input should not be seen as a sign of weak dependence. Rather, it suggests that our take on reality is not unbiased. Just as we use gating and filter mechanisms such as selective attention to somehow cope with the endless stream of input, relevant social norms seem to be superimposed upon our own representations, much like a filter that adjusts and shifts the shape of a function. As our conspecifics are often reliable sources of information (Deutsch & Gerard, 1955), we might have developed an adaptive system that incorporates such potentially beneficial affluxes into our representation of reality (see also Claidière & Whiten, 2012). After all, there is broad consensus that, in a general sense, higher levels of information, such as expectations and prior knowledge, thoroughly infuse earlier levels of representation in a top-down fashion (Friston, 2005; Stokes, Atherton, Patai, & Nobre, 2011; Summerfield & Egner, 2009), thus presumably increasing both the performance and the efficiency of our neural apparatus (see Rao & Ballard, 1999).

In light of the above observations, I would like to return to a remark that I made in the beginning of this synopsis: To many, social influence might have an air of triviality. Clearly, conformity is often obvious and easy to spot, as demonstrated most strikingly in some of the earliest studies in this field (Asch, 1951). I do believe, though, that this newly emerging line of research into social influence, including our own findings regarding the effects of observed judgments on attentional processes (*Publications VI, VII, VIII*), strongly suggests that part of the impact of others on ourselves is actually quite subtle. Perhaps without our knowledge, our perceptual analysis of sensory input, our formation of memories, and our valuation of objects are systematically affected by preceding social norms (Edelson et al., 2011; Germar et al., 2016; Germar et al., 2014; Zaki et al., 2011). What is more, we are apparently even biased during the processing of norms and associated stimuli, focussing more on episodes conveying consent between ourselves and the norm than on episodes that signify our deviance from the social standard

(*Publications VI, VII, VIII*). The reach of social influence thus clearly appears to exceed its openly visible grasp, making it a little less trivial.

5.1.2 The Reinforcing Character of Social Norms

The notion that norms initiate the differentiation of social proof and deviance in neural structures coding behavioral outcomes (Klucharev et al., 2009), as well as the idea that this makes reinforcement a key principle of social influence (Izuma, 2013; Klucharev et al., 2011; Shestakova et al., 2013), have received much attention in the past few years. Not only are we able to convert norms into cues conveying behavioral outcomes, but we are even able to perform this conversion remarkably quickly, as value-related signals during norm processing occur as early as 250 ms after the onset of the norm (*Publications VI, VII*; Chen, Wu, et al., 2012; Kim et al., 2012; Shestakova et al., 2013). Moreover, the outcome signal, presumably generated in cortical and limbic neuron populations whose cell bodies are located in midbrain areas (Klucharev et al., 2009), obviously feeds into a variety of subsequent processes. First, it might directly inform upcoming behavioral adjustments (Klucharev et al., 2009; Klucharev et al., 2011; but see Kim et al., 2012; Shestakova et al., 2013), which corroborates the notion that conformist behavior is the consequence of reinforcement learning (see *Publication I*). Second, the allocation of attentional resources depends on norm-triggered outcome signals (*Publications VI, VII, VIII*), both immediately during the analysis of the norm, which can, but need not, include the simultaneous encoding of the specific stimulus to which the current norm pertains, and during later encounters with norm-related stimuli. As noted earlier, this is in line with the idea that attention works directly as a function of the value associated with a stimulus (Anderson et al., 2011). This extension of the concept of reinforcement principles in social influence relies primarily on the research that my colleagues and I have performed. It expands the scope of this theoretical approach by adding downstream consequences of complex social value signals and suggests that norms exert a subtle influence on how we select and prioritize sensory input. It therefore emphasizes once more the potency of norms to shape how we perceive and represent the world.

We have added a novel perspective in regard to the situational requirements and determinants of how we process other people's judgments in relation to our own (*Publication VIII*). We have shown that not directly juxtaposing a person's own and another person's behavior does change the processes involved, as we did no longer find an FND during the observation of deviant as compared to matching judgments. However, we still found neural signatures of the differentiation of judgments conveying social proof and judgments conveying deviance at early, mid-latency, and late stages of processing. The outcome generated by the rapid comparison of one's own attitude and a social reference was registered during early attentional selection; as variations of the effort to reconcile semantic input; and during the sustained prioritization of stimuli, potentially affecting perceptual encoding. Thus, the general concept that the human mind quickly and thoroughly differentiates social proof and deviance still applies, despite considerable changes to the scenario. During the mere observation of other people's judgments, which is most likely the typical format of how we apprehend social norms (Sherif, 1935), we appear to engage in the evaluation of our own beliefs in light of social references (and potentially vice versa; see *Publication VIII*) just as much as during the assessment of aggregated norms directly juxtaposed to our own decisions.

Most importantly, the effects of social proof on attentional-selection processes suggest that agreeing with a norm is a desirable or at least particularly informative type of social outcome (but see Anderson, 2016). For feedback processing in general, there has recently been a shift to emphasize the importance of reward signals over error or punishment responses (*Publication IV*). Likewise, several of our findings indicate that we are attuned to prioritize stimuli associated with social proof, rather than being in a constant mode to detect our own errors and the need for adjustment (Ridderinkhof et al., 2004). Our data indicate that we do not encode episodes of disagreement, which should eventually trigger conformal adjustment, more thoroughly than those of agreement. On the contrary, we seem to encode with particular emphasis the moments and stimuli associated with consent. This idea is corroborated by findings from a neuroimaging study suggesting that consent-induced positive affect is a predictor of conformity to one's peers (Stallen et al., 2013; but see Deuker et al., 2013). Overly simplified, this might be called a positivity bias in the assessment of social norms. We have recently suggested that safe or predictable environments entail a focus on potential reward, while uncertainty provokes an enhanced focus on errors (*Publication IV*). Making somewhat inconsequential judgments and learning whether or not this is consistent with other people's behavior is not a particularly uncertain or threatening context, which is why we might tend to focus on the possibility of gains rather than the threat of loss (see Anderson, 2013; but see Anderson, 2016). Previous interpretations of the involvement of reinforcement-learning mechanisms in social influence as a sign of the individual's fear of being different and urge to avoid social deviance (Kim et al., 2012) might have been incomplete. Admittedly, our findings in *Publication VI* indicate that deviance-induced neural responses, as indicated by the raw FN, do reflect a functionally significant process of norm evaluation that even feeds into subsequent processing. Therefore, it is not exclusively the determination of social *proof* that dominates our analysis of social norms in relation to our own judgments. However, our findings regarding both early and late attentional selection mechanisms concordantly indicate that we appear to seek social proof at least as much as we try to avoid social deviance. Interestingly, there might be another simple, procedural explanation for this positivity bias in how we attend to norms, as will be discussed further below in a preliminary cognitive model of social proof and deviance.

The fact that reinforcement is an integral part of social influence might entice us to assume that processes of reinforcement learning represent the neural substrates of *normative* influence. Conformity driven by normative concerns is the adjustment of one's views to a norm in order to be affiliated and avoid exclusion; in other words, it is conformity triggered by one's drive to gain social reward and to elude social punishment (Deutsch & Gerard, 1955; Kelman, 1961). Our own assumptions about other people's expectations fuel our longing for agreement and fear of being different, which then shapes our behavior to comply with the norm, much like the anticipation of reward and punishment shapes the likelihood of a certain behavior in general (Thorndike, 1911). However, it should be noted that *informational* influence could and should rely on our tendency to seek reward and avoid punishment just as well. Driven by our will to act accurately, we turn to other people's behavior as a valid source of information. Perceiving a norm that signals that we are in line with others should therefore be perceived as predictive of reward (in the sense of having an appropriate representation of reality), while deviating from the norm should elicit the anticipation of being punished (in the sense of being incorrect in one's views). As explained above, most, perhaps even all studies demonstrating the involvement of reinforcement learning in social influence were based on scenarios that delivered descriptive norms triggering informational influence (see, e.g., Klucharev et al., 2009). Spoken more generally, the rewarding or punishing character of social norms does in no way favor the normative account of social influence over the informational one; both motivational foundations should involve reinforcement-learning processes. The notion that common neurocognitive principles subserve normative and informational influence corroborates the assumption that the two most likely motivational underpinnings of conformal adjustment are not easily separable (Cialdini & Goldstein, 2004).

5.1.3 An Integrative Cognitive Model of Social Proof and Deviance

The process of determining social proof and deviance is less straightforward than might be assumed at first glance. The series of studies published over the course of the last few years, including my own studies, can be integrated into a preliminary model of the computations that contribute to and are sparked by the representation of social proof and deviance. As will be explicated in the following, this represents a consolidation of several claims and ideas that were only loosely connected and have not been integrated before.

As a start, a few general assumptions about how we assess norms need to be made. I propose that the process of norm assessment can be disassembled into several steps: access to content; comparison of individual and norm behavior; translation into an outcome. *First*, when confronted with a descriptive social norm, we need access to its actual content, that is, how exactly the majority behaves. *Second*, based upon that, a comparison of this content with the content of our own behavior is needed. That is, we assess whether the two are the same (which constitutes social proof) or different (which conveys our own social deviance); in case of social deviance, we additionally need to identify in which direction our own behavior deviates from the norm. Of course, this comparison is only necessary if it is not already delivered to us. In several studies (e.g., Klucharev et al., 2009; Klucharev et al., 2011; Shestakova et al., 2013), the numerical distance between individual and group judgment was shown in addition to the norm itself. Therefore, participants did not have to perform a comparison anymore. In other studies, however, only the norm was presented, without a depiction of the degree of agreement with the individual's own judgment (e.g., Chen, Wu, et al., 2012; Kim et al., 2012). Consequently, the comparison still had to be performed. *Third*, the output of this comparison—irrespective of whether it is presented to us or we have to work it out ourselves—needs to be translated into the format of a simple outcome, that is, whether our behavior promises reward (in case of social proof) or punishment (in case of social deviance).

Having established the putative components of norm assessment, I propose a broader functional model of cognitive processes that feed into or are triggered by the representation of social proof and deviance. This preliminary model is shown in Figure 1. The crucial external, overt events involved in the overall process are the appearance of a to-be-judged object, a social norm, and the individual's judgment (see black boxes; to be read from top to bottom along the light blue bar). Note that this model assumes that it is not necessary for the individual's initial appraisal of the to-be-judged object, occurring immediately after the object, to be explicitly stated. The general process of evaluating norms to determine social proof or deviance is not fully contingent upon a direct juxtaposition of one's own view and the norm (see *Publication VIII*). Therefore, we merely assume that the object triggers an appraisal that can, but need not, be uttered. The assessment of norms presumably encompasses several processing threads, as explained above. For the sake of simplicity, these aspects of norm processing are not shown separately in the depiction of the proposed model. We do need to assume, however, that we have almost immediate access to a norm's content and, derived from the comparison of one's own judgment and the norm, to its outcome value. For example, when we learn that a group has rated a face as maximally beautiful, we should not only represent this specific norm ("face is perceived as beautiful"), but also what this means for our own judgment ("my rating was correct" versus "my rating was incorrect, as it was too low").

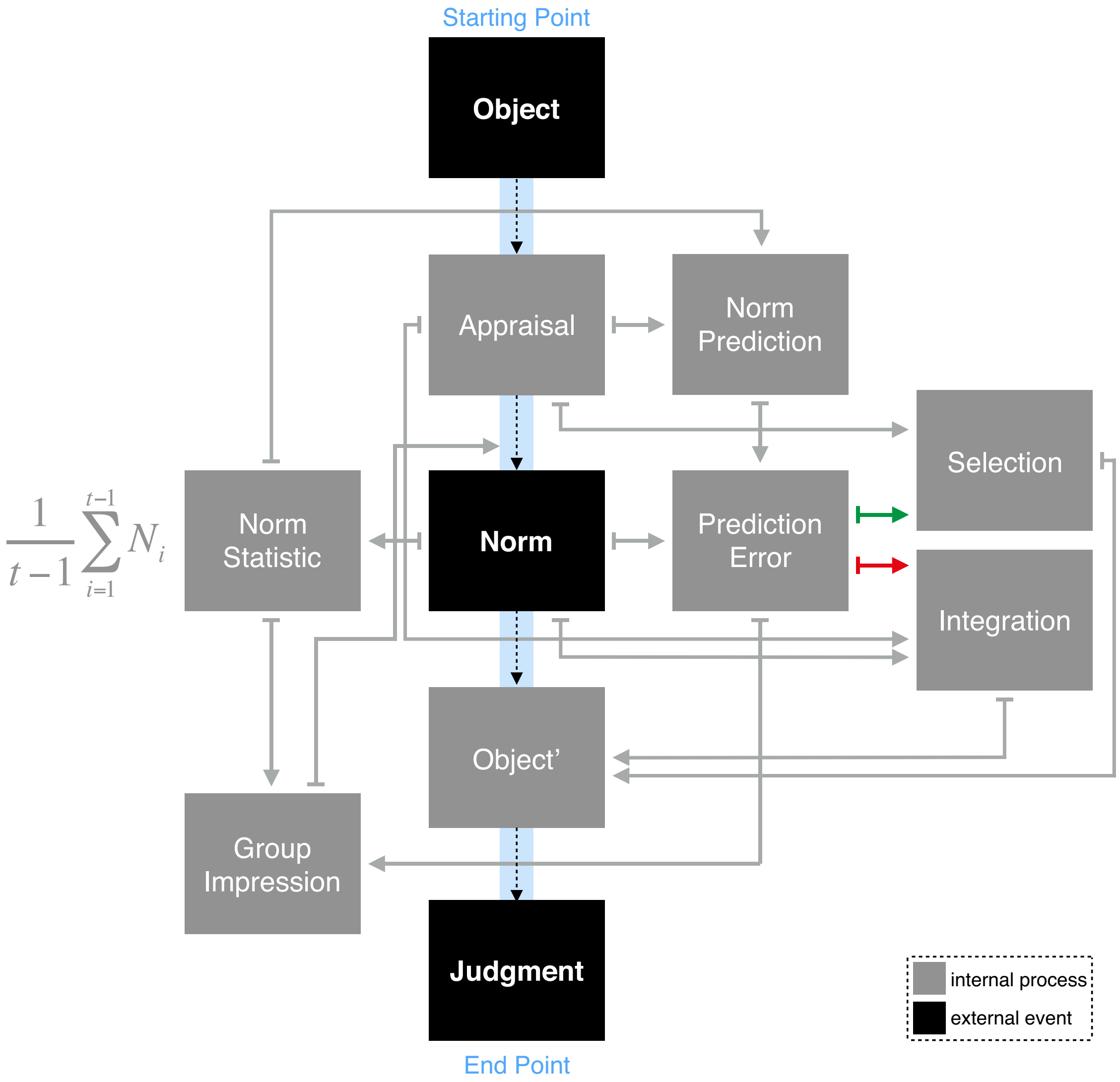

Figure 1: Preliminary model of cognitive processes that contribute to or are sparked by the representation of social proof and deviance. This model is a tentative synthesis of several previous findings as well as some of my own studies presented in this thesis. The figure should be read from top to bottom along the light blue bar. Black boxes represent overt, external events; gray boxes indicate internal, cognitive processes. Note that the putative cognitive components of norm assessment (access to content; comparison of individual and norm behavior; translation into an outcome) are not included in this model for the sake of simplicity. See text for details on the model.

As with other types of outcome (Schultz & Dickinson, 2000), a prediction regarding which norm outcome and, based thereupon, which specific norm content is to be expected should develop, fed by a stored norm statistic, similar to a counter that aggregates the outcome value of all previously processed norms. The norm statistic at time point t could simply be the mean of the outcome value of all norms (N) that have occurred, in a given situation or in regard to a certain issue, up to this point in time (e.g., $+1$ for proof; -1 for deviance). For example, if I received social proof for nine out of ten previous judgments, my norm statistic would be $((9\times1)+(1\times-1))/10 = 0.8$. Thus, I would clearly expect the next norm to provide support for my own judgment. Note that we have proposed that there is some interindividual variation as to how strongly a person responds to social deviance, which should feed into how strongly one attends to this group in the future (*Publication VI*). This variability could be realized in the model by assigning different outcome values to proof and deviance, perhaps even asymmetrically (e.g., $+1$ for proof; -2 for deviance).

When the actual norm is presented, two things are bound to happen: First, its outcome value adds to the norm statistic. Second, the comparison of the actual norm and the norm prediction leads to the determination of the prediction error (actual norm minus predicted norm). The case that the prediction was completely right, that is, identical to the actual norm is not considered in this model due to the fact that this should be far less relevant to the learning process than erroneous predictions (Schultz, 2006). Such errors of the system's predictions can be divided into positive and negative errors. Positive prediction errors occur if the outcome value is higher than predicted. In the above example, my averaged norm statistic, and thus my prediction, is 0.8. Perceiving social proof ($+1$) is an outcome that is still slightly higher than my prediction, which makes this a positive prediction error. Negative prediction errors occur if the outcome is below the prediction. For example, perceiving social deviance (-1) would be way lower than my prediction, making this a strong negative prediction error.

Based on previous studies, it stands to reason that if there is a positive prediction error (see green arrow), this instigates a thorough attentional selection and encoding of the object as well as its obviously correct evaluation. If there is a negative prediction error (see red arrow), this should trigger an increased effort to integrate one's initial assessment, the norm content, and the object. In either case, the individual arrives at an updated representation of the to-be-judged object (*object'*). Moreover, the perception of prediction errors, combined with one's previously aggregated norm statistic, should feed into one's constantly updated impression of the group. As predicted by social-comparison theory (Festinger, 1954), too much deviation—that is, too many negative prediction errors in the present model—should lead to a derogation and subsequent evasion of the group, which would break up the cycle and end the overall process. However, this statistic might not only reflect the actual frequency of proof and deviance, but also an individual's responsiveness to dissent (*Publication VI*). Note that this implies that the organism depicted in this model tends to seek a certain proportion of positive prediction errors: Only if it experiences enough moments of social proof, will it hold a sufficiently favorable view of the group and, hence, not abort the cycle. Crucially, my proposition of a consensus-seeking, self-aborting system could provide an explanation of the well-established notion that we generally tend to expect consensus with our own views (Ross, Greene, & House, 1977). Based on the bias to selectively expose ourselves—in the long run, yet potentially whenever possible—to groups favoring norms that confirm our own views, we might develop a generalized expectation of agreement, even if it is objectively inappropriate (see Sherman, Presson, & Chassin, 1984).

One assumption of this model is that we keep track of the outcome derived from norms, which feeds into our impression of the group and determines our prediction of upcoming norms. Interestingly, in several previous studies, participants were exposed more frequently to conflicting than to agreeing norms (Kim et al., 2012; Klucharev et al., 2009; Klucharev et al., 2011; Shestakova et al., 2013). Therefore, the attentional bias toward social proof might be a consequence of the rareness and, thus, the particular relevance of this type of outcome. Imagine that the system is generally inclined to favor—at least in the long run—sources of feedback that do not contradict our own views too frequently. Under these circumstances in particular, social proof coming from a group whose norms conflict more often with our own views than they agree with them should be a considerably striking and notable event. Considering the idea of prediction errors based on an internal statistic of previous norm outcomes, with deviance representing a negative value (-1) and proof a positive value ($+1$), one should develop a negative overall outcome prediction (e.g., -0.5), such that the positive prediction error triggered by social proof ($+1.5$) should simply be stronger than the negative prediction error triggered by social deviance (-0.5). Our attentional bias toward positive outcomes could therefore be—at least in part—explained by the asymmetry of outcome frequencies.

The proposed model represents a reorganization of our knowledge on the neurocognition of social norms. First, it successfully consolidates several previous findings, particularly those regarding reinforcement-related processes. Second, it incorporates elements from classical psychological and neurophysiological theories, such as on reward processing, on social comparison, and on consensus expectation. Third, it unifies ideas from different methods (behavioral, neuroimaging, ERP) and rather different paradigms. Clearly, however, the model is based on a limited number of observations, as previous research on the neurocognition of social influence is far from exhaustive, and it is speculative in some parts. For example, whether or how we keep track of the overall record of the outcome value of a group's norms has yet to be clarified. Also, the notion of an intensified integration process in the case of deviance is based primarily on one set of data from an idiosyncratic paradigm (*Publication VIII*). In sum, the model is clearly tentative and requires careful examination, testing, and presumably revision in future research. However, even preliminary models have the benefit of synthesizing and explicating conceptual ideas, thus allowing and provoking specific predictions that we can attempt to falsify in dedicated studies to come.

5.2 Extending Our Understanding of the Social Brain

Over the course of the last decade, we have started to unravel the principles of social functions in the human brain (Lieberman, 2007). To some, this has meant exploring elementary functions with an inherently social meaning, often in the service of garnering additional knowledge of the brain itself (Lieberman, 2012; Singer, 2012). Others have understood this as the idea of dissecting complex behavior into its neurocognitive components, thus advancing psychological theory (Amodio et al., 2013; Bartholow & Amodio, 2009). Ideally, both approaches are intertwined and inform each other (Amodio, 2010a). The focus of the present thesis was mainly on understanding the cognitive processes involved and, thus, on extending the psychological approach to social influence. However, our data—mostly of electrophysiological nature—are also suited to contribute to a broader understanding of the social brain.

First, somewhat specific contributions can be made. The medial frontal cortex, including the anterior cingulate cortex, as well as subcortical structures related to affective and value-related processes, such as the basal ganglia, are assumed to play crucial roles in social cognition (Amodio & Frith, 2006; Izuma, 2013). Our ERP experiments confirm this idea, as we have repeatedly shown that signatures of outcome processing, registered at frontocentral recording sites, are pivotal in complex social processes such as assessing proof and deviance (*Publications VI, VII*). Although the nature of the FND is subject to debate and we did not perform any source analyses, two potential generators are typically discussed: the anterior cingulate cortex (Miltner et al., 1997) and the basal ganglia (Foti et al., 2011). Therefore, our data indirectly support the notion that either or both of these regions are crucially involved in human social cognition. Moreover, we propose that the idea of value-driven attention (Anderson et al., 2011; Della Libera & Chelazzi, 2006), potentially involving a direct, top-down modulation of neural responding in sensory areas as well as higher-order biases of selective attention in frontal and parietal brain areas (Serences, 2008), should be considered more explicitly in social cognition (see also Anderson, 2015, 2016). As social information effectively alters our computation of the value that we assign to an object in the basal ganglia and orbitofrontal cortex (Campbell-Meiklejohn et al., 2010; Zaki et al., 2011), and given our findings of norm-dependent early and late prioritization (*Publications VI, VII, VIII*), we reinforce the notion that a key faculty of the human social brain could be a rigorous weighting of information in primary, specialized brain areas and superordinate hubs that govern the deployment of central attentional resources (see Nummenmaa & Calder, 2009).

Second, I believe that one general proposition regarding the principles of the social brain can be made on the basis of what the recent work of many other scholars and myself teaches us about the neurocognition of social influence. While this is not an entirely novel or revolutionary idea, it does underscore contemporary views of the human mind and brain. That reinforcement learning plays a key role in human behavior and, particularly, in decision-making is well established at this point (Lee et al., 2012). What we begin to fathom only now is that this basic system is at the heart of more complex aspects of individual decision-making as well. Our perception of the world is guided by the value that we assign to objects and events, which in turn determines our behavior. In recent years, psychology and closely related disciplines, such as behavioral and neuroeconomics, have strongly bolstered this claim in regard to many types of behavior. What I would like to emphasize, however, is that social behavior does not work much differently. Clearly, we have partly evolved into specialized social operators, featuring a collection of dedicated skills (such as emotions, mentalizing, or empathy), represented by specified neural networks, which mediate how we behave and think socially (Amodio & Frith, 2006; Singer, 2012). But this is not all there is to the human social mind. To put it bluntly, our social environment often simply feeds into generalized systems, utterly non-social in nature. One might even say that much of the social world does not exist as a secluded entity within our mind and brain, represented by dedicated, highly evolved subsystems; rather, it is often merely another instance of basic mental operations that do not seem to differ between complex social experiences (such as learning about other people's opinions) and mundane, more primary types of experience (such as food, sexuality, or money). In fields of research other than social influence, this notion has already gained momentum over the last decade. Perhaps the most notable example is abstract and subjective social pain, caused, for example, by being excluded from a group, which is represented in the human brain very much like physical pain (Eisenberger, 2012; Eisenberger, Lieberman, & Williams, 2003). This could be interpreted as a sign of how vigorous and serious social sensations are to our cognitive system. Alternatively, though, this might just teach us that what we perceive as two utterly different types of stimulation (physical vs. social) are just two exemplars of a common category to our brain. We have elevated one event to be a higher, complex, perhaps innately human experience, while the other one is treated as raw, basic, universal to any type of animal. Our mental computer, however, might represent both events similarly, factually noting a considerable overlap of both types of information, rather than dividing all input into separable, hierarchically organized classes. In the end, social sensitivities, such as affection, longing, deviance, might just be versions of our organism's sensitivities in general, such as satiation, hunger, pain. What we understand as separate qualities of information might be represented almost equally in the human brain.

The fact that one of the major principles of social norms appears to be the modulation of the representation of to-be-judged features corroborates this very idea. As recently demonstrated by Koban and Wager (2016), norms elicit expectations, which in turn prepare our system in a way that eventually leads to changes in how we physically respond to a subsequent stimulus (*Publication I*). Again, complex social experience appears to evoke basic, generalized functions of the cognitive apparatus. It is well established that experiencing regularities in our perceptions quickly leads to expectations (e.g., Doherty, Rao, Mesulam, & Nobre, 2005; Nobre, Correa, & Coull, 2007; Schnuerch, Kreitz, & Lange, 2013), which in turn substantially alter how upcoming sensory input is processed (Esterman & Yantis, 2010; Rohenkohl, Cravo, Wyart, & Nobre, 2012; Summerfield & de Lange, 2014). Typically, we tend to make a thorough distinction between complex, evolved expectations—such as those triggered by social norms—and lowly, plain expectations triggered by perceptual regularities. In the end, though, our system makes use of the same principles, initiates the same type of computation. No advanced, inherently social operator is involved; much of the social world seems to be represented in our mind and brain very much like more mundane stimulations.

In essence, I thus propose that—from a broader perspective—the present thesis reiterates the claim that some of the human faculties that seem to define our most momentous features have evolved around (and are still based upon) quite elementary, universal capacities. This is not to say that our mind is not a wondrous, impressive acquirement; after all, there is no denying how little we know about its workings so far. Yet, perhaps we should not yearn to discover a battery of specialized, higher-order modules (Cosmides & Tooby, 2013); a more promising approach to our enterprise of deciphering the mind might be to understand the small, rudimentary functions that underlie both simple and complex instances of human behavior and experience. This idea of a complex machine whose most advanced capabilities rely on a set of simple, yet essential operations seems to have gained ground in computer science and engineering: Instead of constructing softwares or devices with a myriad of complex functions contrived by the programmer, developers have lately begun to implement only the basic function to learn from feedback (i.e., reinforcement learning) and let the system acquire more complex faculties based upon this central function (Moore & Atkeson, 1993). Clearly, not all that holds true for computer systems applies to the human mind. Perhaps, though, the simple elegance and quick rise of machine learning teaches us a lot about the plain principles of advanced human cognition.

5.3 Limitations

Beyond doubt, no scientific work is ever without limitations. While the convergence of a multitude of investigations in this field increases our confidence in the findings, there are some notable shortcomings in my own studies that should be addressed. First of all, we have used virtual settings of social influence in all of our experiments. While this does capture the recently emerging aspect of mostly anonymous social influence in our everyday virtual interactions (Muchnik, Aral, & Taylor, 2013), it is only an approximation to what social influence actually looks like in the real world. Presumably, the actual presence of others affects these processes to a large degree; for example, it might trigger normative concerns more strongly. Also, the use of ostensibly aggregated group judgments to display norms might artificially increase their informational value. When actually confronted with the opinions of a few other people, we might be much more resistant and resort to their judgments less fiercely and less frequently. Moreover, the presence of dissenters is known to dramatically reduce the impact of the majority on the individual (Asch, 1955); we did not address this relevant limitation of social influence in any of our studies.

A second, yet clearly related concern is that social influence is a complex process that is hard to capture in the laboratory. We have had to simplify, break down, and idealize the scenario. Not only might some of the real-life ingredients of social influence get lost in the process, but this might also overly highlight some aspects. For example, norms might be perceived, or even aggregated across individual observations, in a much more subtle, incidental fashion. Depicting them so bluntly most likely distorts the balance of information in real life. Moreover, the use of ERPs in particular restricted our possibilities regarding which brain processes could be observed. While there are several well-validated signatures of essential cognitive processes, such as attention, several mental operations are not easily mapped to the components of an ERP. Thus, as is usually the case in any type of research, our possibility of insight was severely limited by our choice of method. As explained earlier, fMRI clearly has its own, severe limitations. However, being able to monitor and explore more broadly how the brain responds to certain elements of social-influence scenarios might have allowed us to discover different aspects of its neurocognition.

Third, it is generally hard to assess whether and to what extent our manipulations actually worked in the way we intended them to. Most importantly, we used cover stories in all of our studies and told participants that they would be seeing other people's judgments, which we had collected in previous studies. Such a dishonest approach to psychological research always bears the risk of not actually achieving the type of reality for all participants that we hoped to create. Participants might have asked themselves why we would present the judgments of others. Clearly, most of them saw right through this and assumed that we were interested in processes of social influence. This is not problematic per se, as we are often very much aware of other people's potential influence in real-life situations as well, yet we are still thoroughly influenced by others. Nevertheless, in simplified laboratory scenarios, such as in my series of studies, participants might have been cued to other people's influence *particularly strongly*, much more so than in real life. This knowledge might have distorted participants' behavior. We cannot account for how much a person's explicit knowledge of (and attention to) being exposed to social influence affects their behavior. A related issue is whether participants actually *believed* our cover stories regarding the origin of the ostensible group norms. Of course, we employed paper-pencil and face-to-face debriefings in which we tried to thoroughly assess whether participants doubted the veracity of the norms, which would be much more detrimental to the process than the mere knowledge what the investigation is about. If we had reason to believe that participants did not believe our cover story, we excluded them from all analyses. Moreover, there was indirect evidence that the vast majority of our participants truly believed in the cover stories, as they estimated in a post-experimental survey that in-group norms agreed much more with their own assessments than norms of an out-group—even though the norms were actually manipulated to be the same for both ostensible reference groups (*Publication VII*).

The fourth limitation is that we yet have to connect various theoretical perspectives regarding the principles of social influence. For example, I have mostly ignored or, at least, not specifically focused on the matter of motives. However, to gain a comprehensive understanding of this type of behavior, it is necessary to consider various levels of psychological assessment, rather than restricting one's view to one aspect, such as the neural or cognitive processes. Therefore, my work on this subject—very much like most of the field as a whole—is still a somewhat loose and isolated accumulation of knowledge, concepts, and ideas that is not well-connected to the substantial body of research based on decades of behavioral studies. Clearly, we have based many of our ideas directly on sociopsychological theories (such as in *Publications II, VI, VII*), yet the wealth of findings and interpretations of our own and other scholars' work still needs to be fully and critically integrated into the long tradition of research on social influence.

5.4 Future Avenues

Future studies could build upon what has been uncovered so far and might directly extend some theoretical assertions made in the recent past. For example, the idea that the content of norms can directly shape an individual's mental representation of the to-be-judged feature could be tested directly by varying the tasks and investigating whether it is indeed always norm-related, to-be-judged information that is molded by preceding norms (see *Publication I*). Our theory posits that norms on a certain feature lead to specific expectations regarding this very feature, which then leads to a modification of our internal representation of this specific feature. Perceptual norms thus change how we process to-be-judged items perceptually (Berns et al., 2005), while a norm regarding the value of something can shape our computation of value (Zaki et al., 2011), and a norm regarding the painfulness of an upcoming stimulus alters how we process its painfulness upon its actual occurrence (Koban & Wager, 2016), and so forth. This is, however, a post-hoc explanation that should be put to a direct and critical test. Most importantly, we should attempt to replicate the finding that the *very* neural representation that *directly* governs the individual's judgment is *systematically* shifted by a preceding norm (Zaki et al., 2011).

Also, the somewhat general concept of reinforcement learning in social influence could be extended and specified in several respects. First of all, an appropriate control condition should be used to find out whether neural responses to social proof and deviance are symmetrical or driven mostly by one type of outcome. For example, the ventral striatum is assumed to code both positive and negative reward prediction errors (Schultz & Dickinson, 2000). Thus, by implementing a neutral feedback condition (e.g., without a norm or with a norm unrelated to the current judgments), one could test whether activity in the ventral striatum actually increases for proof and decreases for deviance—or only responds to one type of feedback. Also, the abovementioned issue of reward expectations and prediction-error coding should be clarified. For example, one might manipulate the ratio of proof to deviance and monitor neural responses to both categories in order to find out whether our value-coding system indeed keeps track of social feedback and attunes to the level of predictable reward. Perhaps this is not the case (which would constitute a major deviation from how outcomes are usually coded), and we have a superordinate expectation of social agreement (Ross et al., 1977), such that deviance is generally registered most distinctly and represents the actual driver of the observed differences.

Moreover, several parameters that have been identified over the years as mediating and moderating social influence should be considered more explicitly in future investigations. One could thus gain a more comprehensive overview of the mechanisms of social influence, revealing which cognitive principles apply to whom, when, under what circumstances, to what extent, and so forth. For example, the influence of motives for conformity (Deutsch & Gerard, 1955), the role of a person's identity and self-categorization (Abrams, Wetherell, Cochrane, Hogg, & Turner, 1990), the effect of personality (Crutchfield, 1955; DeYoung, Peterson, & Higgins, 2002), the impact of cognitive mode (Chaiken, 1980; Petty & Cacioppo, 1996), and the role of certainty, investment, and difficulty (Baron et al., 1996) could be controlled for or, even better, manipulated. This would deepen our understanding of social influence and address the abovementioned issue of separate levels of analysis that should be joined, rather than merely coexist.

Also, the mechanisms that have been identified in neuroscientific research to contribute to social influence might be transferred to different domains of psychological inquiry. Specifically, the idea that our basic representations are molded by norms and that reinforcement-learning processes play a key role suggests that individual differences related to these functions should be associated with the susceptibility to social influence. For example, participants who habitually tend to process information intensely might be less prone to the influence of norms (see *Publication II*), possibly even shielding veridical sensory input more vigorously from a fusion with the afflux of norm-based expectations. Therefore, participants scoring high on need for cognition (Cacioppo & Petty, 1982) should tend to conform less to others than those scoring low on need for cognition. Moreover, one of the most influential theories of personality includes the central tenet that individuals differ in the sensitivity to reinforcement, presumably based on functional and structural variations in the underlying neural circuits (for an overview, see, e.g., Corr, 2002, 2004, 2013; Leue & Beauducel, 2008; Reuter, Cooper, Smillie, Markett, & Montag, 2015). Seeing as this appears to be an essential mechanism of social influence, people scoring high on sensitivity to punishment and reward (Torrubia, Avila, Moltó, & Caseras, 2001) should be more likely to conform to others than those scoring low on these variables. In fact, in a preliminary study that has not yet been published, I have tested and confirmed both predictions at the level of trait scales: Using participants' scores on a scale that taps into trait conformity (Mehrabian & Stefl, 1995), I have found need for cognition to be a robust negative predictor of conformity, while sensitivity to punishment and sensitivity to reward were positive predictors. All three variables remained predictive when controlling for sociodemographic variables and basic personality features. However, future research clearly needs to extend these data by testing whether actual conformal behavior—rather than self-reported tendencies to adopt other people's views—is related to personality characteristics such as need for cognition and reinforcement sensitivity. Recent studies from the field of genetic neuroimaging indirectly suggest that a person's neurobiological predisposition to process reinforcement explains part of our susceptibility to social influence (for an overview, see Falk et al., 2012).

Finally, future research should try to apply the insight that we have gained so far to real-world issues related to social influence. As a first step, one might continue the line of basic research, but monitor more relevant aspects of behavior that are known to be prone to social influences, such as eating behavior (Cruwys et al., 2015), reckless driving (Falk et al., 2014), or alcohol intake (Graham, Marks, & Hansen, 1991). Explicitly determining the neurocognitive predictors of such behavior could allow us to adjust measures of prevention or remedy accordingly (Albert et al., 2013; Cascio, Scholz, & Falk, 2015; Stallen & Sanfey, 2015). Only recently, it has been shown that using methods of cognitive neuroscience to validate persuasive communications could be used to improve health-related campaigns (Falk et al., 2016). Likewise, communications based on social norms, such as in anti-drug campaigns, might be tailored more precisely to reach the respective audiences more successfully (Stallen et al., 2013).

5.5 Concluding Remarks

In the present synopsis, I have introduced the long tradition of research into social influence, the novel neurocognitive approach in this field, and my own contributions to the subject. Evidently, much has been achieved and understood thus far, but there is still plenty to explore in the future. The fact that other people's behavior has such a profound and thorough impact on our own overt behavior and even our most basic cognitive operations is as striking as it is unmasking. We are exposed to a rich and complex social environment, yet tend to perceive ourselves as separate entities, navigating through the world in a self-determined, independent way. While this is not entirely inaccurate, it is a gross overstatement. Our construal of reality seems to be the fusion of more or less veridical sensory input and the wealth of social influences surrounding us. Yet this is not an admission of weakness or sloth, but perhaps the simple truth that we have much to gain from using others not only as a means of cohesion and belonging, but also as a source of information that helps us tackle the intricacies of the world.

6 Bibliography

Abe, N. (2011). How the brain shapes deception: An integrated review of the literature. *The Neuroscientist*, *17*, 560–574. doi:10.1177/1073858410393359

Abrams, D., Wetherell, M., Cochrane, S., Hogg, M. A., & Turner, J. C. (1990). Knowing what to think by knowing who you are: Self-categorization and the nature of norm formation, conformity and group polarization. *British Journal of Social Psychology*, *29*, 97–119. doi:10.1111/j.2044-8309.1990.tb00892.x

Albert, D., Chein, J., & Steinberg, L. (2013). The teenage brain: Peer influences on adolescent decision making. *Current Directions in Psychological Science*, *22*, 114–120. doi:10.1177/0963721412471347

Alexander, W. H., & Brown, J. W. (2011). Medial prefrontal cortex as an action-outcome predictor. *Nature Neuroscience*, *14*, 1338–1344. doi:10.1038/nn.2921

Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, *95*, 1082–1095. doi:10.1016/j.jpubeco.2011.03.003

Allport, G. (1959). The historical background of modern social psychology. In G. Lindzey, & E. Aronson (Eds.), *Handbook of social psychology* (2nd ed., Vol. 1, pp. 1–80). Reading, MA: Addison-Wesley.

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, *111*, 256–274. doi:10.1037/0033-2909.111.2.256

Amodio, D. M. (2010a). Can neuroscience advance social psychological theory? Social neuroscience for the behavioral social psychologist. *Social Cognition*, *28*, 695–716. doi:10.1521/soco.2010.28.6.695

Amodio, D. M. (2010b). Coordinated roles of motivation and perception in the regulation of intergroup responses: Frontal cortical asymmetry effects on the P2 event-related potential and behavior. *Journal of Cognitive Neuroscience*, *22*, 2609–2617. doi:10.1162/jocn.2009.21395

Amodio, D. M., Bartholow, B. D., & Ito, T. A. (2013). Tracking the dynamics of the social brain: ERP approaches for social cognitive and affective neuroscience. *Social Cognitive and Affective Neuroscience*, *9*, 385–393. doi:10.1093/scan/nst177

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, *7*, 268–277. doi:10.1038/nrn1884

Anderson, B. A. (2013). A value-driven mechanism of attentional selection. *Journal of Vision*, *13*, 1–16. doi:10.1167/13.3.7

Anderson, B. A. (2015). Social reward shapes attentional biases. *Cognitive Neuroscience*, *7*, 30–36. doi:10.1080/17588928.2015.1047823

Anderson, B. A. (2016). Counterintuitive effects of negative social feedback on attention. *Cognition and Emotion*. doi:10.1080/02699931.2015.1122576

Anderson, B. A., Laurent, P. A., & Yantis, S. (2011). Value-driven attentional capture. *Proceedings of the National Academy of Sciences*, *108*, 10367–10371. doi:10.1073/pnas.1104047108

Anllo-Vento, L., Luck, S. J., & Hillyard, S. A. (1998). Spatio-temporal dynamics of attention to color: Evidence from human electrophysiology. *Human Brain Mapping*, *6*, 216–238.

Arzouan, Y., Goldstein, A., & Faust, M. (2007). Brainwaves are stethoscopes: ERP correlates of novel metaphor comprehension. *Brain Research*, *1160*, 69–81. doi:10.1016/j.brainres.2007.05.034

Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership, and men* (pp. 177–190). Oxford, England: Carnegie Press.

Asch, S. E. (1955). Opinions and social pressure. *Scientific American*, *193*, 31–35. doi:10.1038/scientificamerican1155-31

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied*, *70*, 1–70.

Ashby, F. G. (2015). An introduction to fMRI. In B. U. Forstmann, & E.-J. Wagenmakers (Eds.), *An introduction to model-based cognitive neuroscience* (pp. 91–112). New York, NY: Springer.

Axsom, D., Yates, S., & Chaiken, S. (1987). Audience response as a heuristic cue in persuasion. *Journal of Personality and Social Psychology*, *53*, 30–40. doi:10.1037//0022-3514.53.1.30

Baker, T. E., & Holroyd, C. B. (2011). Dissociated roles of the anterior cingulate cortex in reward and conflict processing as revealed by the feedback error-related negativity and N200. *Biological Psychology*, *87*, 25–34. doi:10.1016/j.biopsycho.2011.01.010

Barnett, A. G. (2004). Regression to the mean: What it is and how to deal with it. *International Journal of Epidemiology*, *34*, 215–220. doi:10.1093/ije/dyh299

Baron, R. S., Vandello, J. A., & Brunsman, B. (1996). The forgotten variable in conformity research: Impact of task importance on social influence. *Journal of Personality and Social Psychology*, *71*, 915–927. doi:10.1037/0022-3514.71.5.915

Bartholow, B. D. (2010). Event-related brain potentials and social cognition: On using physiological information to constrain social cognitive theories. *Social Cognition*, *28*, 723–747. doi:10.1521/soco.2010.28.6.723

Bartholow, B. D., & Amodio, D. M. (2009). Using event-related brain potentials in social psychological research: A brief review and tutorial. In E. Harmon-Jones, & J. S. Beer (Eds.), *Methods in social neuroscience* (pp. 198–232). New York, NY, US: Guilford Press.

Bell, R., Sasse, J., Möller, M., Czernochowski, D., Mayr, S., & Buchner, A. (2015). Event-related potentials in response to cheating and cooperation in a social dilemma game. *Psychophysiology*, *53*, 216–228. doi:10.1111/psyp.12561

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*, *8*, 551–565. doi:10.1162/jocn.1996.8.6.551

Berger, H. (1929). Über das Elektrenkephalogramm des Menschen. *Archiv für Psychiatrie und Nervenkrankheiten*, *87*, 527–570. doi:10.1007/BF01797193

Berns, G. S., Capra, C. M., Moore, S., & Noussair, C. (2010). Neural mechanisms of the influence of popularity on adolescent ratings of music. *NeuroImage*, *49*, 2687–2696. doi:10.1016/j.neuroimage.2009.10.070

Berns, G. S., Chappelow, J., Zink, C. F., Pagnoni, G., Martin-Skurski, M. E., & Richards, J. (2005). Neurobiological correlates of social conformity and independence during mental rotation. *Biological Psychiatry*, *58*, 245–253. doi:10.1016/j.biopsych.2005.04.012

Berns, G. S., McClure, S. M., Pagnoni, G., & Montague, P. R. (2001). Predictability modulates human brain response to reward. *Journal of Neuroscience*, *21*, 2793–2798.

Blass, T. (1991). Understanding behavior in the Milgram obedience experiment: The role of personality, situations, and their interactions. *Journal of Personality and Social Psychology*, *60*, 398–413. doi:10.1037/0022-3514.60.3.398

Blass, T. (1999). The Milgram paradigm after 35 years: Some things we now know about obedience to authority. *Journal of Applied Social Psychology*, *29*, 955–978. doi:10.1111/j.1559-1816.1999.tb00134.x

Blass, T. (2009). From New Haven to Santa Clara: A historical perspective on the Milgram obedience experiments. *American Psychologist*, *64*, 37–45. doi:10.1037/a0014434

Bless, H., Clore, G. L., Schwarz, N., Golisano, V., Rabe, C., & Wölk, M. (1996). Mood and the use of scripts: Does a happy mood really lead to mindlessness? *Journal of Personality and Social Psychology*, *71*, 665–679. doi:10.1037/0022-3514.71.4.665

Blomqvist, N. (1987). On the bias caused by regression toward the mean in studying the relation between change and initial value. *Journal of Clinical Periodontology*, *14*, 34–37. doi:10.1111/j.1600-051X.1987.tb01510.x

Bloom, E. L., Potts, G. F., Evans, D. E., & Drobes, D. J. (2013). Cue reactivity in smokers: An event-related potential study. *International Journal of Psychophysiology*, *90*, 258–264. doi:10.1016/j.ijpsycho.2013.08.005

Blumenthal, A. L. (1975). A reappraisal of Wilhelm Wundt. *American Psychologist*, *30*, 1081–1088. doi:10.1037/0003-066X.30.11.1081

Bodenhausen, G. V., Kramer, G. P., & Süsser, K. (1994). Happiness and stereotypic thinking in social judgment. *Journal of Personality and Social Psychology*, *66*, 621–632. doi:10.1037/0022-3514.66.4.621

Bond, C. F., & Titus, L. J. (1983). Social facilitation: A meta-analysis of 241 studies. *Psychological Bulletin*, *94*, 265–292. doi:10.1037/0033-2909.94.2.265

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, *489*, 295–298. doi:10.1038/nature11421

Bond, R., & Smith, P. B. (1996). Culture and conformity: A meta-analysis of studies using Asch's (1952b, 1956) line judgment task. *Psychological Bulletin*, *119*, 111. doi:10.1037/0033-2909.119.1.111

Burger, J. M. (2009). Replicating Milgram: Would people still obey today? *American Psychologist*, *64*, 1–11. doi:10.1037/a0010932

Cacioppo, J. T., Crites, S. L., & Gardner, W. L. (1996). Attitudes to the right: Evaluative processing is associated with lateralized late positive event-related brain potentials. *Personality and Social Psychology Bulletin*, *22*, 1205–1219. doi:10.1177/01461672962212002

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*, 116–131. doi:10.1037/0022-3514.42.1.116

Campbell, J. D., & Fairey, P. J. (1989). Informational and normative routes to conformity: The effect of faction size as a function of norm extremity and attention to the stimulus. *Journal of Personality and Social Psychology*, *57*, 457–468. doi:10.1037/0022-3514.57.3.457

Campbell-Meiklejohn, D. K., Bach, D. R., Roepstorff, A., Dolan, R. J., & Frith, C. D. (2010). How the opinion of others affects our valuation of objects. *Current Biology*, *20*, 1165–1170. doi:10.1016/j.cub.2010.04.055

Campbell-Meiklejohn, D. K., Kanai, R., Bahrami, B., Bach, D. R., Dolan, R. J., Roepstorff, A., & Frith, C. D. (2012). Structure of orbitofrontal cortex predicts social influence. *Current Biology*, *22*, R123–R124. doi:10.1016/j.cub.2012.01.012

Carretié, L., Hinojosa, J. A., Martín-Loeches, M., Mercado, F., & Tapia, M. (2004). Automatic attention to emotional stimuli: Neural correlates. *Human Brain Mapping*, *22*, 290–299. doi:10.1002/hbm.20037

Cascio, C. N., O'Donnell, M. B., Bayer, J., Tinney, F. J., & Falk, E. B. (2015). Neural correlates of susceptibility to group opinions in online word-of-mouth recommendations. *Journal of Marketing Research*, *52*, 559–575. doi:10.1509/jmr.13.0611

Cascio, C. N., Scholz, C., & Falk, E. B. (2015). Social influence and the brain: Persuasion, susceptibility to influence and retransmission. *Current Opinion in Behavioral Sciences*, *3*, 51–57. doi:10.1016/j.cobeha.2015.01.007

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, *39*, 752–766. doi:10.1037/0022-3514.39.5.752

Chen, J., Zhong, J., Zhang, Y., Li, P., Zhang, A., Tan, Q., & Li, H. (2012). Electrophysiological correlates of processing facial attractiveness and its influence on cooperative behavior. *Neuroscience Letters*, *517*, 65–70. doi:10.1016/j.neulet.2012.02.082

Chen, J., Wu, Y., Tong, G., Guan, X., & Zhou, X. (2012). ERP correlates of social conformity in a line judgment task. *BMC Neuroscience*, *13*, 43. doi:10.1186/1471-2202-13-43

Chen, M., Ma, Q., Li, M., Lai, H., Wang, X., & Shu, L. (2010). Cognitive and emotional conflicts of counter-conformity choice in purchasing books online: An event-related potentials study. *Biological Psychology*, *85*, 437–445. doi:10.1016/j.biopsycho.2010.09.006

Chen, S., Duckworth, K., & Chaiken, S. (1999). Motivated heuristic and systematic processing. *Psychological Inquiry*, *10*, 44–49. doi:10.1207/s15327965pli1001_6

Chen, S., Shechter, D., & Chaiken, S. (1996). Getting at the truth or getting along: Accuracy-versus impression-motivated heuristic and systematic processing. *Journal of Personality and Social Psychology*, *71*, 262–275. doi:10.1037//0022-3514.71.2.262

Cialdini, R. B. (2007). Descriptive social norms as underappreciated sources of social control. *Psychometrika*, *72*, 263–268. doi:10.1007/s11336-006-1560-6

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, *55*, 591–621. doi:10.1146/annurev.psych.55.090902.142015

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, *58*, 1015–1026. doi:10.1037/0022-3514.58.6.1015

Cialdini, R. B., & Trost, M. R. (1998). Social influence: Social norms, conformity and compliance. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (pp. 151–192). New York, NY, US: McGraw-Hill.

Cialdini, R. B., Wosinska, W., Barrett, D. W., Butner, J., & Gornik-Durose, M. (1999). Compliance with a request in two cultures: The differential influence of social proof and commitment/consistency on collectivists and individualists. *Personality and Social Psychology Bulletin*, *25*, 1242–1253. doi:10.1177/0146167299258006

Claidière, N., & Whiten, A. (2012). Integrating the study of conformity and culture in humans and nonhuman animals. *Psychological Bulletin*, *138*, 126–145. doi:10.1037/a0025868

Coles, M. G. H. (1989). Modern mind-brain reading: Psychophysiology, physiology, and cognition. *Psychophysiology*, *26*, 251–269. doi:10.1111/j.1469-8986.1989.tb01916.x

Corr, P. J. (2002). J. A. Gray's reinforcement sensitivity theory: Tests of the joint subsystems hypothesis of anxiety and impulsivity. *Personality and Individual Differences*, *33*, 511–532. doi:10.1016/S0191-8869(01)00170-2

Corr, P. J. (2004). Reinforcement sensitivity theory and personality. *Neuroscience & Biobehavioral Reviews*, *28*, 317–332. doi:10.1016/j.neubiorev.2004.01.005

Corr, P. J. (2013). Approach and avoidance behaviour: Multiple systems and their interactions. *Emotion Review*, *5*, 285–290. doi:10.1177/1754073913477507

Cosmides, L., & Tooby, J. (2013). Evolutionary psychology: New perspectives on cognition and motivation. *Annual Review of Psychology*, *64*, 201–229. doi:10.1146/annurev.psych.121208.131628

Croson, R., Handy, F., & Shang, J. (2009). Keeping up with the Joneses: The relationship of perceived descriptive social norms, social information, and charitable giving. *Nonprofit Management and Leadership*, *19*, 467–489. doi:10.1002/nml.232

Crutchfield, R. S. (1955). Conformity and character. *American Psychologist*, *10*, 191–198. doi:10.1037/h0040237

Cruwys, T., Bevelander, K. E., & Hermans, R. C. J. (2015). Social modeling of eating: A review of when and why social influence affects food intake and choice. *Appetite*, *86*, 3–18. doi:10.1016/j.appet.2014.08.035

Cuthbert, B. N., Schupp, H. T., Bradley, M. M., Birbaumer, N., & Lang, P. J. (2000). Brain potentials in affective picture processing: covariation with autonomic arousal and affective report. *Biological Psychology*, *52*, 95–111. doi:10.1016/S0301-0511(99)00044-7

Cutter, G. R. (1976). Some examples for teaching regression toward the mean from a sampling viewpoint. *American Statistician*, *30*, 194–197. doi:10.2307/2683763

Darvas, F., Pantazis, D., Kucukaltun-Yildirim, E., & Leahy, R. M. (2004). Mapping human brain function with MEG and EEG: Methods and validation. *NeuroImage*, *23*, S289–S299. doi:10.1016/j.neuroimage.2004.07.014

Deffke, I., Sander, T., Heidenreich, J., Sommer, W., Curio, G., Trahms, L., & Lueschow, A. (2007). MEG/EEG sources of the 170-ms response to faces are co-localized in the fusiform gyrus. *NeuroImage*, *35*, 1495–1501. doi:10.1016/j.neuroimage.2007.01.034

Della Libera, C., & Chelazzi, L. (2006). Visual selective attention and the effects of monetary rewards. *Psychological Science*, *17*, 222–227. doi:10.1111/j.1467-9280.2006.01689.x

Deuker, L., Müller, A. R., Montag, C., Markett, S., Reuter, M., Fell, J., … Axmacher, N. (2013). Playing nice: A multi-methodological study on the effects of social conformity on memory. *Frontiers in Human Neuroscience*, *7*, 79. doi:10.3389/fnhum.2013.00079

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *Journal of Abnormal and Social Psychology*, *51*, 629–636. doi:10.1037/h0046408

DeYoung, C. G., Peterson, J. B., & Higgins, D. M. (2002). Higher-order factors of the Big Five predict conformity: Are there neuroses of health? *Personality and Individual Differences*, *33*, 533–552. doi:10.1016/S0191-8869(01)00171-4

Diener, E. (2010). Neuroimaging: Voodoo, new phrenology, or scientific breakthrough? Introduction to special section on fMRI. *Perspectives on Psychological Science*, *5*, 714–715. doi:10.1177/1745691610388773

Doherty, J. R., Rao, A., Mesulam, M. M., & Nobre, A. C. (2005). Synergistic effect of combined temporal and spatial expectations on visual attention. *Journal of Neuroscience*, *25*, 8259–8266. doi:10.1523/JNEUROSCI.1821-05.2005

Edelson, M., Sharot, T., Dolan, R. J., & Dudai, Y. (2011). Following the crowd: Brain substrates of long-term memory conformity. *Science*, *333*, 108–111. doi:10.1126/science.1203557

Eimer, M. (1998). Does the face-specific N170 component reflect the activity of a specialized eye processor? *NeuroReport*, *9*, 2945–2948. doi:10.1097/00001756-199809140-00005

Eimer, M. (2000). The face-specific N170 component reflects late stages in the structural encoding of faces. *NeuroReport*, *11*, 2319–2324. doi:10.1097/00001756-200007140-00050

Eimer, M. (2011). The face-sensitivity of the N170 component. *Frontiers in Human Neuroscience*, *5*, 119. doi:10.3389/fnhum.2011.00119

Eisenberger, N. I. (2012). The neural bases of social pain: Evidence for shared representations with physical pain. *Psychosomatic Medicine*, *74*, 126–135. doi:10.1097/PSY.0b013e3182464dd1

Eisenberger, N. I., Lieberman, M. D., & Williams, K. D. (2003). Does rejection hurt? An fMRI study of social exclusion. *Science*, *302*, 290–292. doi:10.1126/science.1089134

Elbert, T., & Rockstroh, B. (1987). Threshold regulation - a key to the understanding of the combined dynamics of EEG and event-related potentials. *Journal of Psychophysiology*, *1*, 317–333.

Eppinger, B., Kray, J., Mock, B., & Mecklinger, A. (2008). Better or worse than expected? Aging, learning, and the ERN. *Neuropsychologia*, *46*, 521–539. doi:10.1016/j.neuropsychologia.2007.09.001

Esterman, M., & Yantis, S. (2010). Perceptual expectation evokes category-selective cortical activity. *Cerebral Cortex*, *20*, 1245–1253. doi:10.1093/cercor/bhp188

Fabiani, M., Gratton, G., & Coles, M. G. H. (2000). Event-related brain potentials: methods, theory, and application. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (pp. 53–84). Cambridge, UK: Cambridge University Press.

Fahrenberg, J. (2012). Wilhelm Wundts Wissenschaftstheorie der Psychologie. *Psychologische Rundschau*, *63*, 228–238. doi:10.1026/0033-3042/a000141

Falk, E. B., Berkman, E. T., Mann, T., Harrison, B., & Lieberman, M. D. (2010). Predicting persuasion-induced behavior change from the brain. *Journal of Neuroscience*, *30*, 8421–8424. doi:10.1523/JNEUROSCI.0063-10.2010

Falk, E. B., Cascio, C. N., Brook O'Donnell, M., Carp, J., Tinney, F. J., Bingham, C. R., ... Simons-Morton, B. G. (2014). Neural responses to exclusion predict susceptibility to social influence. *Journal of Adolescent Health*, *54*, S22–S31. doi:10.1016/j.jadohealth.2013.12.035

Falk, E. B., O'Donnell, M. B., Cascio, C. N., Tinney, F., Kang, Y., Lieberman, M. D., ... Strecher, V. J. (2015). Self-affirmation alters the brain's response to health messages and subsequent behavior change. *Proceedings of the National Academy of Sciences*, *112*, 1977–1982. doi:10.1073/pnas.1500247112

Falk, E. B., O'Donnell, M. B., Tompson, S., Gonzalez, R., Cin, S. D., Strecher, V., ... An, L. (2016). Functional brain imaging predicts public health campaign success. *Social Cognitive and Affective Neuroscience*, *11*, 204–214. doi:10.1093/scan/nsv108

Falk, E. B., Way, B. M., & Jasinska, A. J. (2012). An imaging genetics approach to understanding social influence. *Frontiers in Human Neuroscience*, *6*, 168. doi:10.3389/fnhum.2012.00168

Farr, R. M. (1983). Wilhelm Wundt (1832–1920) and the origins of psychology as an experimental and social science. *British Journal of Social Psychology*, *22*, 289–301. doi:10.1111/j.2044-8309.1983.tb00596.x

Fehr, E., & Fischbacher, U. (2004). Social norms and human cooperation. *Trends in Cognitive Sciences*, *8*, 185–190. doi:10.1016/j.tics.2004.02.007

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*, 117–140. doi:10.1177/001872675400700202

Festinger, L. (1962). Cognitive dissonance. *Scientific American*, *207*, 93–107. doi:10.1038/scientificamerican1062-93

Fishbein, M., & Ajzen, I. (1975). *Belief, attitudes, intention, and behavior. An introduction to theory and research.* Reading, MA: Addison-Wesley.

Foti, D., Weinberg, A., Dien, J., & Hajcak, G. (2011). Event-related potential activity in the basal ganglia differentiates rewards from nonrewards: Temporospatial principal components analysis and source localization of the feedback negativity. *Human Brain Mapping*, *32*, 2207–2216. doi:10.1002/hbm.21182

Frank, M. J., Seeberger, L. C., & O'Reilly, R. C. (2004). By carrot or by stick: Cognitive reinforcement learning in parkinsonism. *Science*, *306*, 1940–1943. doi:10.1126/science.1102941

Frank, M. J., Woroch, B. S., & Curran, T. (2005). Error-related negativity predicts reinforcement learning and conflict biases. *Neuron*, *47*, 495–501. doi:10.1016/j.neuron.2005.06.020

Friend, R., Rafferty, Y., & Bramel, D. (1990). A puzzling misinterpretation of the Asch 'conformity' study. *European Journal of Social Psychology*, *20*, 29–44. doi:10.1002/ejsp.2420200104

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *360*, 815–836. doi:10.1098/rstb.2005.1622

Furley, P., Schnuerch, R., & Gibbons, H. (2016). The winner takes it all: Event-related brain potentials reveal enhanced motivated attention toward athletes' nonverbal signals of leading. *Social Neuroscience*. doi:10.1080/17470919.2016.1182586

Furley, P., & Schweizer, G. (2014). The expression of victory and loss: Estimating who's leading or trailing from nonverbal cues in sports. *Journal of Nonverbal Behavior*, *38*, 13–29. doi:10.1007/s10919-013-0168-7

Furley, P., & Schweizer, G. (2015). In a flash: Thin slice judgment accuracy of leading and trailing in sports. *Journal of Nonverbal Behavior*, *40*, 83–100. doi:10.1007/s10919-015-0225-5

Gast, A., Gawronski, B., & De Houwer, J. (2012). Evaluative conditioning: Recent developments and future directions. *Learning and Motivation*, *43*, 79–88. doi:10.1016/j.lmot.2012.06.004

Gehring, W. J., & Willoughby, A. R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. *Science*, *295*, 2279–2282. doi:10.1126/science.1066893

Gergen, K. J. (1985). The social constructionist movement in modern psychology. *American Psychologist*, *40*, 266–275. doi:10.1037/0003-066X.40.3.266

Germar, M., Albrecht, T., Voss, A., & Mojzisch, A. (2016). Social conformity is due to biased stimulus processing: Electrophysiological and diffusion analyses. *Social Cognitive and Affective Neuroscience*. doi:10.1093/scan/nsw050

Germar, M., Schlemmer, A., Krug, K., Voss, A., & Mojzisch, A. (2014). Social influence and perceptual decision making: A diffusion model analysis. *Personality and Social Psychology Bulletin*, *40*, 217–231. doi:10.1177/0146167213508985

Giard, M. H., Perrin, F., Echallier, J. F., Thévenet, M., Froment, J. C., & Pernier, J. (1994). Dissociation of temporal and frontal components in the human auditory N1 wave: a scalp current density and dipole model analysis. *Electroencephalography and Clinical Neurophysiology*, *92*, 238–252. doi:10.1016/0168-5597(94)90067-1

Gibbons, H., Bachmann, O., & Stahl, J. (2014). The more you ignore me the closer I get: An ERP study of evaluative priming. *Cognitive, Affective, & Behavioral Neuroscience*, *14*, 1467–1484. doi:10.3758/s13415-014-0289-4

Gibbons, H., Schnuerch, R., & Stahl, J. (2016). From positivity to negativity bias: Ambiguity affects the neurophysiological signatures of feedback processing. *Journal of Cognitive Neuroscience*, *28*, 542–557. doi:10.1162/jocn_a_00921

Gil, S., & Le Bigot, L. (2014). Seeing life through positive-tinted glasses: Color–meaning associations. *PLoS ONE*, *9*, e104291. doi:10.1371/journal.pone.0104291

Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research*, *35*, 472–482. doi:10.1086/586910

Graham, J. W., Marks, G., & Hansen, W. B. (1991). Social influence processes affecting adolescent substance use. *Journal of Applied Psychology*, *76*, 291–298. doi:10.1037/0021-9010.76.2.291

Gray, H. M., Ambady, N., Lowenthal, W. T., & Deldin, P. (2004). P300 as an index of attention to self-relevant stimuli. *Journal of Experimental Social Psychology*, *40*, 216–224. doi:10.1016/S0022-1031(03)00092-1

Greenwood, J. D. (2003). Wundt, Völkerpsychologie, and experimental social psychology. *History of Psychology*, *6*, 70–88. doi:10.1037/1093-4510.6.1.70

Haas, L. F. (2003). Hans Berger (1873–1941), Richard Caton (1842–1926), and electroencephalography. *Journal of Neurology, Neurosurgery & Psychiatry*, *74*, 9. doi:10.1136/jnnp.74.1.9

Hajcak, G., Dunning, J. P., & Foti, D. (2009). Motivated and controlled attention to emotion: Time-course of the late positive potential. *Clinical Neurophysiology*, *120*, 505–510. doi:10.1016/j.clinph.2008.11.028

Haney, C., Banks, W. C., & Zimbardo, P. G. (1973). A study of prisoners and guards in a simulated prison. *Naval Research Reviews*, *9*, 1–17.

Harkins, S. G. (1987). Social loafing and social facilitation. *Journal of Experimental Social Psychology*, *23*, 1–18. doi:10.1016/0022-1031(87)90022-9

Harris, J. J., Reynell, C., & Attwell, D. (2011). The physiology of developmental changes in BOLD functional imaging signals. *Developmental Cognitive Neuroscience*, *1*, 199–216. doi:10.1016/j.dcn.2011.04.001

Haslam, S. A., & Reicher, S. (2012). Contesting the "nature" of conformity: What Milgram and Zimbardo's studies really show. *PLoS Biology*, *10*, e1001426. doi:10.1371/journal.pbio.1001426

Hogg, M. A., & Reid, S. A. (2006). Social identity, self-categorization, and the communication of group norms. *Communication Theory*, *16*, 7–30. doi:10.1111/j.1468-2885.2006.00003.x

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*, 679–709. doi:10.1037/0033-295X.109.4.679

Holroyd, C. B., Hajcak, G., & Larsen, J. T. (2006). The good, the bad and the neutral: Electrophysiological responses to feedback stimuli. *Brain Research*, *1105*, 93–101. doi:10.1016/j.brainres.2005.12.015

Holroyd, C. B., Krigolson, O. E., & Lee, S. (2011). Reward positivity elicited by predictive cues. *NeuroReport*, *22*, 249–252. doi:10.1097/WNR.0b013e328345441d

Huang, Y., Kendrick, K. M., & Yu, R. (2014). Social conflicts elicit an N400-like component. *Neuropsychologia*, *65*, 211–220. doi:10.1016/j.neuropsychologia.2014.10.032i

Hull, C. L. (1933). *Hypnosis and suggestibility.* Oxford, England: Appleton-Century.

Insko, C. A., Smith, R. H., Alicke, M. D., Wade, J., & Taylor, S. (1985). Conformity and group size: The concern with being right and the concern with being liked. *Personality and Social Psychology Bulletin*, *11*, 41–50. doi:10.1177/0146167285111004

Itier, R. J., & Taylor, M. J. (2004). N170 or N1? Spatiotemporal differences between object and face processing using ERPs. *Cerebral Cortex*, *14*, 132–142. doi:10.1093/cercor/bhg111

Ito, T. A. (2004). Tracking the timecourse of social perception: The effects of racial cues on event-related brain potentials. *Personality and Social Psychology Bulletin*, *30*, 1267–1280. doi:10.1177/0146167204264335

Ito, T. A., Larsen, J. T., Smith, N. K., & Cacioppo, J. T. (1998). Negative information weighs more heavily on the brain: The negativity bias in evaluative categorizations. *Journal of Personality and Social Psychology*, *75*, 887–900. doi:10.1037//0022-3514.75.4.887

Ito, T. A., & Urland, G. R. (2005). The influence of processing objectives on the perception of faces: An ERP study of race and gender perception. *Cognitive, Affective, & Behavioral Neuroscience*, *5*, 21–36. doi:10.3758/CABN.5.1.21

Izuma, K. (2013). The neural basis of social influence and attitude change. *Current Opinion in Neurobiology*, *23*, 456–462. doi:10.1016/j.conb.2013.03.009

Izuma, K., & Adolphs, R. (2013). Social manipulation of preference in the human brain. *Neuron*, *78*, 563–573. doi:10.1016/j.neuron.2013.03.023

Jääskeläinen, I. P., Ahveninen, J., Bonmassar, G., Dale, A. M., Ilmoniemi, R. J., Levänen, S., . . . Belliveau, J. W. (2004). Human posterior auditory cortex gates novel sounds to consciousness. *Proceedings of the National Academy of Sciences*, *101*, 6809–6814. doi:10.1073/pnas.0303760101

Jacobson, R. P., Mortensen, C. R., Jacobson, K. J. L., & Cialdini, R. B. (2015). Self-control moderates the effectiveness of influence attempts highlighting injunctive social norms. *Social Psychological and Personality Science*, *6*, 718–726. doi:10.1177/1948550615578463

Jancke, L., & Petermann, F. (2010). Zum Verhältnis von Biologie und Psychologie. *Psychologische Rundschau*, *61*, 175–179. doi:10.1026/0033-3042/a000044

Johnston, V. S., & Oliver-Rodriguez, J. C. (1997). Facial beauty and the late positive component of event-related potentials. *Journal of Sex Research*, *34*, 188–198. doi:10.1080/00224499709551884

Jüttemann, G. (2007). Wundts Psychologiekonzeption ist nicht die Ursache, sondern die Lösung des Problems. *Psychologische Rundschau*, *58*, 267–269.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, *93*, 136–153. doi:10.1037/0033-295X.93.2.136

Kallgren, C. A., Reno, R. R., & Cialdini, R. B. (2000). A focus theory of normative conduct: When norms do and do not affect behavior. *Personality and Social Psychology Bulletin*, *26*, 1002–1012. doi:10.1177/01461672002610009

Kelman, H. C. (1961). Processes of opinion change. *Public Opinion Quarterly*, *25*, 57–78.

Kim, B.-R., Liss, A., Rao, M., Singer, Z., & Compton, R. J. (2012). Social deviance activates the brain's error-monitoring system. *Cognitive, Affective, & Behavioral Neuroscience*, *12*, 65–73. doi:10.3758/s13415-011-0067-5

Kimura, K., & Katayama, J. (2013). Outcome evaluations in group decision making using the majority rule: An electrophysiological study. *Psychophysiology*, *50*, 848–857. doi:10.1111/psyp.12068

Kimura, K., Murayama, A., Miura, A., & Katayama, J. (2013). Effect of decision confidence on the evaluation of conflicting decisions in a social context. *Neuroscience Letters*, *556*, 176–180. doi:10.1016/j.neulet.2013.09.020

Klucharev, V., Hytönen, K., Rijpkema, M., Smidts, A., & Fernández, G. (2009). Reinforcement learning signal predicts social conformity. *Neuron*, *61*, 140–151. doi:10.1016/j.neuron.2008.11.027

Klucharev, V., Munneke, M. A. M., Smidts, A., & Fernández, G. (2011). Downregulation of the posterior medial frontal cortex prevents social conformity. *Journal of Neuroscience*, *31*, 11934–11940. doi:10.1523/JNEUROSCI.1869-11.2011

Klucharev, V., Smidts, A., & Fernández, G. (2008). Brain mechanisms of persuasion: how 'expert power' modulates memory and attitudes. *Social Cognitive and Affective Neuroscience*, *3*, 353–366. doi:10.1093/scan/nsn022

Koban, L., & Wager, T. D. (2016). Beyond conformity: Social influences on pain reports and physiology. *Emotion*, *16*, 24–32. doi:10.1037/emo0000087

Kober, J., & Peters, J. (2012). Reinforcement learning in robotics: A survey. In M. Wiering, & M. van Otterlo (Eds.), *Reinforcement learning* (pp. 579–610). Springer Berlin Heidelberg. doi:10.1007/978-3-642-27645-3_18

Kok, A. (1997). Event-related-potential (ERP) reflections of mental resources: A review and synthesis. *Biological Psychology*, *45*, 19–56. doi:10.1016/S0301-0511(96)05221-0

Kreussel, L., Hewig, J., Kretschmer, N., Hecht, H., Coles, M. G., & Miltner, W. H. (2012). The influence of the magnitude, probability, and valence of potential wins and losses on the amplitude of the feedback negativity: Modulation of feedback negativity. *Psychophysiology*, *49*, 207–219. doi:10.1111/j.1469-8986.2011.01291.x

Krupka, E. L., & Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, *11*, 495–524. doi:10.1111/jeea.12006

Kuhr, B., Schomberg, J., Gruber, T., & Quirin, M. (2013). Beyond pleasure and arousal: Appetitive erotic stimuli modulate electrophysiological brain correlates of early attentional processing. *NeuroReport*, *24*, 246–250. doi:10.1097/WNR.0b013e32835f4eba

Kujawa, A., Smith, E., Luhmann, C., & Hajcak, G. (2013). The feedback negativity reflects favorable compared to nonfavorable outcomes based on global, not local, alternatives: Feedback negativity reflects global outcomes. *Psychophysiology*, *50*, 134–138. doi:10.1111/psyp.12002

Kutas, M., & 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*, 621–647. doi:10.1146/annurev.psych.093008.131123

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

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*, 161–163. doi:10.1038/307161a0

Latané, B. (1981). The psychology of social impact. *American Psychologist*, *36*, 343–356. doi:10.1037/0003-066X.36.4.343

Latinus, M., & Taylor, M. J. (2006). Face processing stages: Impact of difficulty and the separation of effects. *Brain Research*, *1123*, 179–187. doi:10.1016/j.brainres.2006.09.031

Lee, D., Seo, H., & Jung, M. W. (2012). Neural basis of reinforcement learning and decision making. *Annual Review of Neuroscience*, *35*, 287–308. doi:10.1146/annurev-neuro-062111-150512

Leue, A., & Beauducel, A. (2008). A meta-analysis of reinforcement sensitivity theory: On performance parameters in reinforcement tasks. *Personality and Social Psychology Review*, *12*, 353–369. doi:10.1177/1088868308316891

Lieberman, M. D. (2007). Social cognitive neuroscience: A review of core processes. *Annual Review of Psychology*, *58*, 259–289. doi:10.1146/annurev.psych.58.110405.085654

Lieberman, M. D. (2012). A geographical history of social cognitive neuroscience. *NeuroImage*, *61*, 432–436. doi:10.1016/j.neuroimage.2011.12.089

Lilienfeld, S. O., Sauvigné, K. C., Lynn, S. J., Cautin, R. L., Latzman, R. D., & Waldman, I. D. (2015). Fifty psychological and psychiatric terms to avoid: A list of inaccurate, misleading, misused, ambiguous, and logically confused words and phrases. *Frontiers in Psychology*, *6*, 1100. doi:10.3389/fpsyg.2015.01100

Littel, M., Euser, A. S., Munafò, M. R., & Franken, I. H. A. (2012). Electrophysiological indices of biased cognitive processing of substance-related cues: A meta-analysis. *Neuroscience & Biobehavioral Reviews*, *36*, 1803–1816. doi:10.1016/j.neubiorev.2012.05.001

Liu, Y., & Gehring, W. J. (2009). Loss feedback negativity elicited by single- versus conjoined-feature stimuli. *NeuroReport*, *20*, 632–636. doi:10.1097/WNR.0b013e32832a3250

Liu, Y., Nelson, L. D., Bernat, E. M., & Gehring, W. J. (2014). Perceptual properties of feedback stimuli influence the feedback-related negativity in the flanker gambling task. *Psychophysiology*, *51*, 782–788. doi:10.1111/psyp.12216

Liu, Y., Huang, H., McGinnis-Deweese, M., Keil, A., & Ding, M. (2012). Neural substrate of the late positive potential in emotional processing. *Journal of Neuroscience*, *32*, 14563–14572. doi:10.1523/JNEUROSCI.3109-12.2012

Logothetis, N. K. (2002). The neural basis of the blood-oxygen-level-dependent functional magnetic resonance imaging signal. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *357*, 1003–1037. doi:10.1098/rstb.2002.1114

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, *412*, 150–157. doi:10.1038/35084005

Luck, S. J. (2005). *An introduction to the event-related potential technique.* Cambridge, MA: MIT Press.

Luck, S. J., & Kappenman, E. S. (2011). *The Oxford handbook of event-related potential components.* Oxford, UK: Oxford University Press.

Maass, A., & Clark, R. D. (1983). Internalization versus compliance: Differential processes underlying minority influence and conformity. *European Journal of Social Psychology*, *13*, 197–215. doi:10.1002/ejsp.2420130302

Mackie, D. M., & Worth, L. T. (1989). Processing deficits and the mediation of positive affect in persuasion. *Journal of Personality and Social Psychology*, *57*, 27–40. doi:10.1037/0022-3514.57.1.27

Markus, H. R., & Kitayama, S. (1994). A collective fear of the collective: Implications for selves and theories of selves. *Personality and Social Psychology Bulletin*, *20*, 568–579. doi:10.1177/0146167294205013

Mars, R. B., de Bruijn, E. R. A., Hulstijn, W., Miltner, W. H. R., & Coles, M. G. H. (2004). What if I told you: “You were wrong”? Brain potentials and behavioral adjustments elicited by feedback in a time-estimation task. In *Errors, conflicts, and the brain. Current opinions on performance monitoring* (pp. 129–134). Dortmund, Germany: Max Planck Institute for Human Cognitive and Brain Sciences.

Martínez, A., Anllo-Vento, L., Sereno, M. I., Frank, L. R., Buxton, R. B., Dubowitz, D. J., ... Hillyard, S. A. (1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nature Neuroscience*, *2*, 364–369. doi:10.1038/7274

Mason, M. F., Dyer, R., & Norton, M. I. (2009). Neural mechanisms of social influence. *Organizational Behavior and Human Decision Processes*, *110*, 152–159. doi:10.1016/j.obhdp.2009.04.001

McClure, S. M., Berns, G. S., & Montague, P. R. (2003). Temporal prediction errors in a passive learning task activate human striatum. *Neuron*, *38*, 339–346. doi:10.1016/S0896-6273(03)00154-5

McDonald, R. I., & Crandall, C. S. (2015). Social norms and social influence. *Current Opinion in Behavioral Sciences*, *3*, 147–151. doi:10.1016/j.cobeha.2015.04.006

Mehrabian, A., & Stefl, C. A. (1995). Basic temperament components of loneliness, shyness, and conformity. *Social Behavior and Personality*, *23*, 253–263. doi:10.2224/sbp.1995.23.3.253

Milgram, S. (1961). Nationality and conformity. *Scientific American*, *205*, 45–52. doi:10.1038/scientificamerican1261-45

Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology*, *67*, 371.

Milgram, S. (1965a). Liberating effects of group pressure. *Journal of Personality and Social Psychology*, *1*, 127–134. doi:10.1037/h0021650

Milgram, S. (1965b). Some conditions of obedience and disobedience to authority. *Human relations*, *18*, 57–76.

Miltner, W. H. R., Braun, C. H., & Coles, M. G. H. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: Evidence for a “generic” neural system for error detection. *Journal of Cognitive Neuroscience*, *9*, 788–798. doi:10.1162/jocn.1997.9.6.788

Mischel, W. (2004). Toward an integrative science of the person. *Annual Review of Psychology*, *55*, 1–22. doi:10.1146/annurev.psych.55.042902.130709

Mojzisch, A., & Krug, K. (2008). Cells, circuits, and choices: Social influences on perceptual decision making. *Cognitive, Affective, & Behavioral Neuroscience*, *8*, 498–508. doi:10.3758/CABN.8.4.498

Moore, A. W., & Atkeson, C. G. (1993). Prioritized sweeping: Reinforcement learning with less data and less time. *Machine Learning*, *13*, 103–130. doi:10.1007/BF00993104

Moscovici, S., Lage, E., & Naffrechoux, M. (1969). Influence of a consistent minority on the responses of a majority in a color perception task. *Sociometry*, *32*, 365–380.

Muchnik, L., Aral, S., & Taylor, S. J. (2013). Social influence bias: A randomized experiment. *Science*, *341*, 647–651. doi:10.1126/science.1240466

Mussweiler, T., & Epstude, K. (2009). Relatively fast! Efficiency advantages of comparative thinking. *Journal of Experimental Psychology: General*, *138*, 1–21. doi:10.1037/a0014374

Mussweiler, T., & Posten, A.-C. (2012). Relatively certain! Comparative thinking reduces uncertainty. *Cognition*, *122*, 236–240. doi:10.1016/j.cognition.2011.10.005

Mussweiler, T., & Rüter, K. (2003). What friends are for! The use of routine standards in social comparison. *Journal of Personality and Social Psychology*, *85*, 467–481. doi:10.1037/0022-3514.85.3.467

Niederkrotenthaler, T., Voracek, M., Herberth, A., Till, B., Strauss, M., Etzersdorfer, E., … Sonneck, G. (2010). Role of media reports in completed and prevented suicide: Werther v. Papageno effects. *British Journal of Psychiatry*, *197*, 234–243. doi:10.1192/bjp.bp.109.074633

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus–norepinephrine system. *Psychological Bulletin*, *131*, 510–532. doi:10.1037/0033-2909.131.4.510

Nobre, A. C., Correa, Á., & Coull, J. T. (2007). The hazards of time. *Current Opinion in Neurobiology*, *17*, 465–470. doi:10.1016/j.conb.2007.07.006

Nook, E. C., & Zaki, J. (2015). Social norms shift behavioral and neural responses to foods. *Journal of Cognitive Neuroscience*, *27*, 1412–1426. doi:10.1162/jocn_a_00795

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*, 424–430. doi:10.1016/j.tics.2006.07.005

Nummenmaa, L., & Calder, A. J. (2009). Neural mechanisms of social attention. *Trends in Cognitive Sciences*, *13*, 135–143. doi:10.1016/j.tics.2008.12.006

Obleser, J., & Kotz, S. A. (2011). Multiple brain signatures of integration in the comprehension of degraded speech. *NeuroImage*, *55*, 713–723. doi:10.1016/j.neuroimage.2010.12.020

O'Donnell, M. B., & Falk, E. B. (2015). Linking neuroimaging with functional linguistic analysis to understand processes of successful communication. *Communication Methods and Measures*, *9*, 55–77. doi:10.1080/19312458.2014.999751

Ofan, R. H., Rubin, N., & Amodio, D. M. (2011). Seeing race: N170 responses to race and their relation to automatic racial attitudes and controlled processing. *Journal of Cognitive Neuroscience*, *23*, 3153–3161. doi:10.1162/jocn_a_00014

Ofan, R. H., Rubin, N., & Amodio, D. M. (2014). Situation-based social anxiety enhances the neural processing of faces: evidence from an intergroup context. *Social Cognitive and Affective Neuroscience*, *9*, 1055–1061. doi:10.1093/scan/nst087

Pascual-Marqui, R. D. (2002). Standardized low-resolution brain electromagnetic tomography (sLORETA): Technical details. *Methods and Findings in Experimental and Clinical Pharmacology*, *24*, 5–12.

Pauling, L., & Coryell, C. D. (1936). The magnetic properties and structure of hemoglobin, oxyhemoglobin and carbonmonoxyhemoglobin. *Proceedings of the National Academy of Sciences*, *22*, 210–216. doi:10.1073/pnas.22.4.210

Peeters, G., & Czapinski, J. (1990). Positive-negative asymmetry in evaluations: The distinction between affective and informational negativity effects. *European Review of Social Psychology*, *1*, 33–60. doi:10.1080/14792779108401856

Pessoa, L., Kastner, S., & Ungerleider, L. G. (2003). Neuroimaging studies of attention: From modulation of sensory processing to top-down control. *Journal of Neuroscience*, *23*, 3990–3998.

Petty, R. E., & Cacioppo, J. T. (1996). *Attitudes and persuasion: Classic and contemporary approaches*. Boulder, CO, USA: Westview Press.

Pezdek, K., & Eddy, R. M. (2001). Imagination inflation: A statistical artifact of regression toward the mean. *Memory & Cognition*, *29*, 707–718. doi:10.3758/BF03200473

Pfabigan, D. M., Alexopoulos, J., Bauer, H., Lamm, C., & Sailer, U. (2011). All about the money – external performance monitoring is affected by monetary, but not by socially conveyed feedback cues in more antisocial individuals. *Frontiers in Human Neuroscience*, *5*, 100. doi:10.3389/fnhum.2011.00100

Pfabigan, D. M., Alexopoulos, J., Bauer, H., & Sailer, U. (2011). Manipulation of feedback expectancy and valence induces negative and positive reward prediction error signals manifest in event-related brain potentials: FRN, P300, and prediction errors signals. *Psychophysiology*, *48*, 656–664. doi:10.1111/j.1469-8986.2010.01136.x

Pfabigan, D. M., Zeiler, M., Lamm, C., & Sailer, U. (2014). Blocked versus randomized presentation modes differentially modulate feedback-related negativity and P3b amplitudes. *Clinical Neurophysiology*, *125*, 715–726. doi:10.1016/j.clinph.2013.09.029

Phillips, D. P. (1974). The influence of suggestion on suicide: Substantive and theoretical implications of the werther effect. *American Sociological Review*, *39*, 340–354.

Phillips, D. P. (1980). Airplane accidents, murder, and the mass media: Towards a theory of imitation and suggestion. *Social Forces*, *58*, 1001–1024. doi:10.1093/sf/58.4.1001

Phillips, D. P., & Carstensen, L. L. (1986). Clustering of teenage suicides after television news stories about suicide. *New England Journal of Medicine*, *315*, 685–689. doi:10.1056/NEJM198609113151106

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

Polich, J., & Criado, J. R. (2006). Neuropsychology and neuropharmacology of P3a and P3b. *International Journal of Psychophysiology*, *60*, 172–185. doi:10.1016/j.ijpsycho.2005.12.012

Potts, G. F. (2004). An ERP index of task relevance evaluation of visual stimuli. *Brain and Cognition*, *56*, 5–13. doi:10.1016/j.bandc.2004.03.006

Potts, G. F., Martin, L. E., Burton, P., & Montague, P. R. (2006). When things are better or worse than expected: The medial frontal cortex and the allocation of processing resources. *Journal of Cognitive Neuroscience*, *18*, 1112–1119. doi:10.1093/cercor/bhh153

Potts, G. F., Patel, S. H., & Azzam, P. N. (2004). Impact of instructed relevance on the visual ERP. *International Journal of Psychophysiology*, *52*, 197–209. doi:10.1016/j.ijpsycho.2003.10.005

Proverbio, A. M., & Riva, F. (2009). RP and N400 ERP components reflect semantic violations in visual processing of human actions. *Neuroscience Letters*, *459*, 142–146. doi:10.1016/j.neulet.2009.05.012

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*, 79–87. doi:10.1038/4580

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108. doi:10.1037/0033-295X.85.2.59

Ratner, K. G., & Amodio, D. M. (2013). Seeing “us vs. them”: Minimal group effects on the neural encoding of faces. *Journal of Experimental Social Psychology*, *49*, 298–301. doi:10.1016/j.jesp.2012.10.017

Reicher, S., & Haslam, S. A. (2011). After shock? Towards a social identity explanation of the Milgram ‘obedience’ studies. *British Journal of Social Psychology*, *50*, 163–169. doi:10.1111/j.2044-8309.2010.02015.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. Prokaky (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Reuter, M., Cooper, A. J., Smillie, L. D., Markett, S., & Montag, C. (2015). A new measure for the revised reinforcement sensitivity theory: Psychometric criteria and genetic validation. *Frontiers in Systems Neuroscience*, *9*, 38. doi:10.3389/fnsys.2015.00038

Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science*, *306*, 443–447. doi:10.1126/science.1100301

Rilling, J. K., & Sanfey, A. G. (2011). The neuroscience of social decision-making. *Annual Review of Psychology*, *62*, 23–48. doi:10.1146/annurev.psych.121208.131647

Rohenkohl, G., Cravo, A. M., Wyart, V., & Nobre, A. C. (2012). Temporal expectation improves the quality of sensory information. *Journal of Neuroscience*, *32*, 8424–8428. doi:10.1523/JNEUROSCI.0804-12.2012

Rose, R. L., Bearden, W. O., & Teel, J. E. (1992). An attributional analysis of resistance to group pressure regarding illicit drug and alcohol consumption. *Journal of Consumer Research*, *19*, 1–13. doi:10.1086/209281

Ross, L., Greene, D., & House, P. (1977). The "false consensus effect": An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, *13*, 279–301. doi:10.1016/0022-1031(77)90049-X

Rossion, B., & Jacques, C. (2008). Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? Ten lessons on the N170. *NeuroImage*, *39*, 1959–1979. doi:10.1016/j.neuroimage.2007.10.011

Rugg, M. D., & Coles, M. G. H. (Eds.). (1995). *Electrophysiology of mind: Event-related brain potentials and cognition.* New York: Oxford University Press.

Rule, N. O., Adams, R. B., Ambady, N., & Freeman, J. B. (2012). Perceptions of dominance following glimpses of faces and bodies. *Perception*, *41*, 687–706. doi:10.1068/p7023

Russell, N. J. C. (2011). Milgram's obedience to authority experiments: Origins and early evolution. *British Journal of Social Psychology*, *50*, 140–162. doi:10.1348/014466610X492205

Sandholm, T. W., & Crites, R. H. (1996). Multiagent reinforcement learning in the Iterated Prisoner's Dilemma. *BioSystems*, *37*, 147–166. doi:10.1016/0303-2647(95)01551-5

Sarinopoulos, I., Grupe, D. W., Mackiewicz, K. L., Herrington, J. D., Lor, M., Steege, E. E., & Nitschke, J. B. (2010). Uncertainty during anticipation modulates neural responses to aversion in human insula and amygdala. *Cerebral Cortex*, *20*, 929–940. doi:10.1093/cercor/bhp155

Schacht, A., & Sommer, W. (2009). Emotions in word and face processing: Early and late cortical responses. *Brain and Cognition*, *69*, 538–550. doi:10.1016/j.bandc.2008.11.005

Schmid, P. C., & Amodio, D. M. (2016). Power effects on implicit prejudice and stereotyping: The role of intergroup face processing. *Social Neuroscience*. doi:10.1080/17470919.2016.1144647

Schnuerch, R., & Gibbons, H. (2014). A review of neurocognitive mechanisms of social conformity. *Social Psychology*, *45*, 466–478. doi:10.1027/1864-9335/a000213

Schnuerch, R., & Gibbons, H. (2015). Social proof in the human brain: Electrophysiological signatures of agreement and disagreement with the majority. *Psychophysiology*, *52*, 1328–1342. doi:10.1111/psyp.12461

Schnuerch, R., Koppehele-Gossel, J., & Gibbons, H. (2015). Weak encoding of faces predicts socially influenced judgments of facial attractiveness. *Social Neuroscience*, *10*, 624–634. doi:10.1080/17470919.2015.1017113

Schnuerch, R., Kreitz, C., & Lange, K. (2013). Independent effects of temporal expectation and stimulus intensity in audition. *Attention, Perception, & Psychophysics*, *75*, 1520–1532. doi:10.3758/s13414-013-0491-4

Schnuerch, R., Richter, J., Koppehele-Gossel, J., & Gibbons, H. (2016). Multiple neural signatures of social proof and deviance during the observation of other people's preferences. *Psychophysiology*, *53*, 823–836. doi:10.1111/psyp.12636

Schnuerch, R., Schnuerch, M., & Gibbons, H. (2015). Assessing and correcting for regression toward the mean in deviance-induced social conformity. *Frontiers in Psychology*, *6*, 669. doi:10.3389/fpsyg.2015.00669

Schnuerch, R., Trautmann-Lengsfeld, S. A., Bertram, M., & Gibbons, H. (2014). Neural sensitivity to social deviance predicts attentive processing of peer-group judgment. *Social Neuroscience*, *9*, 650–660. doi:10.1080/17470919.2014.934393

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological Science*, *18*, 429–434. doi:10.1111/j.1467-9280.2007.01917.x

Schultz, W. (1997). Dopamine neurons and their role in reward mechanisms. *Current Opinion in Neurobiology*, *7*, 191–197. doi:10.1016/S0959-4388(97)80007-4

Schultz, W. (2000). Multiple reward signals in the brain. *Nature Reviews Neuroscience*, *1*, 199–207. doi:10.1038/35044563

Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron*, *36*, 241–263. doi:10.1016/S0896-6273(02)00967-4

Schultz, W. (2006). Behavioral theories and the neurophysiology of reward. *Annual Review of Psychology*, *57*, 87–115. doi:10.1146/annurev.psych.56.091103.070229

Schultz, W. (2007). Reward signals. *Scholarpedia*, *2*, 2184. doi:10.4249/scholarpedia.2184

Schultz, W. (2010). Dopamine signals for reward value and risk: Basic and recent data. *Behavioral and Brain Functions*, *6*, 1–9. doi:10.1186/1744-9081-6-24

Schultz, W., & Dickinson, A. (2000). Neuronal coding of prediction errors. *Annual Review of Neuroscience*, *23*, 473–500. doi:10.1146/annurev.neuro.23.1.473

Schupp, H. T., Cuthbert, B. N., Bradley, M. M., Cacioppo, J. T., Ito, T. A., & Lang, P. J. (2000). Affective picture processing: The late positive potential is modulated by motivational relevance. *Psychophysiology*, *37*, 257–261. doi:10.1111/1469-8986.3720257

Schweinberger, S. R., Pickering, E. C., Jentzsch, I., Burton, A. M., & Kaufmann, J. M. (2002). Event-related brain potential evidence for a response of inferior temporal cortex to familiar face repetitions. *Cognitive Brain Research*, *14*, 398–409. doi:10.1016/S0926-6410(02)00142-8

Serences, J. T. (2008). Value-based modulations in human visual cortex. *Neuron*, *60*, 1169–1181. doi:10.1016/j.neuron.2008.10.051

Shah, A. S., Bressler, S. L., Knuth, K. H., Ding, M., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2004). Neural dynamics and the fundamental mechanisms of event-related brain potentials. *Cerebral Cortex*, *14*, 476–483. doi:10.1093/cercor/bhh009

Sherif, M. (1935). A study of some social factors in perception. *Archives of Psychology*, *27*, 23–46.

Sherman, L. E., Payton, A. A., Hernandez, L. M., Greenfield, P. M., & Dapretto, M. (2016). The power of the like in adolescence: Efects of peer influence on neural and behavioral responses to social media. *Psychological Science*. doi:10.1177/095679761 6645673

Sherman, S. J., Presson, C. C., & Chassin, L. (1984). Mechanisms underlying the false consensus effect the special role of threats to the self. *Personality and Social Psychology Bulletin*, *10*, 127–138. doi:10.1177/0146167284101015

Shestakova, A., Rieskamp, J., Tugin, S., Ossadtchi, A., Krutitskaya, J., & Klucharev, V. (2013). Electrophysiological precursors of social conformity. *Social Cognitive and Affective Neuroscience*, *8*, 756–763. doi:10.1093/scan/nss064

Shiraev, E. B. (2014). *A history of psychology: A global perspective* (2nd ed.). Thousand Oaks, CA: Sage.

Shuler, M. G., & Bear, M. F. (2006). Reward timing in the primary visual cortex. *Science*, *311*, 1606–1609. doi:10.1126/science.1123513

Singer, T. (2012). The past, present and future of social neuroscience: A European perspective. *NeuroImage*, *61*, 437–449. doi:10.1016/j.neuroimage.2012.01.109

Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., … Sanchez-Vives, M. V. (2006). A virtual reprise of the Stanley Milgram obedience experiments. *PLOS ONE*, *1*, e39. doi:10.1371/journal.pone.0000039

Stallen, M., & Sanfey, A. G. (2015). The neuroscience of social conformity: Implications for fundamental and applied research. *Frontiers in Neuroscience*, *9*, 337. doi:10.338 9/fnins.2015.00337

Stallen, M., Smidts, A., & Sanfey, A. G. (2013). Peer influence: Neural mechanisms underlying in-group conformity. *Frontiers in Human Neuroscience*, *7*, 50. doi:10.3389 /fnhum.2013.00050

Steinberg, L. (2008). A social neuroscience perspective on adolescent risk-taking. *Developmental Review*, *28*, 78–106. doi:10.1016/j.dr.2007.08.002

Stigler, S. M. (1997). Regression towards the mean, historically considered. *Statistical Methods in Medical Research*, *6*, 103–114. doi:10.1177/096228029700600202

Stokes, M. G., Atherton, K., Patai, E. Z., & Nobre, A. C. (2011). Long-term memory prepares neural activity for perception. *Proceedings of the National Academy of Sciences*, E360–E367. doi:10.1073/pnas.1108555108

Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: Neural and computational mechanisms. *Nature Reviews Neuroscience*, *15*, 745–756. doi:10.1038/nrn3838

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, *13*, 403–409. doi:10.1016/j.tics.2009.06.003

Sutton, S., & Ruchkin, D. S. (1984). The late positive complex. *Annals of the New York Academy of Sciences*, *425*, 1–23. doi:10.1111/j.1749-6632.1984.tb23520.x

Tanford, S., & Penrod, S. (1984). Social influence model: A formal integration of research on majority and minority influence processes. *Psychological Bulletin*, *95*, 189–225. doi:10.1037/0033-2909.95.2.189

Thorndike, E. L. (1927). The law of effect. *American Journal of Psychology*, *39*, 212–222.

Thorndike, E. L. (1933). A proof of the law of effect. *Science*, *77*, 173–175. doi:10.1126/science.77.1989.173-a

Thorndike, E. L. (1911). *Animal intelligence: Experimental studies.* London, UK: Macmillan.

Titchener, E. B. (1921). Wilhelm Wundt. *American Journal of Psychology*, *32*, 161–178.

Toelch, U., & Dolan, R. J. (2015). Informational and normative influences in conformity from a neurocomputational perspective. *Trends in Cognitive Sciences*, *19*, 579–589. doi:10.1016/j.tics.2015.07.007

Tong, E. M., Tan, C. R., Latheef, N. A., Selamat, M. F., & Tan, D. K. (2008). Conformity: Moods matter. *European Journal of Social Psychology*, *38*, 601–611. doi:10.1002/ejsp.485

Torrubia, R., Avila, C., Moltó, J., & Caseras, X. (2001). The sensitivity to punishment and sensitivity to reward questionnaire (SPSRQ) as a measure of Gray's anxiety and impulsivity dimensions. *Personality and Individual Differences*, *31*, 837–862.

Tracy, J. L., Shariff, A. F., & Cheng, J. T. (2010). A naturalist's view of pride. *Emotion Review*, *2*, 163–177. doi:10.1177/1754073909354627

Trautmann-Lengsfeld, S. A., & Herrmann, C. S. (2013). EEG reveals an early influence of social conformity on visual processing in group pressure situations. *Social Neuroscience*, *8*, 75–89. doi:10.1080/17470919.2012.742927

Trautmann-Lengsfeld, S. A., & Herrmann, C. S. (2014). Virtually simulated social pressure influences early visual processing more in low compared to high autonomous participants. *Psychophysiology*, *51*, 124–135. doi:10.1111/psyp.12161

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131. doi:10.1126/science.185.4157.1124

van den Brink, D., Berkum, J. J. A. V., Bastiaansen, M. C. M., Tesink, C. M. J. Y., Kos, M., Buitelaar, J. K., & Hagoort, P. (2012). Empathy matters: ERP evidence for inter-individual differences in social language processing. *Social Cognitive and Affective Neuroscience*, *7*, 173–183. doi:10.1093/scan/nsq094 van Hooff, J. C., Crawford, H., & van Vugt, M. (2011). The wandering mind of men: ERP evidence for gender differences in attention bias towards attractive opposite sex faces. *Social Cognitive and Affective Neuroscience*, *6*, 477–485. doi:10.1093/scan/nsq066

Voss, A., Nagler, M., & Lerche, V. (2013). Diffusion models in experimental psychology: a practical introduction. *Experimental Psychology*, *60*, 385–402. doi:10.1027/1618-3169/a000218

Voss, A., Rothermund, K., & Brandtstädter, J. (2008). Interpreting ambiguous stimuli: Separating perceptual and judgmental biases. *Journal of Experimental Social Psychology*, *44*, 1048–1056. doi:10.1016/j.jesp.2007.10.009

Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly high correlations in fmri studies of emotion, personality, and social cognition. *Perspectives on Psychological Science*, *4*, 274–290. doi:10.1111/j.1745-6924.2009.01125.x

Wagner, M., Fuchs, M., & Kastner, J. (2004). Evaluation of sLORETA in the presence of noise and multiple sources. *Brain Topography*, *16*, 277–280. doi:10.1023/B:BRAT.0000032865.58382.62

Walther, E., Bless, H., Strack, F., Rackstraw, P., Wagner, D., & Werth, L. (2002). Conformity effects in memory as a function of group size, dissenters and uncertainty. *Applied Cognitive Psychology*, *16*, 793–810. doi:10.1002/acp.828

Wasserman, I. M. (1984). Imitation and suicide: A reexamination of the Werther effect. *American Sociological Review*, 427–436. doi:10.2307/2095285

Weinberg, A., & Hajcak, G. (2011). The late positive potential predicts subsequent interference with target processing. *Journal of Cognitive Neuroscience*, *23*, 2994–3007. doi:10.1162/jocn.2011.21630

West, R., Bailey, K., Anderson, S., & Kieffaber, P. D. (2014). Beyond the FN: A spatiotemporal analysis of the neural correlates of feedback processing in a virtual Blackjack game. *Brain and Cognition*, *86*, 104–115. doi:10.1016/j.bandc.2014.02.003

West, R., Bailey, K., Tiernan, B. N., Boonsuk, W., & Gilbert, S. (2012). The temporal dynamics of medial and lateral frontal neural activity related to proactive cognitive control. *Neuropsychologia*, *50*, 3450–3460. doi:10.1016/j.neuropsychologia.2012.10.011

Wood, W. (2000). Attitude change: Persuasion and social influence. *Annual Review of Psychology*, *51*, 539–570. doi:10.1146/annurev.psych.51.1.539

Xu, Q., Shen, Q., Chen, P., Ma, Q., Sun, D., & Pan, Y. (2011). How an uncertain cue modulates subsequent monetary outcome evaluation: An ERP study. *Neuroscience Letters*, *505*, 200–204. doi:10.1016/j.neulet.2011.10.024

Yeung, N., & Sanfey, A. G. (2004). Independent coding of reward magnitude and valence in the human brain. *Journal of Neuroscience*, *24*, 6258–6264. doi:10.1523/JNEUROSCI.4537-03.2004

Yoo, S.-S., Talos, I.-F., Golby, A. J., Black, P. M., & Panych, L. P. (2004). Evaluating requirements for spatial resolution of fMRI for neurosurgical planning. *Human Brain Mapping*, *21*, 34–43. doi:10.1002/hbm.10148

Yu, R., & Chen, L. (2015). The need to control for regression to the mean in social psychology studies. *Frontiers in Psychology*, *5*, 1574. doi:10.3389/fpsyg.2014.01574

Yu, R., & Sun, S. (2013). To conform or not to conform: Spontaneous conformity diminishes the sensitivity to monetary outcomes. *PLoS ONE*, *8*, e64530. doi:10.1371/journal.pone.0064530

Zajonc, R. B. (1965). Social facilitation. *Science*, *149*, 269–274. doi:10.1126/science.149.3681.269

Zaki, J., Schirmer, J., & Mitchell, J. P. (2011). Social influence modulates the neural computation of value. *Psychological Science*, *22*, 894–900. doi:10.1177/0956797611411057

Zebrowitz, L. A., & Collins, M. A. (1997). Accurate social perception at zero acquaintance: The affordances of a Gibsonian approach. *Personality and Social Psychology Review*, *1*, 204–223. doi:10.1207/s15327957pspr0103_2

Zhang, W., Lu, J., Ni, Z., Liu, X., Wang, D., & Shen, J. (2013). Harm avoidance in adolescents modulates late positive potentials during affective picture processing. *International Journal of Developmental Neuroscience*, *31*, 297–302. doi:10.1016/j.ijdevneu.2013.03.009

Zhang, Y., Li, X., Qian, X., & Zhou, X. (2012). Brain responses in evaluating feedback stimuli with a social dimension. *Frontiers in Human Neuroscience*, *6*, 29. doi:10.3389/fnhum.2012.00029

Zhang, Z., & Deng, Z. (2012). Gender, facial attractiveness, and early and late event-related potential components. *Journal of Integrative Neuroscience*, *11*, 477–487. doi:10.1142/S0219635212500306

A Appendix

The present thesis is based upon eight original publications (see Table 1). To avoid copyright infringement, the appendix containing these articles has been removed from the online version of this thesis. The articles can be found online via the respective publishers or in the hardcopies of the present thesis at the *Universitäts- und Landesbibliothek Bonn.*

[Die vorliegende Dissertation basiert auf acht Originalpublikationen (siehe Tabelle 1). Aus urheberrechtlichen Gründen ist der Anhang mit diesen Artikeln aus der Online-Version dieser Dissertation entfernt worden. Die Artikel sind online bei den jeweiligen Verlagen oder in den Druckexemplaren der vorliegenden Dissertation an der *Universitäts- und Landesbibliothek Bonn* verfügbar.]