Doctoral thesis submitted to
the Faculty of Behavioural and Cultural Studies
Heidelberg University
in partial fulfillment of the requirements of the degree of
Doctor of Philosophy (Dr. phil.)
in Psychology

Title of the publication-based thesis

Communication as a Meta-Theory for Judgment and Decision-Making

presented by
Karolin Salmen

year of submission
2023

Dean: Prof. Dr. Guido Sprenger
Advisor: Prof. Dr. Klaus Fiedler

Acknowledgments

This thesis is dedicated to those who have supported, challenged and influenced me.

To the academics whose example and mentorship shaped my understanding of research and drew me toward social psychology. To name only a few, Prof. Dr. Gerald Echterhoff roped me into the shared reality approach before I had even written my first university exam and set me on the path that led here. Prof. Dr. Florian Kutzner gave me the best possible introduction to JDM. And, most importantly, Prof. Dr. Klaus Fiedler. You taught me so much through the open-hearted exchange of ideas and views and countless hours of discussions, sharing your passion for research, teaching, and collaboration, and your unwavering kindness and optimism. Thank you for everything.

To fellow and former CRISPies, particularly Florian Ermark, without whom my time at the PI would have been so much less – collaborative, thought-provoking, friendly, and fun. And to Ute Lorenz, thank you for your skill, dependability, humor, and friendliness that let me enjoy the time in your office even when faced with the absurdities of bureaucracy.

To the institutions which made it possible for me to study abroad and learn foreign languages, particularly the Studienstiftung des deutschen Volkes and the DAAD. These experiences shaped my understanding of communication, language, and semiotics far beyond the intellectual level.

To my friends, and everyone else who listened patiently to my enthusiastic rambling and frustrated rants. Especially to Jakob Mertesacker, Anna Hänig, and Miriam Pfister, whom I can always depend on. Thank you for putting up with me – and for understanding me so well.

To my family, who still wonder about what it might be that I do, exactly, and still gifted me with unwavering support. Thank you for trusting me to know my own way and to adapt to any challenge, in the end.

Content

1. Introduction

Within the behavioral sciences, a substantial and influential body of research concerns *judgment and decision-making* (JDM). In general terms, this field investigates the process of *decision-making*, with its three core elements: evaluations or estimates of alternatives (*judgments*), how they are weighed according to personal characteristics (*preferences*) and integrated to pursue a course of action (*choice*; Fischhoff & Broomell, 2020). The psychological inquiry into JDM has been dominated by economic theories (mainly expected utility theory, Von Neumann & Morgenstern, 1944), social cognition (particularly dual process theories, e.g., Chaiken, 1980) and ecological approaches (e.g., Gigerenzer & Selten, 2002). Meanwhile, the investigation of communication processes only played a minor role in JDM theorizing, with some notable exceptions (e.g., Hilton, 1995).

Spanning several phenomena and experimental paradigms central to JDM research, this thesis uses influential conceptualizations of communication introduced in Chapter 2 for novel theorizing about judgment and decision-making. In Chapter 3, seeing the research procedure as communication suggests that participants expect the researcher to communicate cooperatively and thus give responses corresponding to a classic cognitive illusion, the hindsight bias. Chapter 4 then focuses on competitive uses of communication by subtle means and how they can be used to discourage participants from following a testable, normative mental model for decision-making. Extending beyond the influence of cooperative and competitive pragmatics, Chapter 5 explores how a focus on the structure of the sign system, particularly that concepts may share the same cues, sheds new light on illusory correlations in judgments. In summary, this work furthers a communication perspective within the theoretical landscape of JDM research, in which social cognition, economic theories, and the ecological approach prevail.

2. Theorizing about Judgment and Decision-Making

An extensive review of the rich landscape of JDM research, as well as an exhaustive discussion of the manifold views on communication that may be relevant to this field, is beyond the scope of this thesis. Instead, this chapter will highlight some influential conceptualizations of communication that inspired theorizing in the empirical work presented in the following chapters. The article “Major Theories in Social Psychology” (Fiedler & Salmen, 2021) then provides an overview of theoretical landscape of social psychology and JDM research, and shows examples of communication as a meta-theory in other domains of social psychology.

2.1. Communication as a Meta-Theory

Some constructs and structural relationships between them appear again and again in theories of widely different explananda – they represent a *meta-theory*, an overarching theme in scientific theorizing. For instance, the elaboration likelihood model of attitude change (Petty & Cacioppo, 1986) and dual-process theories of JDM (for an overview, see, e.g., Chaiken & Trope, 1999) both use terms of information processing, storage and retrieval (see the article below for more examples). These theories belong to different domains of inquiry but are also parallel in the sense of a simile or metaphor and connect to other applications of principles from information theory across scientific disciplines. Such a meta-theoretical metaphor provides a lens, which focuses our view on aspects that may have never caught our attention when approaching from other theoretical viewpoints. The meta-theory provides a structural pattern – the interplay between adopting and adjusting this structure provides a fertile ground for creative theorizing and experimentation.

This thesis presents communication as one among many useful meta-theoretical lenses in socio-psychological research. Indeed, it is virtually impossible to come across an example of a phenomenon of interest to social psychology that does not involve communication – in and outside the laboratory. Consequently, recognizing the role of communication and focusing on communicative mediators and moderators can add valuable insight. Take a prototypical example of decision-making, such as buying a house. It is apparent how judgment, preferences and choice are embedded in the meaning-making process of communication. A buyer learns about options from what owners and realtors write and answer to their questions, gathers judgments and advice on how to weigh different features of the houses from others, and may even make a choice jointly with their family. In JDM research, paradigms employed in the study of advice taking and group decision-making – to name only two examples – are often inherently communicative, as participants exchange information and integrate it to make their choices. While communicative processes are an important part of theorizing on group decision making (see, e.g., Schulz-Hardt & Mojzisch, 2012), observations in advice taking research are mainly explained in terms of cognitive mechanisms and the striving for accuracy to maximize rewards (see, e.g., Bonaccio & Dalal, 2006). This prominence of individual information processing and behaviorist theories is even more pronounced in the vast field of cognitive illusions research (see, e.g., Pohl, 2022). As informative and important as these theoretical lenses are, they often ignore the striving for belonging, cooperation and a shared reality achieved through communication. They also commonly ignore many of the challenges that decision-makers face in less cooperative communication contexts inside and outside the laboratory, where they encounter highly indeterminate, misleading and even deceptive messages (see Chapter 4).

Importantly, many contributions even exhibit a blind spot to the fact that whenever someone becomes a participant in an experiment and answers questions, as in the vast majority of JDM experiments, participant and experimenter communicate with each other. In some cases, ignoring the communicative processes at work within any self-report study can lead to incomplete or even unwarranted inferences from experimental observations (some examples are presented in Chapter 3). The basic insight that what researchers intend to say may not be what participants understand, and that these deviations are systematic and can be anticipated, should motivate every JDM researcher to hone their awareness for communicative processes.

Which aspects do we focus on when we use the meta-theoretical lens of communication?

For a start, communication involves a process in which two sides engage to share and construct meaning. What one side means may differ from what the other side understands – to repeat the popular tenet, the more one studies communication, the more it seems like a miracle that we manage to construct common meaning at all. Among the many ways to conceptualize this joint meaning-making process, the Shannon-Weaver model (Shannon & Weaver, 1949) is one of the most influential. It describes communication as a one-way process of information transmission, in which a sender (source) encodes information in a message (signal) that a receiver then decodes to recover the information (see Figure 1). This understanding, a very selective simplification of the original mathematical model, is a popular assumption behind social cognition approaches – it reduces the communicative process to individual cognitive processes of encoding, sending, receiving, and decoding.

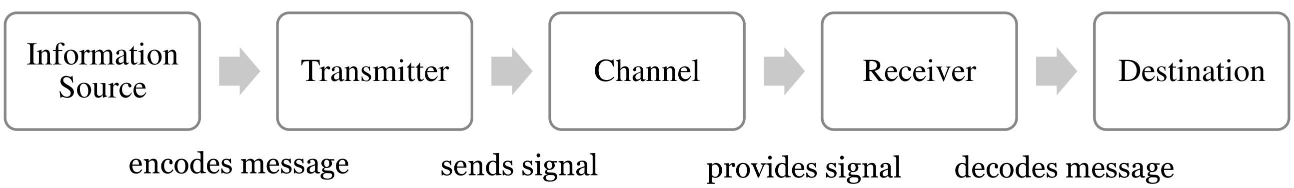


Figure 1. Shannon-Weaver model of communication.

However, this reductionist approach ignores the core part of the Shannon-Weaver model – to transmit information, a communicator depends on the use of a set of signs that the addressee can decode and a (noisy) channel that carries the signal. In the overwhelming majority of JDM research, this is achieved by using language. Typically, participants write in their first language – rarely, as in some group tasks, participants engage in face-to-face spoken conversation. When we accept the central role of language for JDM in and outside the laboratory, it becomes a natural next step to utilize knowledge from linguistics and communication research to deepen our understanding of decision-making.

The Shannon-Weaver-model is helpful to remind us how what is meant and what is understood may come to differ. Nevertheless, the empirical work in this thesis also aims to show that it can be highly valuable to swap this model from information theory, which pictures communication as unidirectional and information as given within the sender, for approaches that put the joint construction of meaning and the social functions of communication into focus. One of the most influential on psychology is the late work of Wittgenstein (1953) on *language games*. To take Wittgenstein's most famous example, imagine that someone says: "Water!" In a restaurant, when a waiter is taking orders, everyone will understand this as a request for a glass of water. Now imagine that the person exclaims this when they draw a card in a board game. Will the other players fetch a glass of water? Or will they rather understand that their fellow player has drawn a card that represents a much-needed resource for their gameplay? This simple example demonstrates that the meaning of an utterance is inseparably entwined with the context it occurs in. Chapter 3 will discuss this conceptualization of communication and its influence on psychological research in more detail.

In this example, communication is far from a mere exchange of information – in many cases, it does not only change the mind of the listener but is an extension of and preparation for action (Hörmann, 1976). Taking this notion seriously, communication becomes an adaptive tool to achieve goals in a social world (Fiedler, 2008). This attitude is embodied in *speech act theory* as a functionalist approach to communication that focuses on "how to do things with words" (Austin, 1962). It classifies utterances according to their consequences for speakers, listeners and bystanders, with some types of speech acts that are of particular interest to decision-making, such as advice giving, negotiation, persuasion, and deception. This approach focuses the theoretical lens on communication as goal-directed, that may not always (expected to) be an innocent and cooperative exchange of meaning and that may hinder rationality in intended and unintended ways, as is discussed further in Chapter 4.

So far, the discussed approaches focus on the pragmatics of language use within social contexts – what one could term *parole* in the terms of Saussure (1964). While this focus may come natural to social psychologists, this thesis also aims to show that guiding our attention to the structure of the language proper, *la langue*, offers new possibilities for JDM theorizing and research. While language is shaped by its users and the functions it fulfills, it is undeniable that language also shapes what we know about the world and how we think about it (see Fiedler, 2008, for a discussion of the complex interplay between language, cognition, and environment). The signs we use in joint meaning-making (semiosis) relate to the things they represent, their users and each other in a structured, systematic way that is the object of *semiotics*. Even if signs frequently are verbal, it is possible to apply a similar semiotic analysis to any situation in which something stands for something else (see, e.g., Morris, 1946; Eco, 1979). Such a semiotic approach to theorizing about judgments is demonstrated and discussed further in Chapter 5.

The presented models of communication guide our theorizing towards communication as information transmission, context-based, goal-directed action, and shaped by the sign system – a list that the reader should understand as open-ended. The following article helps situate communication as a meta-theory in the landscape of theorizing about JDM. It also shows how models of communication and language have already inspired theorizing in other areas of social psychology. For instance, construal level theory (Trope & Liberman, 2003) shows how abstract and concrete language interacts with several other dimensions central to motivation regulation. The following article presents this theory in more detail. Meanwhile, a matrix of major theories in social psychology, which differentiates between areas of inquiry and employed meta-theories, shows a blank spot for the combination of JDM and communication. The subsequent chapters aim to demonstrate that this does not indicate limited generative power of communication as a meta-theory for JDM research, but rather that there is still much to explore and contribute using the lens of communication theories.

2.2. Major Theories in Social Psychology (Fiedler & Salmen, 2021)[1]

> A synopsis of major theories of social psychology is provided with reference to three major domains of social-psychological inquiry: attitudes and attitude change, motivation regulation, and group behavior. Despite the heterogeneity of research topics, there is considerable overlap in the basic theoretical principles across all three domains. Typical theories that constitute the common ground of social psychology rely on rules of good Gestalt consistency, on psychodynamic principles, but also on behaviorist learning models and on semantic-representation and information-transition models borrowed from cognitive science. Prototypical examples that illustrate the structure and the spirit of theories in social psychology are dissonance theory, construal-level, regulatory focus, and social identity theory. A more elaborate taxonomy of pertinent theories is provided in the first table in this article.

Any attempt to provide a comprehensive review of all theories proposed in more than half a century of flourishing research in social psychology would exceed the scope of the present article. It is necessarily confined to a selected set of most impactful theories, which are most characteristic of the social psychological approach. While the diverse literature published in social psychological journals covers virtually all kinds of theories—from psychodynamic to behaviorist, sociological,

---

[1] This material was originally published in *Oxford Research Encyclopedia of Psychology* edited by Johnsrude and has been reproduced by permission of Oxford University Press. For permission to reuse this material, please visit http://global.oup.com/academic/rights.

neurobiological, evolutionary, and psychonomic approaches inspired by cognitive psychology and computer science—the typical, formative theories that define the discipline of social psychology are clearly distinct from other disciplines. They differ from the functional models in psychophysics (Gopher & Kimchi, 1989), from model-fitting in cognitive science (Trueblood & Busemeyer, 2011), the genetic and biological theories in developmental psychology (Zhang & Meaney, 2010), or the mathematical models in game theory and behavioral economics (Ortmann & Rydval, 2004).

Theories that constitute common ground among social psychologists rarely rely on formal notation implemented in computer or robot systems, or on axiomatic or fundamental laws. Rather, they are typically encoded in ordinary language used to subsume behavioral phenomena as manifestations of a basic set of intrapsychic needs, goals, motives, reasons, dispositions, and knowledge structures. Despite their frequently expressed emphasis on situational influences (Zimbardo, 2006) and their aspiration to go beyond the individual to groups and cultures as truly social units of analysis (Frable, 1997), there is no doubt that almost every theory has the individual as its focus. The power of the situation is typically treated as part of an individual's world knowledge; group dynamics and intergroup relations are explained in terms of individuals' striving for social identity (Ellemers et al., 2002); and even cultural determinants of social behavior are treated as culture-specific differences in world view and the self (Kitayama & Miyamoto, 2000).

Because of this focus on individual-level concepts with common meanings in ordinary language, theoretical accounts often resemble laypeople's explanations of everyday behavior. It is no coincidence that most public discussions of psychological findings in the media and in popular science refer to social psychological topics, such as attitudes, stereotypes, aggression, altruism, or gender roles. The technical meaning of these concepts is often overlooked, not only in lay-epistemological discourse but sometimes even in scientists' circular explanations in terms of goals or motives that sound like synonyms of the behaviors to be explained.

Overcoming the temptation to "explain" hostile behavior by a hostile disposition or cooperative behavior by a sense of fairness, and moving from intra-psychic to more inclusive ecological theories at a higher level of explanatory distance (Fiedler, 2014), is therefore an essential criterion of good theorizing in social psychology. Conversely, showing that meaningful social behaviors lend themselves to the same theoretical scrutiny and precision as basic physiological variables and environmental stimulus conditions constitutes an ultimate challenge for theory-minded social psychologists.

A Framework for Social Psychological Theories

A convenient and instructive framework to represent the common and distinctive features of theories must consider two facets: a set of theoretical principles or metaphors and a set of behavioral domains (Roediger, 1980). The latter facet, behavioral domains, consists of an open-ended list of research topics, such as attitudes, stereotypes, causal inferences, aggressive behaviors, or group decisions. All of these topics can be explained and systematized in various ways, with reference to distinct principles or metaphors, which are often borrowed or inferred by analogy from physics (balance, activation), chemistry (diffusion), geometry (symmetry, Euclidean distance), propositional logic (subsumption, implication), statistics (contingency, redundancy), physiology (synapsis, reflex), computer science (store, retrieval), and other disciplinary sources. No theory can claim to afford a unique account of a particular topic or domain. The fertility of the most successful theories often reflects the joint operation of more than a single principle, and different theories overlap to varying degrees in their underlying principles, as will soon be apparent. Theories that share the same principles (e.g., associations, co-occurrence) are related and likely to converge in their implications. More interesting perhaps is the phenomenon that non- overlapping theories that are rooted in different principles may also converge.

Table 1. Meta-theoretical approaches across areas of social psychology, with prominent theory exemplars

	ATTITUDES AND ATTITUDE CHANGE	PERSON PERCEPTION	JUDGMENT AND DECISION-MAKING	MOTIVATION AND BEHAVIOUR	SOCIAL GROUPS
GESTALT PSYCHOLOGY AND PSYCHODYNAMIC THEORIES	Balance Theory (Heider, 1946) Social Comparison Theory (Festinger, 1954) Cognitive Dissonance (Festinger, 1957) Reactance Theory (Brehm, 1966)	Attribution Theory (Heider, 1958)	Majority Conformity (Asch & Guetzkow, 1951)	Field Theory (Lewin, 1951) Rubicon Model (Heckhausen & Gollwitzer, 1987) Attribution Theory of Motivation (Weiner, 1985) Terror Management Theory (Greenberg et al., 1986) Need for Closure (Kruglanski, 1990)	Theory of Cooperation and Competition (Deutsch, 1949) Need to Belong (Baumeister & Leary, 1995)
LEARNING AND BEHAVIORISM	Classical Conditioning (Pavlov, 1927) Operant Conditioning (Skinner, 1938) Instrumental learning (Hovland et al., 1953)		Expected Utility Theory (Von Neumann & Morgenstern, 1944)	Social Learning Theory (Bandura, 1977) Neoassociation Theory of Aggression (Berkowitz, 1969)	Interdependence Theory (Thibaut & Kelley, 1959)
(SELF-) CATEGORISATION	Social Identity Theory (Tajfel, 1978) Social Role Theory (Eagly, 1978)	Infra- and Dehumanisation (Haslam, 2006) Continuum Model (Fiske & Neuberg, 1990)		Self-Categorization Theory (Turner et al., 1987)	Social Identity Theory (Tajfel, 1978) Optimal Distinctiveness Theory (Brewer, 1991) Uncertainty-Identity Theory (Hogg, 2007) Common ingroup identity model (Gaertner et al., 1993)
INFORMATION THEORY AND SOCIAL COGNITION	Elaboration Likelihood Model (Petty & Cacioppo, 1986)	Covariation Model (Kelley, 1973) Social Information Processing Theory (Wyer & Srull, 1986) Stereotype Content Model (Fiske et al., 2002)	Prospect Theory (Kahneman & Tversky, 1979) Dual-process theories: (Chaiken, 1980) (Strack & Deutsch, 2004) (Kahneman, 2003)	Theory of Planned Behavior (Ajzen, 1985) Action Identification Theory (Vallacher & Wegner, 1987)	Intergroup Emotions Theory (Smith, 1993)

			Uni-Model (Kruglanski & Thompson, 1999) Feelings-as-Information Theory (Schwarz & Clore, 1983)		
ADAPTIVE COGNITION / ECOLOGICAL APPROACH	Lens Model (Brunswik, 1952) Illusory Correlations (Chapman, 1967)	Sampling Approaches (Fiedler & Juslin, 2006)	Adaptive Toolbox (Gigerenzer & Selten, 2002)	Focus Theory of Normative Conduct (Cialdini et al., 1990)	
COMMUNICATION	Social Transmission (e.g., Kashima, 2000) Iterated Learning (Kalish et al., 2007) Communication Game Approach (Higgins, 1981) Shared Reality Theory (Hardin & Higgins, 1996)	Linguistic Category Model (Semin & Fiedler, 1991)		Construal Level Theory (Trope & Liberman, 2003) Goal Setting Theory (Locke & Latham, 1990)	Communication Accommodation Theory (Giles et al., 1991)
CONSTRUCTIVIST, COMPARATIVE AND CULTURAL PSYCHOLOGY	Generalized Other Hypothesis (Mead, 1934) Social Representation Theory (Moscovici, 1976)	Implicit Theories (Dweck, 1999)	Accessibility Theory (Bruner, 1957) Minority Influence Theory (Nemeth, 1986)	Regulatory-Focus Theory (Higgins, 1998) Individualism/ Collectivism (Markus & Kitayama, 1991)	Ingroup Projection Model (Mummendey & Wenzel, 1999) Integrated Threat Theory (Stephan & Renfro, 2002) System Justification Theory (Jost & Banaji, 1994)
BIOLOGICAL THEORIES: NEUROCOGNITION AND EVOLUTION	Evaluative Space Model (Cacioppo & Berntson, 1994)			Behavioral activation/ inhibition system (Carver & White, 1994) Tend and befriend theory (Taylor et al., 2000)	
HUMANISTIC AND POSITIVE PSYCHOLOGY	Client-centered Approach (Rogers, 1951)			Self-determination Theory (Deci & Ryan, 1987)	

While the remainder of this article is devoted to a discussion of only a few most prominent and influential theories, as mentioned, a synopsis of a more inclusive set of social-psychological theories is given in Table 1. In this two-dimensional schema, each theory is located in the intersection of its most typical domain (column) and its leading principle (row). Let us now enter the discussion of competing or complimentary theories in three major domains: attitudes, motivation regulation, and social groups.

Attitudes and Attitude Change

> "The concept of attitude is probably the most distinctive and indispensable concept in contemporary American social psychology"
> (Allport, 1935, p. 798).

When reading introductory texts on attitudes in social psychology, one usually finds this or a similar citation, maybe with the remark that "American" can easily be removed, and that this statement is as true in the early 21st century as it was a century earlier. But what is an attitude? The commonplace understanding is that an attitude is held by an individual, is directed toward an object, and is evaluative—it can be positive or negative. Apart from this minimal (triadic) definition, what is considered an attitude depends as much on the choice of theory as on measurement tools, experimental paradigms, and statistics.

Some theoretical approaches focus on how stable attitudes are formed, structured, and maintained, and how they relate to manifest behavior, whereas others focus on how attitudes change, particularly under social influence. In any case, most pertinent research refers to attitude objects with high societal or economic relevance, such as products, politicians and institutions, medical procedures (e.g., vaccination), or other persons and social groups (as in stereotypes and prejudice). The literature dealing with the long history of attitude research is replete with methods and theoretical constructs that illuminate attitudes from ever new angles. If one searches long enough, any theoretical metaphor can likely be found in this literature. This article, however, is limited to the most influential and productive approaches.

Learning Theories

Theoretical accounts of attitude acquisition often start with the notion of classical conditioning (Pavlov, 1927). When pairing a formerly neutral conditional stimulus (CS) with an affectively significant, unconditional stimulus (US) that elicits an unconditioned response (UR), the originally neutral CS will over time produce a conditioned response (CR) that resembles UR, even CS is presented alone and not accompanied by US. Such a stimulus- pairing process can result in a stable tendency of CS to solicit a similar affective and behavioral response as the original US, which corresponds to the behaviorist definition of an attitude. This principle is refined in theories of evaluative conditioning (Staats & Staats, 1958; for a review see Hofmann et al., 2010). It proposes that a positive evaluation of a stimulus can be directly transferred to a neutral stimulus through paired presentation. In comparison to classical conditioning, it recognizes attitudes as evaluations and not behavioral tendencies and assigns greater flexibility to the relationship between stimulus and resulting response through cognition.

In operant conditioning (Skinner, 1938), reinforcement is contingent on the individual's behavior. Following Thorndike's (1898) law of effect, operant conditioning predicts that reinforced behavior will be repeated, while punished behavior is extinguished. If reward or punishment is reliably contingent on operant behavior over time, individuals acquire a corresponding tendency to exhibit the attitude.

A genuinely social perspective on conditioning originated in Bandura's (1977) theory of social learning. It proposes that individuals can acquire behaviors (and attitudes) in a truly social fashion, by instruction or by observing the contingency of reward and punishment on operant behavior in other people. The chief theoretical innovation here lies in the notion of vicarious reinforcement. Learning is not bound to first-hand experience of a significant consequence; it extends to second-hand observation of a social model's learning process. Thus, if a child observes that a peer steals a cookie, followed by punishment, social learning theory predicts that the child will be less likely to steal a cookie in the future. In contrast, the child is more likely to imitate behaviors that lead to satisfaction and enjoyment rather than punishment.

The purpose of conditioning theories is to develop distinct predictions of how to change attitudes and behavior and how to implement manifest changes, but the application of learning theories to attitude change research is not limited to conditioning. Another learning theory that is commonly known as the Yale model of persuasion (Hovland et al., 1953) aimed to predict and explain attitude change as a function of attributes of the communicator (e.g., expertise), the message (e.g., strong arguments), the channel, and the audience. It needs to be noted that this work, as well as later contributions it inspired, did not rely on explicit assumptions about learning mechanisms but simply assumed, at the functional level, that presenting attitude objects along with suitable stimulus arguments in an optimal time frame will foster persuasion.

Over decades of empirical research, it became obvious that learning theories do not provide a complete explanation of some of the phenomena observed in experiments on attitude change, such as the unreliable effects of reinforcement magnitude (e.g., Bonem & Crossman, 1988) or repeated exposure (e.g., Pechmann & Stewart, 1988). Nevertheless, learning theories enjoy unbroken popularity. Learning is so fundamental to one's concept of the acquisition and maintenance of attitudes that its contribution is taken for granted. Yet, the following theoretical ideas that originate in other metatheoretical metaphors complement and qualify the principles of learning in contemporary social psychology of attitudes and attitude change.

Psychodynamic Theories: Dissonance and Beyond

Comparable to learning theory both in scientific productivity and in popularity among practitioners, theories of attitudes and attitude change rely on a blend of the psychodynamic notion of equilibrium and the Gestalt notion of cognitive consistency. Consistency can refer to the interrelations between different attitudes people hold (as in balance theory by Heider, 1946), between one's own attitudes and others' opinions (e.g., in Lewin's 1951 field theory, and in Asch's 1951 majority conformity), and between attitudes and behaviors.

In what would become one of the most prominent and widely used ideas in social psychology, Festinger (1957) proposed the theory of cognitive dissonance. It poses that when individuals encounter a discrepancy between two or more mutually incompatible cognitions (such as an attitude and an attitude-inconsistent behavior) they experience a state of arousal and discomfort, eliciting a drive to reduce dissonance. There are many different methods of re-establishing a comfortable equilibrium, of which the most efficient method is most likely chosen.

For instance, if someone dislikes rap music, but finds himself or herself recommending rap songs to friends, the resulting dissonance might be reduced by inventing justifications for the attitude-discrepant behavior, downplaying the importance of rap music, repressing (forgetting) the discrepancy, or adding consonant cognitions (about many other attitude-consistent behaviors). However, if the discrepancy is too salient to be ignored or repressed and no justifying arguments suggest themselves, the individual must change his or her attitude as a final resort, telling himself or herself that rap music is actually not that bad.

Thus, one distinct implication of dissonance theory that has inspired hundreds of experiments in the forced-compliance paradigm is that engaging people in attitude-discrepant behavior affords an effective means of attitude change, especially when attitude-discrepant behavior is not enforced and hard to justify and when no other methods of dissonance reduction are available.

Indeed, when the level of reward or sanctions for desired or undesired behavior are manipulated in another prominent paradigm, a typical finding says that the greatest change in attitudes may occur not with the highest amount of reward. If a parent treats children to an ice-cream every time they clean up their room, the principle of over-justification predicts that intrinsic motivation will be undermined. An extra reward for an intrinsically motivated behavior that a child exhibits spontaneously, without an external incentive, may turn a positive motive ("I cleaned up, so I must like cleaning or at least a clean room") into an extrinsically motivated instrumental act ("I cleaned up to get ice-cream").

In the remarkable monograph, Deterrents and Reinforcement, Lawrence and Festinger (1962) applied the same principle of under-justification to animal learning, showing that dissonance theory can explain a variety of behaviorist learning principles (such as the partial-reinforcement effect or the wisdom that domesticated animals must be treated strictly and not rewarded abundantly). In a noteworthy application of dissonance theory in economics, motivation-crowding theory (Frey & Jegen, 2001) contends that token systems in organizational settings must avoid over-justification effects. To optimize co-workers' sustainable performance, it is essential not to provide extra incentives for intrinsically motivated achievements.

Several other "Festingerian" theories constitute special variants of the basic dissonance-theory idea. For example, social comparison theory (Festinger, 1954) presupposes that uncertainty about one's own social rank or about one's emotional reactions is unpleasant, similar to a state of dissonance, so that an explanatory need will motivate people to seek opportunities to compare themselves with other people. Specifically, individuals will prefer comparison partners that are close to their own attitudes (as comparison becomes less informative with increasing discrepancies), and will act to reduce any remaining discrepancy in attitudes by either changing their own attitude or attempting to change others' attitudes.

Still another Festingerian theory advanced by Brehm (1966) is reactance theory. It proposes that individuals experience a state of arousal and discomfort when they experience a restriction of their freedom. For instance, when one takes a toy away from a child, reactance will render the toy more attractive to the child than when it was freely available. The child will try to regain access to the toy and regret its loss even if he or she rarely used to play with the toy previously. The notion of reactance is widely received in the scientific and applied literature (see Miron & Brehm, 2006); it has become common knowledge beyond psychological science.

This sketch of dissonance theory and its derivatives reflects a metatheory that is fundamentally different from the metatheory of conditioning. Rather than imposing strong constraints on predictable behaviors in highly structured situations, the regulation of cognitive consistency and of a psychodynamic equilibrium leaves plenty of room for flexibility and degrees of freedom for regulation under weak constraints. Nevertheless, the literature demonstrates that it can induce similar attitudes as conditioning, with testable predictions of immense practical value.

Cognitive Process Theories

When psychology turned its back on behaviorism and embraced the "cognitive revolution" in the 1970s, the concept of attitudes changed from "stable behavioral tendency" to an evaluative judgment of an object represented in memory. This also induced a shift of research goals from the acquisition of attitudes and their relationship to deeper understanding of the underlying cognitive architecture. Nevertheless, the roots of these cognitive approaches can already be found in the earlier work on learning theory: Bandura's (1977) social learning theory already drew heavily on cognitive-process accounts. Likewise, Hovland et al. (1953) proposed four cognitive functions that mediate the influence of communicated messages: attention, comprehension, inference, and retention. Taking the processing of objective stimulus properties by the audience even further, proponents of cognitive response theory suggested that the persuasive impact of a message is primarily determined by the thoughts the audience generates in response to it, not the message itself (Greenwald, 1968).

Many cognitive process theories constitute dual-process approaches, which had a particularly strong impact on social-cognitive theorizing in the 1990s. Whether they divide processes into heuristic and systematic (Chaiken, 1980), impulsive and reflective (Strack & Deutsch, 2004), or intuition and reasoning (Kahneman, 2003), the common denominator of dual-process theories is that, in a given moment, individuals process information in one of two fundamentally different modes, and that there is a trade-off between frugality and accuracy in processing. The more systematic processing becomes, the better results it is supposed to yield, but also the more time and effort it takes.

In attitude-change research, this notion has inspired one of the seminal theories taught in any introductory social psychology course: the elaboration likelihood model (ELM; Petty & Cacioppo, 1986). It suggests that if involvement and ability to process persuasive communication are low, processing will take the peripheral route (which is supposed to be low-effort and heuristics-driven) and lead to relatively temporary, unreliable attitude change. If motivation and resources are sufficient, however, information is likely processed in the central route (effortful and systematic), which leads to enduring and reliable attitude change. For instance, an advertisement can entail cues that trigger influence along the peripheral route (such as attractive pictures and music) or elaborate on contents that appeal to central processing (such as advantages of the brand over its competitors).

The advent of dual-process approaches was linked to new forms of attitude measurement. Unlike psychodynamic theories that construe attitudes as result of a deliberate mental reasoning process, dual-process theories locate much of the processes connected to attitude acquisition and change and their influence on behavior outside of awareness (unless the preconditions for systematic processing are met). They do, however, assume that attitudes are reflected in behavioral measures such as response latencies (reminiscent of the stable tendencies suggested by learning theorists in the behaviorist tradition) that can be used for implicit measurement of attitudes, like the well-known implicit association test (IAT, Greenwald et al., 1998).

This shift toward new methods and measures offered a convenient solution to some of the most pertinent problems of attitude research: social desirability and demand effects. Some socially relevant attitudes (like sexist or racist prejudice or preferences for political candidates unpopular in the academic environment) may be held, but not reported honestly for fear of social repercussions. Implicit measurement tools promise to afford a remedy to self-presentation strategies and self-censure. However, subsequent research showed that implicit measures are also subject to voluntary control of expressed attitudes (e.g., Fiedler & Bluemke, 2005). Similarly, the assumption of two exclusive processing routes has been continuously challenged (Melnikoff &

Bargh, 2018). There are good reasons to assume that either more than two processing modes may be at work (e.g., quad model, Sherman, 2006), or the assumption of a single mode may be enough (e.g., uni-model, Kruglanski & Thompson, 1999).

Interpersonal and Constructivist Theories

Virtually all theories presented so far assume that to acquire and change attitudes some input from the environment is necessary. An adaptive-cognition approach to attitude research and evaluative learning illuminates how attitudes are shaped by the structure of the environment and the individual's experiences in this environment as well as the usefulness of attitudes to navigate the natural and social world. Brunswik's (1952) lens model highlights that distal properties of attitude objects (e.g., another person's sincerity or honesty) are not amenable to immediate perception but can only be inferred from proximal cues (e.g., verbal, nonverbal, and para-verbal symptoms of sincerity and deception; see Ambady & Rosenthal, 1992; Hartwig & Bond, 2014). This Brunswikian theory program, called probabilistic functionalism, compliments traditional intra-psychic approaches to attitude acquisition and change, as it puts both the multivariate distribution of environmental cues and the individual's multi-cue learning process into focus.

From a similar vantage point, several sampling theories (Fiedler & Juslin, 2006; Fiedler & Kutzner, 2015) highlight that attitudes and social cognition are predetermined by the structure and distribution of environmental stimulus information that impinges on the individual's mind, and by the individuals' strategies of information search (Klayman & Ha, 1987). For instance, the acquisition and maintenance of objectively unfounded attitudes can reflect the skewness and the unequal availability and density of information in the environment (Hogarth & Soyer, 2011; Kutzner & Fiedler, 2017; Parducci, 1968; Unkelbach et al., 2008).

In the early research history, attitudes were conceived as inherently super-individual constructs, as evident from Weber's (1947) linking attitudes to societal institutions, the social attitude concept propagated by Allport (1924), or symbolic interactionism (Mead, 1934). This genuinely social, environmental perspective, which transcended the individual and which was also inherent in psychodynamic and behaviorist learning theories, is largely disregarded in many contemporary theories of attitudes and social cognition. Likewise, Moscovici's (1976) highly influential social representation theory proposes that attitudes are shared among groups and communities to enable social groups to function through communal action and communication. In turn, communal action and communication produce similarity of attitudes within groups, while substantial discrepancies may arise between different groups, communities, and cultures.

Other approaches share the assumption that the formation and change of attitudes often takes place through communication and social interaction, reflecting more than mere information transmission. Bandura (1977) highlighted the role of instruction in learning, and argued that any widely shared cultural representation necessitates dissemination of socially reproducible information (Kashima, 2000; Schaller et al., 2002) across individuals, social groups, and generations. Social transmission accounts (e.g., Kashima, 2000) and iterated learning theory (Kalish et al., 2007) use the intriguing research paradigm of serial reproduction (Bartlett, 1932) to investigate how communication shapes and maintains attitudes. The most compelling prediction of this approach is that what may appear as a very small effect in one individual may be amplified when information is communicated repeatedly and travels through the social network. In this way, tiny biases in what is understood, remembered, and chosen for retelling may exert a strong and persistent influence on socially shared attitudes and stereotypes.

Communication follows a set of rules—communicators engage in a communication game (Higgins, 1981)—that shapes the environment in which attitudes form and change. This view implies that by communicating with others and tuning messages to meet the knowledge of their communication partners, speakers not only manage to change others' attitudes. They also tune and update their own attitudes. Under the influence of repeated perspective-taking, switching between speaker and listener roles, participants of the "communication game" form joint representations of reality according to shared reality theory (Hardin & Higgins, 1996).

This fascinating coordination process between individual communication partners, their environment, and the semiotic medium through which they interact provides the core of a comprehensive metatheory representing the core of Grice's (1975) principle of cooperative communication.

Conclusion and Future Directions

This brief and necessarily incomplete summary gives an impression of the richness and diversity of social psychology theorizing about attitudes and attitude change. What is striking, however, is that some metatheoretical ideas have sparked intense productivity in research and practice. Other principles and metaphors are conspicuously missing. For instance, biological and neurocognitive metaphors were rarely used, and if so, mostly to illuminate the physiological basis of influential theories such as learning, dissonance theory, or dual-process models. While some quickly abandoned theoretical paths may just not be suited to the study of attitudes, other metatheories may prosper in the future, once a new generation of methodological instruments have been established (e.g., network theory, computational linguistics, multimedia tools). In any case, it seems safe to say that despite the achievements of the past, attitudes and their conceptual siblings will remain a central area of social psychological theorizing in the future.

Motivation Regulation

Following the focus on attitudes in the prime age of dissonance theories in the 1960s (Brehm & Cohen, 1962; Festinger, 1962; Lawrence & Festinger, 1962), the attribution theories in the 1970s (Kassin, 1979; Kelley, 1973), the heuristics-and-biases research program in the 1980s, and the dual-process approaches in the 1990s, a new class of theories are devoted to motivation regulation. The term motivation regulation highlights the insight gained from several decades of industrial empirical research that distinct stimuli, situations, person attributes, group membership, and cultural norms hardly ever impose strong deterministic constraints on social inferences, judgments, and decisions. They rather allow for a considerable flexibility and malleability, enabling individuals to construe the same constellation in manifold ways. The same verbal utterance can be construed as a joke, an insulting speech act, a clumsy dating attempt, or a symptom of linguistic incompetence.

Consumers, managers, or basketball players can rely on risk-avoiding or on risk-abiding strategies; communications and instructions can be abstract or concrete; and actions can be careful and cautious based on large amounts of information or quick and intuitive, relying on spontaneous impulses and gut reactions. These and many similar settings are characterized by constructive flexibility and more uncertainty and indeterminacy than anticipated in former theorizing on attitude formation and change.

Theories of motivation regulation offer reasonable frameworks and research paradigms to study how individuals try to reduce uncertainty and to develop adaptive behavioral strategies under uncertainty. Depending on their personal goals and preferences and on the motivational structure of the problem situation, individual strategies vary on such dimensions as liberal versus conservative, abstract versus concrete, or spontaneous versus careful. The present section will focus on three prominent and comprehensive theories of motivation regulation, which have exerted a huge impact on social psychology since the late 1990s: regulatory-focus theory (RFT), construal-level theory (CLT), and need-for-closure. All three approaches focus on motivated strategies that trigger systematically different ways of regulating individuals' adaptive struggle with a complex and uncertain environment.

Regulatory-Focus Theory

Crucial to the RFT propagated by Higgins (1998, 2002) and by Pham and Avnet (2004) is the distinction between promotion focus and prevention focus, conceived as two opposite (or complementary) regulatory orientations. While a promotion focus emphasizes the motive to win (or else to survive, to succeed, to dominate, to be healthy and happy), a prevention focus entails the motive not to lose (or else not to die, not to fail, not to give up, to not be unhealthy or unhappy). Most task settings allow for both framing perspectives, making it possible to conceive the same task from a promotion or prevention focus. And yet, the resulting emotional experience and adaptive strategies are systematically different for individuals guided by a promotion or a prevention focus.

While the basic motivation underlying a promotion focus is concerned with advancement, growth, and accomplishment, a prevention focus emphasizes security, safety, and responsibility concerns (Crowe & Higgins, 1997). RFT assumes that fundamentally different regulation and information-processing strategies administrate these opposite regulatory orientations: “For example, to keep a slim figure, a promotion-focused individual would be more likely to exercise (approach a match), whereas a prevention-focused individual would be more likely to avoid eating fatty foods (avoiding a mismatch)” (Florack et al., 2010, p. 193).

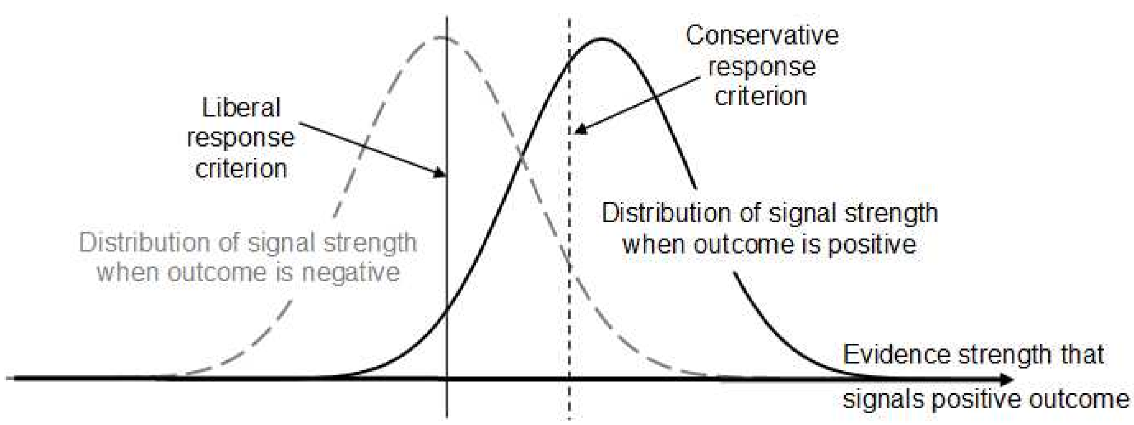


Figure 2.2./1. Using signal-detection analysis to pinpoint testable implications of RFT.

Because promotion-focused regulation means finding ways to approach desired states whereas prevention-focused regulation means avoiding mismatches with desired states, promotion-focused people are eager not to miss positive outcomes, whereas prevention-focused people are vigilant to avoid negative outcomes. This assumption imposes testable constraints on decision strategies that can be visualized through signal-detection analysis (see fig. 1). A promotion focus calls for liberal response biases that serve to maximize "hits" and to minimize "false negatives" (i.e., undetected positive outcomes), whereas a prevention focus gives rise to a conservative bias, trying to avoid "false positives" and to enhance "correct rejections."

Assume that a signal is stronger when an expected outcome is positive rather than negative. For example, a stronger signal discriminates between consumer products of high quality (solid distribution curve) versus low quality (dashed curve). Promotion-focused consumers, who buy all products with a signal strength higher than the liberal criterion on the left will achieve a very high hit rate, as over 90% of the area under the solid curve is right of the solid criterion line. They do not care much about the sizeable rate of false positives (almost 50% of the area under the dashed curve exceeding the liberal criterion). In contrast, the conservative criterion of prevention-focused consumers (dashed vertical line on the right) produces an opposite bias, characterized by high correct rejection rates (more than 80% of the area under the dashed curve is left of the dashed criterion). Such a conservative bias increases willingness to accept higher false negative rates (almost 40% of the area under the solid curve left of dashed criterion).

Specifying the distinction of promotion versus prevention focus, respectively, in terms of a liberal versus conservative response bias in signal detection analysis offers a parsimonious and elegant account of a plethora of experimentally tested hypotheses inspired by RFT. Thus, the assumption of a more liberal, less cautious response bias can explain that promotion-focused compared to prevention-focused people give higher weight to speed than to accuracy (Förster et al., 2003) and exhibit more risky and less careful responses informed by internal feelings and mood states (Pham & Avnet, 2004). Moreover, a promotion focus fosters creativity (Friedman & Förster, 2001), whereas a prevention focus encourages more systematic and extensive information processing (Pham & Avnet, 2004). In a series of experiments, Florack et al. (2010) showed that promotion-focused consumers were more likely to follow their implicit preferences for hamburgers and other food items, as measured by the IAT (Greenwald et al., 1998).

In these and in many similar experiments, regulatory focus functions as the chief independent variable, either manipulated experimentally (e.g., by letting participants write about their hopes and aspirations or about the duties and obligations) or measured as a personality trait, using scales developed by Lockwood et al. (2002) or by Higgins et al. (2001).

Later extensions of the theory are concerned with the notion of regulatory fit (Higgins, 2002), which emphasizes the feeling of rightness or fit that people experience when they reach a successful outcome in their own preferred regulatory orientation (i.e., when promotion-focused people succeed through eagerness or when prevention-focused people succeed through vigilance). It is important to note that the enhancement gained from regulatory fit refers to subjective experience rather than objective performance.

Construal-Level Theory

A similarly prominent motivation regulation approach with numerous implications for social judgment, inference making, and adaptive behavior in multiple domains is CLT, advocated by Trope and Liberman (2003, 2010). CLT can be conceived as a theory of psychological distance. In a nutshell, it assumes that from a high temporal, spatial, social, or factual distance, social targets and decision objects are construed in more abstract, holistic, and idealized value-related ways. In contrast, short-distance construals are concrete, detailed, and feasibility-related. For instance, when one plans a holiday journey, a new research project, or a dating episode next year, in the remote future, one is concerned with idealized goals and their value, detached from many details and concrete constraints, which only become visible when the same behavioral objectives are visible from a shorter distance. Now the holiday plans are no longer exclusively determined by an idealized hierarchy of favorite countries or destinations, but also depend on such mundane feasibility constraints as the availability of flights, the money in one's bank account, or the weather in one's favorite destination. Likewise, as distance shrinks, a seemingly fascinating research project has to be revised in the light of reviewers' feedback to one's grant proposal. The dating outcome turns out to depend not only on one's own ideals but even more so on the envisaged partner's preferences, the role of dating rivals, and such mundane constraints as geographical distance and vocational compatibility.

Although CLT appears to rely on concepts clearly different from RFT, it also assumes that social behavior can be flexibly construed and regulated and it has similar implications about the dimensions of adaptive regulation. From a distance, at high levels of construal, the dominant regulatory strategies (curiosity, risk-seeking, desirability-driven, creativity, positive mood) resemble the typical strategies in promotion-focus states, whereas prevention-focus strategies (carefulness, risk avoidance, feasibility concerns, security, negative mood) are typical of low-distance settings construed in concrete detail. More specifically, the construal of distal and proximal goals and behavioral options varies both in terms of what aspects one attends to and in terms of how one mentally represents these aspects. Regarding the former aspect, it is possible to concentrate from a distance on the idealized desirability and the evaluation of one's behavioral objectives. As one gets closer, more weight is given to unexpected and uncontrollable feasibility constraints that restrict the realization of these idealized motives (Trope & Liberman, 2003, 2010). Consequently, distal construal focuses on the inherent value of objects, on individuals' internal traits and dispositions, and on the utility component of decision options, whereas proximal construal has to take unwanted side aspects, situational forces, and the probability components of decision options into account (Nussbaum et al., 2003; Sagristano et al., 2002).

Regarding the latter aspect of how we encode and represent information, the metaphor of a zoom objective is helpful. Just as it is possible to zoom in or zoom out a picture of the same environment and thereby decrease or increase the experienced distance, respectively, the same goal or behavioral objective can be construed in all its (zoomed-in) detail, at a high level of resolution. Conversely, zooming out produces a much lower level of resolution, yielding a less detailed representation but a richer context of the surrounding world. In line with this metaphor, distant construal uses abstract language and large measurement units, referring to global characteristics, whereas proximal construal uses concrete language and small measurement units referring to local characteristics (Krüger et al., 2014; Trope & Liberman, 2010).

Note that these implications and corollaries of CLT follow naturally from the a-priori notion of distance and the zooming metaphor. They do not create a burden of a non-parsimonious theory that suffers from too many arbitrary assumptions. Just as a world map must present more inclusive and less detailed representations of the same geographical objects than a local city map, a distant psychological perspective naturally calls for higher abstractness and requires larger units of analysis than a proximal perspective. Likewise, distal construal consists of idealized representations of expected goals, abstracted from many details and feasibility constraints that become only visible as one gets closer to the goal and its many unpredictable concomitants.

What may be less predictable on a priori grounds is that seemingly independent aspects of temporal, spatial, social, and evidential distance all converge to one unitary dimension of psychological distance. Thus, regardless of whether experiments manipulate temporal distance (e.g., judgments or decisions taking place right now or in the far-away future, or past), spatial distance from the participant's current location, social distance (e.g., oneself vs. others), or evidential distance (e.g., judgments or decisions about probable vs. improbable events), the influences of psychological distance are very similar (Fiedler et al., 2015; Trope & Liberman, 2010). Upon closer reflection this is plausible, because what happened long ago is likely to involve other social partners and to have taken place in another place than what happens right now, in one's current social and ecological setting.

Figure 2 provides a synopsis of some of the most prominent CLT findings. High-distance settings, relative to low-distance settings, give rise to more simplified, low-dimensional cognitive representations (Liberman et al., 2002), more pronounced fundamental attribution errors (Nussbaum et al., 2003), higher levels of self-control (Fujita & Carnevale, 2012), and lesser risk and uncertainty aversion (Lermer et al., 2015). Note that the behavioral implications of high-distance settings largely overlap with the implications of promotion focus whereas the feasibility focus under low distance resembles the prevention-focus orientation. Although both theories use different component concepts, they converge in many implications and empirical findings concerning major dimensions of motivation regulation.

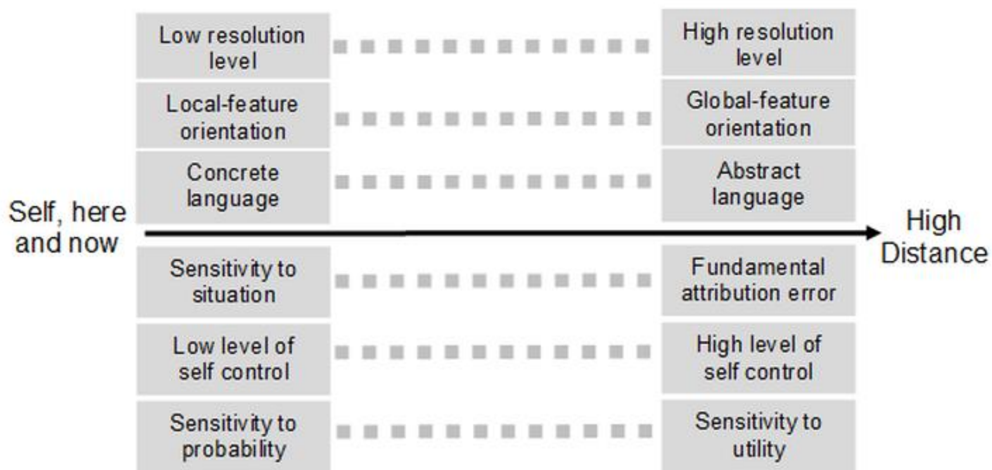


Figure 2.2./2. Psychological manifestations of high versus low egocentric distance according to contrual-level theory.

Need for Closure

According to the lay-epistemological approach propagated by Kruglanski (1990, 2004; Kruglanski & Freund, 1983; Webster & Kruglanski, 1994), the strength of a need for cognitive closure (NFC) varies between individuals and between situations. Depending on the perceived costs of uncertain or undecided states and the perceived benefits of reaching "an answer on any given topic, or any answer" (Kruglanski, 1990, p. 337), need for closure is supposed to increase in taxing and personally involving situations and when time pressure calls for resolution of an uncertain or ill-understood situation. Under these conditions, and when closure is not charged with costs through evaluation apprehension, individuals strive for quick and premature answers to open questions and to make judgments and decisions based on insufficient information. For example, Kruglanski and Freund (1983) found that elevated NFC through time pressure fostered primacy effects in impression formation and facilitated stereotypical judgments and anchoring effects on quantitative estimations.

Reaching closure can be satisfactory not only because uncertainty is avoided but also because epistemic efforts of hypothesis generation come to rest. While incomplete processes keep mental processes busy and cause Zeigarnik effects (Atkinson, 1953), closure means that epistemic activities are “frozen” (Lewin, 1947), freeing the individual from the need to process relevant stimulus information. Conversely, the epistemic system is “unfrozen” when a discrepancy between experienced and desired states interferes with closure.

Motivation regulation is primarily concerned with an adaptive balance of exploration and exploitation (Mehlhorn et al., 2015). Elevated closure serves to decrease patience and exploration time and fosters premature exploitation. Whether such a strategy is functional and effective depends on the payoff structure of the situation and on the social context.

Like RFT and CLT, the lay-epistemological NFC perspective also highlights the malleability of problem representations. Although the chief dimension of regulation appears to be fundamentally different from psychological distance (CLT) and regulatory-focus orientation (RFT), there is a good deal of overlap in predictions and behavioral downstream consequences. The quick and insufficiently reflected strategies triggered by high (vs. low) NFC resemble in many ways the liberal response strategies under promotion (vs. prevention) focus or the idealized, simplifying worldviews fostered by high (vs. low) psychological distance. Thus, it appears as if different theories of motivation regulation, however independent they appear at surface level, converge on a similar spectrum of (simplifying vs. refined) behavioral styles and strategies as the dual-process models that grew out of the attitude research program.

Social Groups

In a third major domain of social psychological research, theories focus on behavior within and between social groups. It is not too surprising to see that group-level theorizing often refers to the same typical social psychological principles that we have already encountered in the attitude domain or in motivation regulation approaches. Despite the radical shift from individual to group behavior and from attitudes and motives to such concepts as group norms and intergroup conflict and discrimination, theorizing continues to draw on such principles as consistency, categorization, representation of social knowledge in (collective) memory, uncertainty coping through social comparisons, and the regulation of motives and goals.

There is hardly any genuinely group-specific theoretical module, or axiom, that cannot also be found in theories of individual behavior. Other theory modules that might ideally apply to groups but not to individual behavior—such as linguistic or semiotic rule systems (Nakassis, 2005), ethological approaches to group and herd behavior (like incest barriers; Bischof, 1972), formal foraging models (Hills et al., 2015), or agent-based simulation of the Schelling type (Zhang, 2004)—are conspicuously rare, if not completely neglected. In any case, the convergence in theorizing across seemingly different domains and units of analysis testifies to the existence of a compact repertoire of preferred theoretical tools, which dominate social psychological journals, research designs, and curricula.

This section begins with the developmental history of the most prominent group-level theory, social identity theory (SIT) (Tajfel & Turner, 1979), as it is rooted in basic individual-level processes of categorization (Bruner, 1957; Tajfel, 1969), social comparison (Festinger, 1954), and motivated regulation processes (Hogg, 2007). The remainder of this section is devoted to more subsequent theories that are either derived from, or related to, social identity.

Social Identity Theory

In a nutshell, SIT (Tajfel, 1978; Tajfel & Turner, 1979) can be conceived as a compound of two individual-level theories, namely Tajfel's (1957) theory of accentuation and a group-level analogue of Festinger's (1954) social comparison theory. Accentuation is a universal principle that is reminiscent of the principle of redundancy in information theory (see Tajfel, 1957). It says that when a primary task calls for the discrimination of stimulus objects in one attribute dimension X, discrimination is facilitated when the same objects also vary in another, correlated (i.e., redundant) attribute Y. For instance, differences in line length (Tajfel & Wilkes, 1963) or in the size of coins (Bruner & Postman, 1948) are accentuated, if long versus short lines also vary in an irrelevant attached letter and if large versus small coins also vary in value. This facilitates discrimination regarding the relevant attribute.

Tajfel (1969) and others (Eiser, 1971; Eiser & Stroebe, 1972) soon recognized that the principle of accentuation is applicable to all kinds of social categories and groups. Thus, when statements or newspaper headlines not only vary in the preferred attitude (e.g., liberal vs. restrictive statements about education) but opposite attitude statements are allegedly taken from two (fictitious) newspapers (The Messenger vs. The Gazette), the perceived difference in attitudes is accentuated (Eiser, 1971). By analogy, evaluative judgments of moral behavior, likeability, or performance exhibit accentuation effects when variation in task-relevant attributes is correlated (redundant) with group membership. For instance, it is easier to discriminate students' performance in science if high and low performance is observed in students who have a different gender or socioeconomic background. The resulting accentuation of between-group differences and within-group similarities is not confined to task settings in which groups actually correlate with behavior but also extend to situations in which correlations are merely expected stereotypically. Thus, regardless of whether girls actually perform lower than boys in science, it is sufficient that teachers expect or believe gender to correlate with performance.

This universal principle of categorization and discrimination (in a literal sense of the word) is at the heart of SIT's assumption that individuals' feelings of social identity (i.e., their self-conception and personal feeling of distinctiveness) depend crucially on the groups to which they belong, as distinguished from the rival groups to which they do not belong. Thus, the kernel of social identity is a generalized, multivariate accentuation effect, which highlights the distinctive features of people's ingroups and their outgroups. The crucial role played by SIT's basic categorization component received particular support from the memorable evidence on minimal groups (Billig, 1973; Tajfel et al., 1971), showing that even artificial ad-hoc groups without a common history and future (like the random allocation of people to the "groups" of overestimators vs. underestimators) are sufficient to evoke discrimination and ingroup favoritism.

While the accentuation component of SIT has all typical features of a good Gestalt (grouping; inclusion/exclusion) or consistency theory, the second component assumption that social identity triggers social-comparison processes at the group level, is clearly of the psychodynamic kind that characterizes Festingerian theories (Festinger, 1954). Accordingly, people have a fundamental interest in understanding who they are, and how they compare relative to other people. Because no objective metric is available to measure their traits and capacities (e.g., how honest, how physically strong, how humorous they are) and even ordinal scales are hard to construct via physical comparisons (e.g., lifting or "mastering" stones of different size), people resort to social comparisons, preferably with similar others, as a major source of identity information. SIT assumes that social comparisons at group level (e.g., involving one's favorite soccer team, political party, or gender group) are crucial for the negotiation of one's relative social rank and identity.

Similar to dissonance theory in attitude research or to motivation regulation theories, SIT is not bound to a single silver-bullet road to attaining a positive social identity. Much like attitude profiles or motivation orientations, the theory provides individuals with many degrees of freedom in how to construe a positive social identity. Strategies can vary between selective engagement in upward versus downward social comparisons, selective use of abstract versus concrete words in describing desirable and undesirable ingroup and outgroup behaviors (Maass, 1999), ascription of primary (infrahuman) and secondary (human) emotions (Demoulin et al., 2009), or selective memory and biased representation of ingroup versus outgroup behaviors (Rothbart & Lewis, 1994).

Descendants of SIT

Since the appearance of SIT, a growing class of related theories extends and refines its principles (e.g., self-categorization theory; Turner et al., 1987; and optimal distinctiveness theory;

Brewer, 1991). In most of these cases, theory derivatives were formed by combining SIT with other theoretical metaphors and moderating assumptions, arriving at new compelling theories that go beyond SIT.

Integrating SIT with cognitive approaches to judgment under uncertainty and motivation, uncertainty-identity theory poses that individuals experience uncertainty about their own identity and the actions they should take as averse and uncomfortable (Hogg, 2007). In accordance with psychodynamic principles, this motivates behavior to reduce uncertainty. Crucially, belonging to a social group and activating its social identity promises to efficiently reduce uncertainty about the self, particularly when the group is highly distinctive, homogenous, and unambiguous in their behavioral expectations (as, e.g., typical of extremist political or religious groups).

While the theories discussed so far are mainly concerned with when and why individuals identify with social groups, they also have implications about the attitudes held toward ingroup and outgroup (often referred to as stereotypes or prejudice). In a combination of social identity with emotion theory, intergroup emotions theory predicts shifts in group-related attitudes across contexts (Smith, 1993). It proposes that events are appraised in the light of the currently activated social identity, and that this influences cognitive appraisals, action tendencies (e.g., discrimination), or group-based emotions as three interrelated aspects of group relations.

A substantial reformulation of the SIT, the ingroup projection model puts prejudice between groups on a firmly interpersonal footing (Mummendey & Wenzel, 1999). It suggests that people may find a common superordinate group (e.g., students) to integrate the representation of ingroup and outgroup (e.g., psychology and law students), provided a superordinate group label is available and salient. Crucially, the theory poses that individuals project ingroup characteristics on the superordinate group; the ingroup thus appears representative and prototypical of the broader category. In contrast, any outgroup that differs from the ingroup is likely to appear as a deviation from a “good” member of the subordinate group. If the evaluative direction of this perceived deviation is negative, it motivates prejudice, outgroup derogation, and discrimination.

What is notable about ingroup-emotions theory and the ingroup-projection model is the subjective and constructive quality of the suggested comparisons. With the same flexibility as social comparison theory (Festinger, 1957) and RFT (Higgins, 2002) outlined in the sections on attitudes and motivation regulation, individuals may shift between several available social identities and highlight different characteristics, finding either similarities or differences between and within groups.

Interpersonal and Constructivist Theories

This constructive quality of social groups within the individual is complemented by the interpersonal level of analysis. The role of the social environment, and particularly communication, in forming and maintaining social groups and intergroup relations is obvious. According to communication accommodation theory membership in social groups is reflected in and negotiated through differences in how communication is used (Giles et al., 1991), ranging from small distinguishing idiosyncrasies (e.g., a special greeting shared among a group of friends) to differences in the way language and speech are employed between nationalities and ethnic groups. Individuals are thought to seek out contact with others who have a similar communication style, and adjust their communicative behavior to enhance how they are perceived by the communication partner (convergence) or to render the self and the ingroup distinct from them (divergence). Similarities in communication style are supposed to facilitate the formation of social groups but, vice versa, social groups also cause similarity of communication style among their members, increasing closeness within ingroups and social distance to outgroups.

Not only the communication style but also the lexical contents moderate group dynamics. Communication serves as the means to share and uphold a common concept of the ingroup and stereotypes toward outgroups, as noted with reference to shared reality (Hardin & Higgins, 1996) and serial reproduction of stereotypes (Kashima, 2000). In addition, work on the linguistic category model highlights the key role of language for interpersonal and intergroup relations (Semin & Fiedler, 1991). The model holds that the flexibility of language allows language users to describe the same social behavior using different classes of words. Linguistic categories (i.e., verbs and adjectives) that vary in abstractness elicit systematically different inferences about the behavior they describe and about the acting individual. In an intergroup context, a typical finding is that negative behaviors are described in more abstract terms if they refer to the outgroup than to the ingroup, whereas descriptions of positive behaviors tend to be more abstract when they belong to the ingroup than to the outgroup.

This pattern, which has been termed the linguistic intergroup bias (Maass, 1999), serves to frame positive ingroup and negative outgroup behavior as typical and expected. By raising expected behaviors to an abstract language level and downplaying unexpected behaviors in concrete terms, stereotypes are upheld in the cultural sphere, regardless of whether the factual evidence is inconclusive or even incongruent.

Prejudice and hostile intergroup relations not only constitute challenges for empirical research but, first and foremost, a pressing societal problem. Over the history of research in social psychology, research has aimed to develop effective interventions to improve intergroup relations. The well-known contact hypothesis holds that through added personal experience with the outgroup under auspicious conditions (equal status, common goals, no competition, and supportive authorities and regulations), prejudice will decrease and intergroup relations will improve (Allport, 1954). This mirrors the implication of sampling theories assuming that negative attitudes may often reflect biased and impoverished experience samples.

Integrating the contact hypothesis with SIT, the common ingroup identity model suggests that prejudice can be reduced by nudging members of different social groups to develop a common social identity (Gaertner et al., 1993). For instance, when athletes compete in mixed-nation teams, their identity as fellow team members reduces prejudice against foreign nationalities.

Concluding Remarks

As contended at the outset, it was impossible to provide a comprehensive review of all major theories that have been advanced in a psychological subdiscipline as fertile and multiplex as social psychology. Rather than listing minimal sketches of a maximal number of theories, the article focused on a selected set of most prominent theories that can be considered typical and representative and that—from a historical perspective—had a profound influence on research and theorizing in social psychology. This overview deliberately outlined and elaborated on these most prototypical theories in some detail, while several other theories were mentioned in Table 1.

As a result, several solitaire theories were neglected, including some rather prominent examples, which may have deserved more attention. For instance, readers may have missed a closer discussion of terror-management theory by Greenberg et al. (1986) or Gollwitzer's (1999) notion of implementation intentions, or Schwarz and Clore's (1983) mood-as-information approach, or perhaps Berkowitz's (1969) neo-behaviorist theory of aggression.

Dozens of other noteworthy theories might be added, which simply did not fit the present organized overview of major theories of social psychology. To compensate for the deliberate selectivity of the present article, Table 1 includes a more inclusive taxonomy of theories, organized as a "periodic table" or Cartesian product of theory principles and domains of social psychological research.

Interested readers are invited to contemplate and elaborate on the taxonomy, adding distinct theories that are missing in Table 1, generating new rows (original theory modules) and columns (novel domains), and reflecting on white spots, that is, theories that are sorely missing although the taxonomy guarantees a place for them in the future development of social psychology. Readers are particularly encouraged to contemplate on the common ground and also on the points of divergence between different theories. One possibility is that understanding the relations between theories amounts to understanding the historical origins of pioneer theorists and their disciples—the Hovlanders, the Festingerians, or the followers of Bruner and Bartlett. Another possibility is to examine the theoretical principles shared with, or borrowed from, neighboring disciplines such as psychophysics, psychoanalysis, developmental psychology, biology and ethology, and philosophy. It should be particularly enlightening to systematically search for convergent as well as divergent implications and applications of the various historical sources.

With regard to applied social psychology, this article argues for a symbiotic relationship between fundamental and applied research. On one hand, basic research and theorizing is inspired and fertilized by such practical challenges as personnel selection, group performance, group decisions in democratic systems, migration, social media, law and justice, and health behavior. On the other hand, practitioners can profit a lot and build on insights gained from basic research, as articulated in Kurt Lewin's famous statement that "there is nothing as practical as a good theory" (Lewin, 1943, p. 118). There can be no doubt that social psychology is at the heart of a great number of societal, political, ecological, and health- related problems of the new millennium. In some of these domains (e.g., migration, legal practices, aggression, sports), social psychology has been applied abundantly and with remarkable success. Other domains have been neglected (e.g., sustainable behavior, pandemic interventions, social rationality). In any case, the key to proper applied research lies in the further development of translational science, which tackles the ambitious task of translating nomological rules discovered in basic research into controlled practical interventions based on a deliberate analysis of costs and benefits in existing societal systems. The propensity to bridge the gap between basic and applied settings is perhaps the most ambitious and responsible aspect of a good and practically useful theory.

3. Hindsight Bias and Pragmatics

Among psychologists who study decision-making, cognitive illusions have been a central topic since the landmark paper by Tversky and Kahneman in 1974. One of the most well-known and intensely studied, the hindsight bias or "knew it all along" effect describes the observation that after an event occurred, we judge it as more likely and predictable than before it occurs. While the hindsight bias is a phenomenon that empirical studies show to be both robust and widespread, the mechanisms behind it are still not fully understood (Bernstein et al., 2016). Prior theoretical accounts often explain the bias in terms of social cognition: failures in memory storage, retrieval, and information processing, as well as metacognitive shortcomings. Others present it as motivated by needs to perceive the self as competent and the environment as orderly, in the tradition of psychodynamic theories (see, e.g., Pohl & Erdfelder, 2022, for an overview).

This chapter presents a radically different approach that construes hindsight bias studies as a communication process between participant and experimenter. It will first provide some background on pragmatics within cognitive illusions research and the influence of communication on the speaker, before the article "Pragmatic, constructive, and reconstructive memory influences on the hindsight bias" (Salmen, Ermark, & Fiedler, 2023) applies the presented ideas to form a new theory of a mechanism behind hindsight bias.

3.1. Cooperative Communication and Cognitive Illusions

To provide adequate answers to experimental questions, a participant needs to recover the meaning and intention of the experimenter. Let us assume that the participant will interpret the exchange of question and answer in an experiment like any other everyday conversation – then they will use not only *what is said* but also *what is implicated* (Grice, 1975) by the context and tacit rules that govern conversations in general to give appropriate answers (see, e.g., Schwarz, 1999). Describing these tacit rules is the objective of *pragmatics*. According to Grice's (1975) *cooperative principle*, contributions to a conversation should follow four maxims: quantity, manner, relation and quality. What we say to others, and what they answer in return, should be as brief as possible, comprehensible, relevant to the current context, and truthful. In listening to others, we expect their contributions to satisfy these maxims. If what they say appears to violate them, we use the utterance and its context to reconstruct what could actually be meant (*conversational implicatures)*. Dozens of pragmatic theories have developed to extend, revise or refute Grice's work. To name only one influential example, *relevance theory* (Sperber & Wilson, 1986) reduces the four maxims to one dimension of relevance that is described in terms of information theory and social cognition.

What pragmatics research in the tradition of Grice demonstrates again and again is that understanding of an utterance goes beyond what is said. Over the years, some impressive and thought-provoking attempts have been made to show the importance of conversational implicatures in cognitive illusions research (for overviews, see, e.g., Bless et al., 1993; Hilton, 1995; Schwarz, 1999). For instance, Hertwig and Gigerenzer (1999) could demonstrate that because the term "probable" can have different meanings, participants in experiments on the conjunction fallacy (Tversky & Kahneman, 1983) will use contextual, pragmatic cues to disambiguate the meaning of the instruction and construct their answers. Accordingly, when the question's meaning is clarified, they follow the mathematical interpretation of the term and provide corresponding answers. At the same time, in the original version of the task, the instructions lead to conversational implicatures that favor non-mathematical interpretations of the question, and thus, their corresponding answers violate the mathematical model the researchers presume.

Unfortunately, studies of this kind did not spark a wide-spread recognition that conversational implicatures play an important role in cognitive illusions research - and that construing cognitive illusions through the lens of communication deepens our understanding of mechanisms behind the most well-known JDM findings. As Norbert Schwarz (1999) poignantly put it in the title of his seminal paper, "the questions shape the answers." Understanding a study as communication between the researcher and the participants allows us to form many more novel theories of cognitive illusions, as the study presented in Chapter 3.3. aims to show for the hindsight bias.

A typical research paradigm on the hindsight bias involves a prompt (e.g., the description of a situation and possible outcomes), the revelation of the true answer, and then the instruction to make a judgment as if the true answer was unknown (hypothetical design) or retrieve a prior answer from before the true answer was revealed (memory design). Using the perspective of Gricean pragmatics, the following hypothesis can be formed: Even if what is said is “ignore the true answer, it is irrelevant to the question we ask you”, what is implicated by the mere presence of the true answer is that it should be used to answer the question that appears next. After all, why else would the researcher present this information? The experimenter violates the maxim of relation with presenting the true answer before a question that does not call for it, while the participant will assume the experimenter to be cooperative and thus relevant in their instructions. In general, "communicated information comes with a guarantee of relevance" (Sperber & Wilson, 1986, p. vi). How does the participant solve the dilemma of such contradictory explicit and implicit affordances? Participants in the study presented in Chapter 3.3. give answers biased towards the true answer when it appears in the instruction – but give unbiased answers to the same question on the next page, where the true answer is absent from the page, and the maxim of quantity further propels them to give a different answer from the one they just gave. That such a minimal manipulation, and not knowledge of the true answer, would determine how participants responded, was unfathomable from prior theoretical accounts. Here, a communication perspective inspired a novel research design that illustrates a new view on mechanisms behind the hindsight bias.

3.2. Communication as a Two-Way Process

Apart from experimental pragmatics, the following article also puts the focus on an under-appreciated feature of communication: it may change the mind of the listener as well as the mind of the speaker (Hörmann, 1976). This assumption is a departure from classic linear, one-way conceptualizations of communication, as underlies the Shannon-Weaver model and Gricean pragmatics. Rather, the investigation of speaker and listener who jointly construct meaning in a complex feedback loop that involves changes in both minds to form what may be termed a *shared reality* (Echterhoff et al., 2009) stands firmly in a Wittgensteinian tradition.

The hypothesis that what someone says will influence themselves has already been put to a test in the investigation of attitudes and attitude change with the *communication-game approach* (Higgins et al., 1982). In its defining audience tuning paradigm, participants read either positively or negatively valenced person descriptions. Their task is to summarize the description for another person, who they know either likes or dislikes the described person. The typical finding (e.g., Higgins & Rholes, 1978) is that the summaries were more positive when the recipient liked the target than when they disliked the target – and that after communicating their summary to the recipient, participants recalled the original description as more positive when they had written a positive summary for a recipient who liked the target, and more negative when their summary was negative, for someone who already disliked the target.

This tuning of both message content and the communicator’s memory representations towards the recipient's attitude – the saying-is-believing effect – even increased after two weeks, and persisted to affect the speaker's attitude towards the person they described. No such effects were observed when participants only read the same descriptions and did not communicate them to another person. In the communication game approach, communicative mechanisms explain how socially shared attitudes are formed and maintained and how attitude change happens through contact with others who hold specific beliefs.

The communication game approach is a major theoretical contribution to the domain of attitudes, as displayed in Chapter 2.2. Other research paradigms that investigate memory recall and the emergence of false memories provide even more evidence that communication shapes the memory representations of the speaker. For instance, in the famous study on eye-witness testimony by Loftus and Palmer (1974), all participants saw an identical video of a car crash and were asked how they estimated the speed of the car. Participants who answered the question „How fast were the cars going when they smashed into each other?" believed the cars to go faster on average than participants who received the question „How fast were the cars going when they hit each other?". One week later, participants who were asked if the cars smashed into each other were much more likely to falsely remember broken glass in the video they saw. In another study, participants falsely remembered to have seen a yield sign instead of a stop sign in accident videos after the experimenter asked questions that presupposed a yield sign (Loftus et al., 1978). This misinformation effect suggests that memory is not only reconstructed for communication (Bartlett, 1932) rather than recalled like data from computer storage, but conversational exchanges also re-construes memory representations time and again.

The communication game approach and research on constructive and reconstructive memory demonstrate how saying something shapes the memory representations of the speaker. Meanwhile, there have been no similar demonstrations of the two-way effect of communication for judgment and decision-making. As we show in the following study, cooperatively given answers go beyond "mere artifacts" of research – they lastingly influence memory representations of judgments. This also holds true for the condition in which participants give answers that are unaffected by the (usually almost unavoidable) hindsight bias – making the following study an example of how a communication approach can not only uncover new mechanisms behind cognitive illusions, but also guide debiasing attempts.

3.3. Pragmatic, constructive, and reconstructive memory influences on the hindsight bias (Salmen, Ermark, & Fiedler, 2023)[2]

In hindsight, when the outcome of an uncertain scenario is already known, we typically feel that this outcome was always likely; hindsight judgments of outcome probabilities exceed foresight judgments of the same probabilities without outcome knowledge. We extend prior accounts of hindsight bias with the influence of pragmatic communication inherent in the task and the consolidation of self-generated responses across time. In a novel 3 × 2 within-participants design, with three sequential judgments of outcome probabilities in two scenarios, we replicated the within-participants hindsight bias observed in the classic memory design and the between-participants hindsight bias in a hypothetical design simultaneously. Moreover, we reversed the classic memory design and showed that subjective probabilities also decreased when participants encountered foresight instructions after hindsight instructions, demonstrating that previously induced outcome knowledge did not prevent unbiased judgments. The constructive impact of self-generated and communicated judgments ('saying is believing') was apparent after a two-weeks consolidation period: not outcome knowledge, but the last pragmatic response (either biased or unbiased) determined judgments at the third measurement. These findings highlight the short-term malleability of hindsight influences in response to task pragmatics and has major implications for debiasing.

Hindsight is 20/20: we just knew that Donald Trump would win the U.S. election in 2016, or we always thought that a global pandemic would occur in the foreseeable future. That an outcome appears more likely after it occurred than in foresight is known as hindsight bias (Fischhoff, 1975 ; for reviews, see Bernstein et al., 2016; Guilbault et al., 2004; Roese & Vohs, 2012). To demonstrate hindsight bias in the laboratory, two predominant research designs are used (e.g., Pohl, 2007; Pohl & Erdfelder, 2016). In the *hypothetical design*, one group of participants provides naïve answers in foresight, while another group answers in hindsight. For instance, participants read about a newly discovered respiratory virus, which may spread globally (outcome A) or be eradicated (outcome B). In the hindsight group, participants learn that a pandemic, outcome A, occurred (*actual outcome*) but are asked to judge how likely outcomes A and B are as if they did not know the actual outcome. Nevertheless, participants in the hindsight group typically judge the actual outcome as more likely than participants in the foresight group, who did not learn about the actual outcome. In the *memory design*, participants first give answers in foresight. The same participants then learn about the actual outcome. Subsequently, their task is to recall the answers they gave in foresight. Typically, the recalled answers assign a higher probability to the actual outcome than the original answers given in foresight.

Why does hindsight bias occur? The most prominent and advanced models of hindsight bias, SARA (*Selective Activation, Reconstruction, and Anchoring*; Pohl et al., 2003) and RAFT (*Reconstruction After Feedback With Take The Best*; Hoffrage et al., 2000), agree in the assumption that knowledge of the actual outcome alters the memory representation of the scenario (Blank & Nestler, 2007). When participants make judgments from hindsight, the assumption is that outcome knowledge causes an irreversible change in the memory representation that subsequently precludes judges from forming unbiased judgments. These memory-based accounts offer viable theories of sufficient conditions for hindsight bias. However, the term hindsight bias encompasses a variety of empirical observations on different measures of inevitability, foreseeability, and recollection (e.g., Nestler et al., 2010) that may stem from fundamentally different mechanisms (see, e.g., Blank et al., 2008; Roese & Vohs, 2012).

Until now, the hindsight bias has proven to be robust and to resist various de-biasing attempts (for a comprehensive review, see Son et al., 2021). Are people thus, from the second they receive outcome knowledge, incapable of taking a foresight perspective? The present research aims to add to this discussion and extend prior models of the hindsight bias with two aspects: the influence of communicative pragmatics, and memory consolidation of task information and self-generated responses across time. We outline how hindsight bias can be explained as a malleable response to pragmatic demands of the task, and how the same process can enable unbiased judgments even while having outcome knowledge. However, memory consolidation constrains this assumed flexibility and leads to long term representations of either biased or unbiased responses.

Preview of the Present Research

To illustrate the experimental approach used to demonstrate this, consider the repeated-measures design depicted in Figure 1. For two different scenarios with two possible outcomes (A vs. B), each participant provides probability estimates at three measurement points (M1- M3), but in a different order of instructions per scenario. In the *hindsight-first* order (upper sequence of Figure 1), participants first receive hindsight instructions, which include the actual outcome. However, they are told to always respond as if they did not have outcome knowledge. Directly after that, at M2, they encounter the same scenario without the actual outcome (foresight instruction). Of course, they already possess outcome knowledge from the prior trial; we thus call this condition "foresight" (with exclamation marks). The *foresight-first* sequence for the other scenario (lower sequence in Figure 1) starts with a genuine foresight judgment at M1, followed by a hindsight judgment when participants are informed about the actual outcome at M2. Two weeks later at M3, participants respond to both scenarios again under foresight instructions. Naturally, this is a "foresight" judgment, as participants possess outcome knowledge from either M1 (hindsight-first) or M2 (foresight-first).

Thus, comparing M1 and M2 judgments in the foresight-first condition allows us to test whether subsequent hindsight judgments reproduce previous foresight judgments. This is a variant of what Pohl (2007) calls a memory design that instructs participants to exclude outcome knowledge from their judgments but does not directly ask for recall of prior judgments. The hindsight-first condition offers a completely novel reversal of this design: Do participants manage to construe typical foresight judgments at M2, even though they already have outcome information from preceding hindsight judgments at M1?

Within the same design, a comparison of M1 judgments in the hindsight-first and foresight-first conditions resembles what Pohl (2007) calls a hypothetical design. Here, hindsight bias means that participants who receive a scenario with the actual outcome at M1 (hindsight-first) judge this outcome as more likely than participants who receive the same scenario with foresight instructions at M1 (foresight-first).

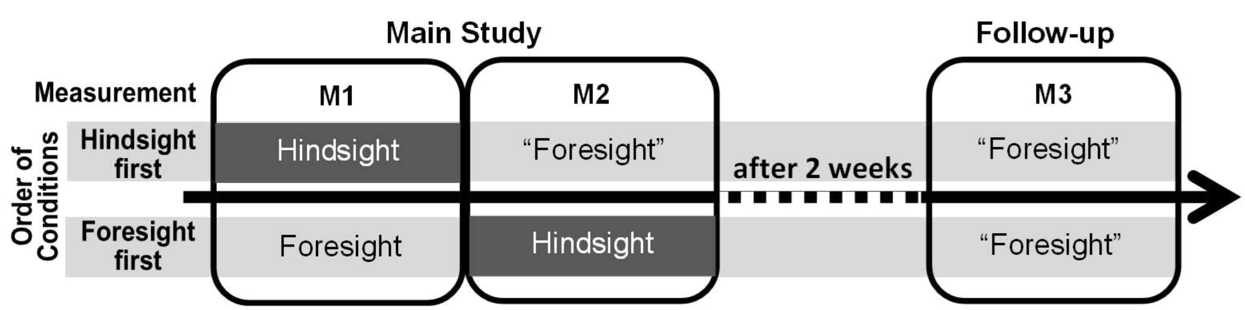


Figure 3.3./1: Within-participants design calling for repeated outcome probability judgments of two scenarios across three measurement points (M1–M3). Participants estimate the outcome probabilities for the two scenarios in different sequential order conditions (foresight-first vs. hindsight-first). We denote measurements in which participants receive foresight instructions, but already possess outcome knowledge from a prior measurement, as "foresight".

With the final M3 judgment, we test participants' long-term problem representation, after a delay of two weeks. If the crucial determinant is outcome knowledge, which participants either received at M1 or M2, both M3 judgments should resemble a typical hindsight judgment. If, however, M3 judgments are sensitive to memory for the participants' own previous judgments and to pragmatic influences of previously communicated probabilities, the influence of their own prior behavior may dominate the impact of existing outcome knowledge.

Let us elaborate on the pattern of empirical findings that can be expected on theoretical grounds. The mechanistic accounts cited at the outset agree in predicting that in the crucial hindsight-first condition, participants should exhibit hindsight bias in M1 and M2. They could hardly explain why "foresight" judgments at M2, after genuine outcome knowledge is induced at M1, should resemble naïve foresight judgments given in the absence of outcome knowledge. However, exactly such a pattern of malleable and reversible hindsight effects, regardless of prior outcome knowledge can be derived from a pragmatic theory perspective, if we assume that participants can frame their judgments according to given linguistic-pragmatic constraints.

Consistent with principles of pragmatic communication (Grice, 1975; Schwarz, 1996), participants may not be bound to generating factual judgments that reflect the entirety of all relevant information about the scenarios. Rather, they may be capable of construing the same judgment in different ways, depending on pragmatic task constraints, such as communication partners' attitudes and an understanding of the intentions guiding their communication (e.g., Higgins & McCann, 1984).

Specifically, regarding the repeated-measures design in Figure 1, we assume that prior to the two-week memory consolidation between M2 and M3, probability judgments are quite malleable, allowing the participants to construe the outcome probabilities of the same scenarios in flexible ways. It is likely that when they receive the typical hindsight instruction, participants believe it to be pragmatically appropriate to construe their judgments in line with the outcome knowledge, which their communication partner, the experimenter, features so prominently. In such a task setting, even an explicit instruction to ignore the outcome may only reinforce, rather than reduce, the focus to the outcome as a relevant cue for mental construal. Meanwhile, at M2 in the hindsight-first order, the experimenter asks participants for the same judgment again, while the outcome notably disappears from the task. This presumably communicates to participants that giving a judgment that differs from their prior answer and excludes outcome knowledge is now appropriate.

When judges indeed respond to pragmatic task constraints, they should judge the actual outcome as more likely under hindsight instructions (with actual outcome) than under foresight instructions (without actual outcome), regardless of the order of instructions. Thus, from this perspective, we not only expect to replicate the typical finding of a memory design (i.e., increasing probability judgments of the actual outcome from foresight at M1 to hindsight at M2) but also the seemingly paradoxical reversal (decreasing probability judgments from hindsight at M1 to "foresight" at M2), despite the available outcome knowledge.

Hypothesis 1: Likelihood judgments of the actual outcome are higher under hindsight instructions than under foresight (or "foresight") instructions, independent of the sequential ordering of both judgment conditions.

Constraints on constraints. Yet, the malleability of probability judgment construal is not unlimited. Referring to Hastie and Park's (1986) distinction of memory-based and on-line judgments, we expect judgments to be malleable while the on-line construal process is still ongoing rather than complete. The time range of the malleable phase can be expected to cover M1 and M2, during which participants receive the same scenario twice in direct succession, without any extraneous mental activities and consolidation period in between. However, once a two-week delay between M2 and M3 provides ample time for consolidation, we assume that a strong pre-formed judgment in memory becomes inevitable. Accordingly, we predict a distinct shift from malleable on-line construal of judgments (during M1 and M2) to pre-formed memory-based judgments at M3, with preceding M2 judgments being the strongest predictor of the final M3 judgments.

We propose that repeated responses in a hindsight paradigm are subject to a pronounced generation effect, that is, a selective memory bias in subsequent judgments toward self-generated information from preceding judgments (Bertsch et al., 2007; Slamecka & Graf, 1978). A canonical exemplar of such a self-generation effect in cooperative communication is the saying-is-believing effect (Hardin & Higgins, 1996; Higgins, 1992, 1999; Higgins & Rholes, 1978). It is typically observed within the audience-tuning paradigm, where participants describe a person to another person. When participants tune this description towards the attitude of the audience, this also shapes their own memory representation and attitude towards the described person. For this effect to occur, participants need not initially believe in their answer or hold it as the single best estimate (Hardin & Higgins, 1996; Higgins, 1992, 1999). People are not only accountable for what they communicate to others (Tetlock, 1983); the well-known memory advantage of self-generated information also implies that one's own previously communicated self-generated judgments may exert a similarly strong, or even stronger, impact on one's subsequent knowledge than experimenter-provided outcome information. Thus, the high or low probability judgment induced by the pragmatic prompts at M2 should determine the final M3 judgment.

Hypothesis 2: After two weeks, likelihood judgments of the true outcome made under "foresight" instructions (M3) resemble the last judgments that participants provided (M2).

Note that the predicted pattern diverges in distinct ways from prior mechanistic accounts, which converge in predicting that participants cannot stop using available outcome knowledge. The pragmatic perspective does not require a distinction between "legitimate" and "forbidden" information (e.g. Fischhoff, 1975; Pohl & Hell, 1996) or a rationalist assumption that the intrusion of (forbidden) outcome knowledge is in line with Bayesian updating processes (Dietvorst & Simonsohn, 2019). To set our pragmatic account apart from the assumption that hindsight effects constitute deliberate attempts to overcome inaccuracy and bias, we included a measure of participants' metacognitive insights in judgment biases in the following research.

Method

Materials

We adopted the four stimulus scenarios from Slovic and Fischhoff (1977). Each scenario involved two possible outcomes (A and B). For instance, in one scenario researchers successfully influenced a hurricane with a chemical, with the outcomes "the strength of the hurricane increased" (A) and "the strength of the hurricane decreased" (B). Under foresight instructions, participants read the vignette and then immediately judged the probability of both outcomes. In the hindsight instructions, the sentence "The actual outcome was (A/B)" was appended to the vignette. The actual outcome was A for two scenarios and B for the two other scenarios. The computer dialogue forced participants to enter two percentages (one for each outcome), which had to sum up to 100%. Within this paper, we always report the judged probability of the actual outcome as the relevant judgment, meaning that an increase in judgments under hindsight instructions compared to foresight instructions shows hindsight bias. The scenario and actual outcome for hindsight instructions remained visible during the judgments. The entire experiment was administered as an online questionnaire using SoSci Survey (Leiner, 2020). Instruction provision and data collection were fully automated, warranting complete anonymity and eliminating all experimenter influences.

Design and Procedure

The repeated-measures design is summarized in Figure 1. Every participant received two scenarios, for which three measurement points (M1, M2, M3) were collected: M1 and M2 in the main study and M3 two weeks later in the follow-up. For one scenario, participants received hindsight instructions at M1 and then foresight instructions at M2 (*hindsight-first*, upper sequence in Figure 1). For the other scenario, they started with foresight instructions at M1, followed by hindsight instructions at M2 (order called *foresight-first*, lower sequence in Figure 1).

The allocation of scenarios to the order of instructions was counterbalanced, as were the scenario pairs allocated to a given participant. Whether the foresight-first or hindsight-first condition appeared first was also counterbalanced. In summary, we employ a 2x3 within-participants design with the factors order of instructions (foresight-first, hindsight-first) and measurement (M1, M2, M3).

In the main study (collection of M1 and M2), participants first received detailed instructions. We explicitly informed all participants about the task and the typical differences between probability judgments made in foresight and hindsight (the bias regularly occurs even if participants are aware of it, see, e.g., Pohl & Hell, 1996). They were also informed that the study would continue two weeks later. They then responded twice to each of two scenarios. Judgment prompts for each scenario appeared on a new page, asking participants to provide two types of ratings: the probability of each outcome and a metacognitive self-report measure in which we asked participants to indicate on a 7-point scale which of the typical responses their judgment resembled more (from 1 = *foresight* to 7 = *hindsight*). Please note that the initial instructions about the hindsight bias enabled participants to base this comparison on a common standard. When participants received the same scenario for a second time (M2), they started with the following instructions: "Now, reread the scenario. If you received outcome information on the previous page, try not to use it for your judgments". Upon completion of two judgments for both scenarios, participants provided demographic data. The duration of the first session was about 7 minutes (3 to 12).

Roughly two weeks later, participants were invited again to the follow-up (collection of M3). Now all participants received all four scenarios under foresight instructions, with the following additional instruction: "You already know two of these scenarios from the first survey on this topic. Please respond as you did in the foresight condition." Again, each scenario appeared on a new page, in random order. For each of the scenarios, we asked participants to indicate their meta-memory experience. Three response options were provided: "*This scenario is new to me*," "*I know this scenario,*" and "*I remember my prior responses to this scenario exactly*." This question includes two relevant aspects: whether an item is recognized (yes/no) and, if so, whether participants merely know that they have seen the scenario or have a recollective experience, remembering the precise episode and which judgments they had given in the main study. Note that the latter distinction is aimed to discriminate between semantic and episodic memory (Tulving, 1972), operationalized as "remember" versus "know" (Rajaram, 1993). The total follow-up lasted about 4 (3–7) minutes.

Participants

The main study included 90 participants (68 female, age: *m* = 24.74, 20 – 55); 64 of them remained in the follow-up (48 female, age: *m* = 25.09, 20 – 55). The target sample size was determined by a simulation-based power analysis documented in the Appendix. We recruited participants from social psychology courses for Bachelor students at Heidelberg University. Participation was unpaid. Participants could choose, both at the start and at the end of the session, to preclude us from analyzing their data (3 excluded for this reason). Five further participants were excluded due to irregularities (repeated participation, less than a day from the main study to follow-up). Incomplete attempts were also deleted (main study: n = 12, follow-up: n = 1). On average, the delay between main study and follow-up was 13.5 days (9 – 21). Participants provided informed consent before each study part and were debriefed at the end. We completed data collection before data analysis commenced.

Statistical Analyses

For our analyses, we used R 3.6.1 (R Core Team, 2019), as well as the packages lme4 (Bates et al., 2015), lmerTest (Kuznetsova et al., 2017), sjPlot (Lüdecke, 2019), simr (Green & MacLeod, 2016), and ggplot2 (Wickham, 2016). We used mixed-effects modelling to account for our longitudinal repeated-measures design. To aid readability, we provide the results in the more common ANOVA format within the main text. The interested reader finds the full mixed-effects model analysis in the Appendix.

Results

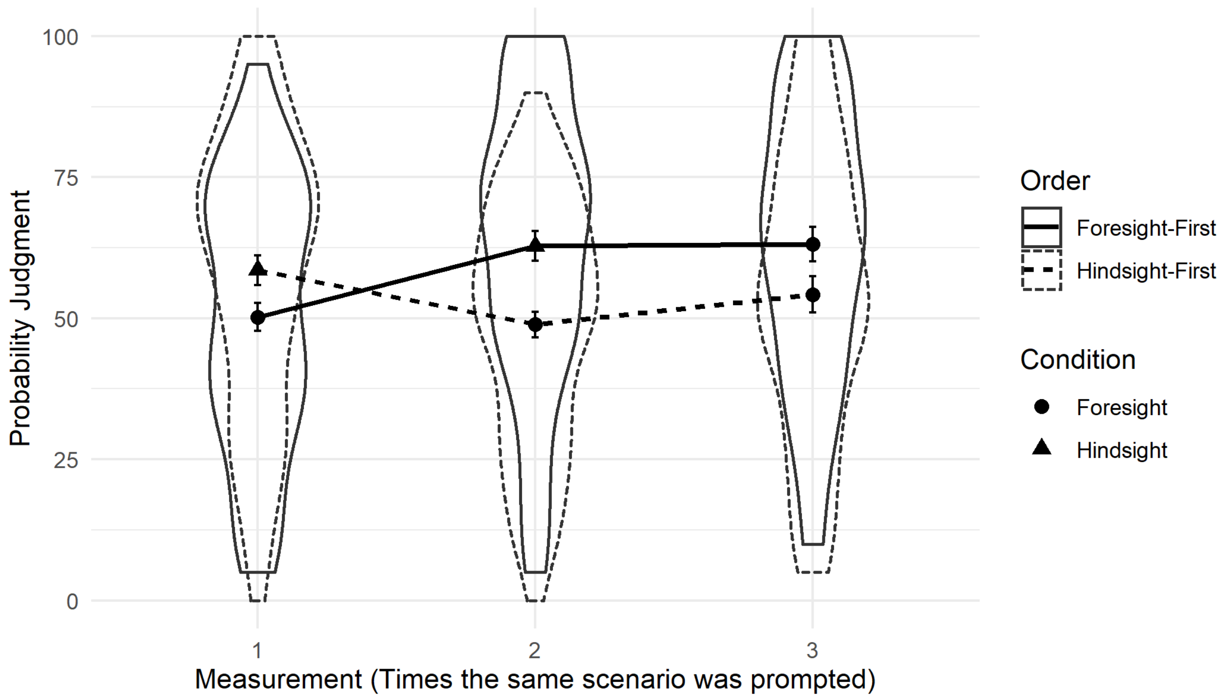


Figure 3.3./2: Violin plot with corresponding line graph depicting the means, standard errors and distributions of outcome probability judgments as a function of measurement (M1, M2, M3), condition (circles vs. triangles), and ordering of foresight and hindsight condition (solid vs. dashed lines).

Probability Judgments – Main Study and Follow-Up

Figure 2 displays the judged probability of the actual outcome across conditions and measurements. We replicated the classic between-participant hindsight bias in the hypothetical design. At M1, probability judgments under the hindsight instruction (m = 58.85%) were significantly higher than under the foresight instruction, with no outcome knowledge (m = 49.83%), b = 8.60, t(88.37) = 2.25, p = .01, d = .36. The classic within-participant finding in the memory design was also replicated: After participants had provided judgments under foresight instructions at M1 (m = 49.83%), they judged the actual outcome as more likely under hindsight instructions at M2 (m = 62.79%), b = 12.60, t(89) = 5.92, p < .001, d = .52.

Hypothesis 1. In the critical hindsight-first order, the mean probability estimate for the actual outcome decreased from 58.85% with hindsight instructions at M1 to 49.22% with foresight instructions at M2. This decline from hindsight to "foresight" constitutes a well-tuned downward adjustment, despite formerly induced outcome knowledge, matching the strength of the classic between- and within-participants hindsight bias.

These results were reflected in a two-way repeated-measures ANOVA with the factors instruction (foresight, hindsight) and measurement (M1, M2). There was a significant main effect of instruction: likelihood judgments were generally higher under hindsight instructions (m = 60.66) than under foresight instructions (m = 49.54), $F(1, 264.93) = 27.82$, $p < .001$. The measurement (M1, M2) did not have a significant influence on judged likelihood, $F(1, 266.19) = 1.18$, $p = .28$. There was also no interaction between instruction and measurement, $F(1, 264.93) = 0.50$, $p = .48$. A stable hindsight bias was obtained regardless of whether the foresight condition was followed or preceded by the hindsight condition (see Figure 2). Put differently, participants provided judgments that resembled genuine foresight even when they possessed outcome knowledge.

Hypothesis 2. Two weeks later at M3, likelihood judgments of the true outcome made under foresight instructions mirrored the judgments that participants provided at M2. The mean judged probabilities of the actual outcomes in the foresight-first (m = 63.16%) and the hindsight-first order (m = 54.22%) were still significantly different, $b = 8.61$, $t(57.36) = 2.25$, $p = .028$, $d = .34$.

Thus, when participants concluded the main study with a foresight judgment (hindsight-first order), there was less hindsight bias after two weeks than when the main study concluded with hindsight instructions (foresight-first order). When participants last judged under hindsight instructions before the consolidation period of two weeks, they showed a pronounced hindsight bias despite receiving foresight instructions at M3.

Metacognitive Insight – Main Study

The responses to the metacognitive insight scale followed a pronounced bimodal distribution. We therefore dichotomized responses (1-3 changed to 0, close to foresight, and 5-7 to 1, close to hindsight) and excluded responses on the median value of the scale (4), as their interpretation is ambiguous. After transformation, 180 responses were coded as "similar to foresight" (0) and 133 responses as "similar to hindsight" (1).

In the foresight-first order, participants indicated less frequently that their judgment was "similar to hindsight" ($m_{\text{Foresight}}$= 39%, $m_{\text{Hindsight}}$= 36%) than in the hindsight-first order ($m_{\text{Foresight}}$= 46%, $m_{\text{Hindsight}}$= 49%). This is at odds with participants' actual judgments. Their responses in the main study depended only on the judgment condition, not on the presentation order. Suppose participants were fully sensitive to the difference in their provided judgments. In that case, the received instruction (hindsight *vs.* foresight) and not the presentation order should guide their responses. Still, participants seem to think that the outcome influences them in the hindsight-first order, but not the foresight-first order, independent of the trial's actual condition – thereby contradicting their probability judgments.

Meta-Memory – Follow-Up

Recognition of the scenarios at the follow-up was almost perfect, with two false alarms in 256 responses. Therefore, we attempted no further analyses of recognition rates. Of the 128 recognized items, participants indicated 92 as known (72%) and 36 as remembered precisely (28%).

When participants reported that the scenario was new to them, the mean judged probability of the true outcome was 42.96%, 95% *CI* = 35.35 – 50.56, p <.001. It increased by 11.38% for trials in which participants knew the scenario, 95% *CI* = 5.50 – 17.27, p <.001, d = .50. With 24.61%, this increase was highest when participants reported that they remembered their main study's responses exactly, 95% *CI* = 16.47 – 32.76, p <.001, d = 1.07. The persistent influence of the outcome information at the follow-up is visible in the descriptive data provided in Figure 3.

Surprisingly, reported recollective memory of preceding judgment scenarios did not lessen the hindsight bias but even strengthened it.

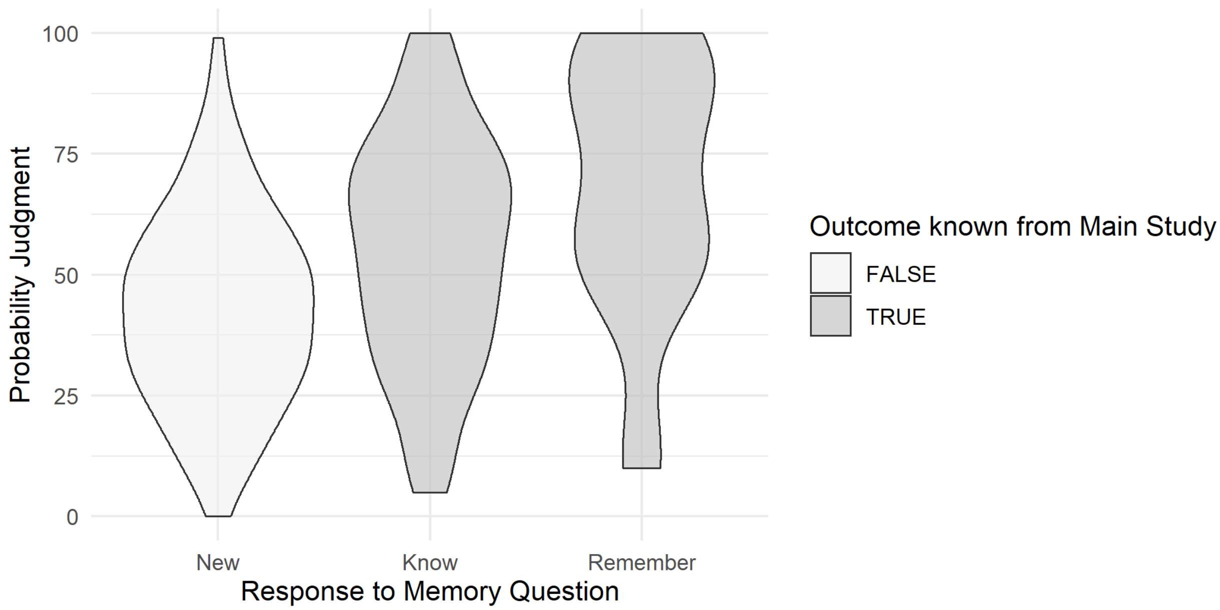


Figure 3: Distribution of probability judgments for the true outcome at the follow-up as a function of outcome knowledge and self-reported meta-memory experience.

Discussion

Within our novel experimental design, we find a completely reversible, symmetrical pattern of hindsight effects at M1 and M2, while the judgments made after a two-weeks consolidation period at M3 mirror the judgments at M2. Thus, we find a malleable construal effect in response to pragmatic demands while judgments are made on-line: participants succeeded in providing unbiased responses, even if they possessed outcome knowledge and just gave biased responses under hindsight instructions. After consolidation, this malleability is replaced by memory-based judgments consistent with participants last responses. This finding aligns with research on constructive memory (e.g., Loftus, 1975), audience tuning, and the saying-is-believing effect (Hardin & Higgins, 1996; Higgins, 1992, 1999). When no other communicative or self-generated prompts occur, the labile representations formed in cooperative communication become part of a consolidated, stable memory representation.

This distinct pattern of findings is plausible from our theoretical approach, which considers pragmatics and constructive memory, but it can hardly be explained by prior mechanistic accounts of hindsight effects. Interestingly, participants had minimal metacognitive insight into their responses, which does not align with recent claims that the biased responses in the hindsight paradigm reflect a deliberate choice of participants (Dietvorst & Simonsohn, 2019).

We explored the interplay of pragmatics, constructive, and reconstructive memory using predictions for fictional experimental settings as an exemplary task from the literature. Further studies should replicate these findings with other material, such as almanac questions and historical events, while also including foreseeability, inevitability and confidence in judgments (Ackerman et al., 2020) as dependent variables. Additionally, our understanding of cooperative communication as an influence on hindsight bias can be broadened by introducing different linguistic-pragmatic constraints, like the instructions of dialectical bootstrapping (Herzog & Hertwig, 2009). Further research is also needed to determine whether the long-term influence of pragmatic answers within our study shares the typical moderators of saying-is-believing effects, such as audience characteristics and communicative goals (Echterhoff et al., 2005, 2008).

The flexibility of answers given in communication and the long-term constructive influence of these answers on memory oppose a narrow definition of rationalist updating towards more accurate knowledge. However, we argue that both the malleability of answers to fit the communicative context and their long-term influence should be considered adaptive in a world in which accurate predictions, but also social bonds and acceptance by others, determine success. Reacting to communicative-pragmatic constraints and forming a stable shared representation allows us to further our social goals and effectively act together.

Importantly, we also demonstrated how this mechanism can lead to more accurate judgments. Sometimes, for instance after receiving inaccurate outcome information, learning from feedback leads to less accurate judgments. When social sharing prompts someone to exclude inaccurate information directly after receiving it, our findings suggest that they will often succeed. Crucially, we demonstrated that this correction can be temporally stable. If we aim to debunk false claims and enhance prediction accuracy, knowledge of the subtle, low-effort manipulation used in the present research could inspire effective communicative strategies and interventions.

This investigation has demonstrated how communicative malleability in response to communicative-pragmatic constraints, over time, becomes an inextricable part of our memory. Hindsight effects can stem both from the reconstructive nature of memory and the constructive malleability necessary in an uncertain and social world.

4. Rational Choice and Pragmatics

The prior chapter presented research inspired by Gricean pragmatics, where communication is construed as a cooperative language game in which everyone follows the rules. Considering pragmatic inferences in experiments, deviations from what is defined as rational within a research paradigm suddenly appear as a misunderstanding between experimenter and participant. Once this misunderstanding is cleared up, participants may provide perfectly rational responses, if we define rational as following a model of reality that is normatively sound and testable, as is discussed in more detail in the article below. At the same time, it demonstrates how easily communication and its influence on judgments and memory complicates rational decision-making.

In this chapter, I will argue that this is not a dysfunctional deviation from the norm of faithful information transmission, but rather an indication of communication as geared towards taking action and achieving goals in a social world, sometimes to the detriment of rationality. The article in Chapter 4.2. then demonstrates how we can easily use communication to hinder participants in making decisions that follow normative mental models based on sound judgments.

4.1. Competitive Communication and Rationality

Chapter 3 characterized communication as a tool to share meaning and partake in rule-based social situations. Another concept of communication highly useful for JDM theorizing is that use of language extends our ability to act in the world (Hörmann, 1976) – a notion popularized and investigated by *speech act theory* (Austin, 1962; Searle, 1969). This pragmatics theory proposes that when I ask someone to take a walk with me on a rainy day, and they answer, "I have forgotten my umbrella.", this will not only lead me to understand what is said (the *locutionary* act that is described by syntax and semantics). In this context, I will probably understand the speaker to mean that they would like to come but cannot, as they would need an umbrella – the listener also represents the *illocutionary* act, which Grice (1975) described as "what is implicated". Crucially, speech act theory adds a third level of analysis, the *perlocutionary* act. This is the effect of the utterance on the listener and in the world. In our example, I might offer the other person my spare umbrella so that they can accompany me. As is easily imaginable, the perlocutionary act may differ immensely from what the speaker meant to do – maybe the forgotten umbrella was the excuse of a person who wanted to reject my offer of a walk gently and did not want me to take away their excuse. In this case, I misread the illocutionary force. In other cases, speech acts may have incidental, surprising, and unwanted perlocutionary results that do not follow directly from the illocutionary act (see Fiedler's discussion of non-referential language effects, 2008). While Austin (1962) started with the assumption that only some speech acts contain perlocutionary force – act in the world – he later concluded that all speech acts have constative (state things) and performative (do things) features.

Speech act theory provides us with a focus on communication as action, which is directed towards goals. In this view, communication is successful when its goals are met (Luhmann, 1984) – this need not be faithful information transmission or supporting the other person in rational choice (Fiedler, 2008)! Indeed, both communication partners may join in a language game that is not geared towards logic, reason and truthfulness. For instance, they might rather support a positive relationship and self-image with white lies required by etiquette. But more importantly, the speaker's goals may clash with a striving towards adaptive choice of the listener, turning a cooperative communication game into a competitive one. Take our example from Chapter 2. Buying property can be a difficult endeavor, as realtors and homeowners are usually motivated not to provide the potential buyer with the most truthful impression to inform a choice that is best for them, but to sell their own property. Arguably, language provides sellers with the tools to favorably influence the judgments and preferences of the potential buyer and achieve their goal. "To do things with words" (Austin, 1962), communicators can use locutionary acts that appear harmless – but still apply an illocutionary force that brings about the goal-congruent perlocutionary results. In a competitive situation, that may as well be having them form biased judgments and decisions that are detrimental to the listener. I subsume this instrumental use of pragmatics under the term competitive communication, in contrast to the purely cooperative concept of communication presented in the prior chapter.

It is important to note that the competitive use of communication works precisely because communication is usually fairly cooperative, and extreme violations of conversational rules are comparatively rare (see, e.g., a discussion of this for deception in truth default theory, Levine, 2022). Communication would break down if the cooperative principle was usually violated. Like for the shepherd boy in Aesop's fable, who cried wolf until no-one believed him and came to his rescue, language fails to achieve the desired perlocutionary force if the listener notices to be misled time and again. This favors the use of inconspicuous communicative means to achieve the goal of the speaker in a competitive situation – successful social influence, persuasion and deception are based on subtlety. Subtle competitive communication not only avoids detection and resistance by the listener, but also hedges against accountability if the attack against the goals of the listener is detected. After all, the unwanted perlocutionary force may as well be incidental or based on a misunderstanding of the locutionary or illocutionary act. If a potential buyer interprets the realtor's advertisement that a flat is comparatively quiet as such that they are disappointed by the actual noise level, is the realtor at fault for misleading them or did the buyer just assume a standard of comparison not implied by the realtor?

Communication, particularly through language, provides speakers with a great variety of subtle means that can become a complicating condition for rational choice of the listener. In Chapter 3, the influence of any given information, even if marked as irrelevant, and the constructive memory influences of leading questions already provide tools to have listeners form biased representations. Leading questions are a specific example of how *presuppositions* within utterances affect the listener, even if they are negated, doubted or refuted by the listener or even the speaker. Research on *misinformation* and its correction has amassed considerable evidence of similar perseverance effects (e.g., PeConga et al., 2022). Another notable example are shifts in *implicit verb causality* that can affect person perception (linguistic category model; Semin & Fiedler, 1991) and attitudes toward social groups (Maass, 1999), as elaborated on in Chapter 2.2.

Within JDM research, the central example are *framing* effects (for an overview, see Kühberger, 2022). Most frequently, they are demonstrated using the Asian disease problem (Tversky & Kahneman, 1981): To combat an epidemic, participants choose between a riskier (A) and a less risky (B) option. In one condition, these options are framed as survival rates (option A: "with 1/3 probability, 600 people are saved, with 2/3 probability, 600 people are not saved"; option B: "200 out of 600 people are saved"), in the other condition as mortality rates (option A: "with 1/3 probability, none of 600 people die, with 2/3 probability, 600 people die; option B: "400 out of 600 people die"). When the same rates are expressed by the number of deaths, participants make riskier choices than when the number of surviving people is focused – what someone may perceive as a small linguistic change considered irrelevant in economic decision-making theories leads to a preference reversal. This reversal is usually explained by prospect theory, which employs a social cognition meta-theoretical lens, but several authors (e.g., Bless et al., 1998; Mandel, 2014; McKenzie & Nelson, 2003) also suggest possible mechanisms based on pragmatics. These studies provide an additional example of how conversational implicatures may lead to cognitive illusions, as discussed in Chapter 3.

These examples are only the tip of the iceberg – communication is replete with subtle means to systematically influence decision-making. Particularly in more competitive social situations, these means are available to divert choice from following a normative mental model – communication becomes a complicating condition for rationality. The following article shows how easily, by receiving a more elaborate story to provide the same basic details about choice options, participants abandon their apparently rational choices. This leads to the question whether a standard for adaptive rationality should be that it is robust against influences of competitive communication. After all, decision-making outside the laboratory happens in an environment where people are surrounded by others who intuitively use communication to reach their goals.

4.2. Toward stronger tests of rationality claims: Spotlight on the rule of succession. (Fiedler, Salmen & Prager, 2022)[3]

A distinct statistical inference rule, the Laplacian rule of succession (RoS), is used to illustrate a general problem of decision research, namely, empirical tests of human rationality. The RoS relates the probability p of a dominant outcome in a population of binary events to an observed proportion P of the dominant outcome in a sample. The inferred probability $\hat{p}$ is generally regressive; it deviates more from the sample proportion P when samples are small rather than large. Based on computer simulations of $\hat{p}$ related to P under various boundary conditions (measurement error; prior odds; payoff expectation, population inference vs. sample estimation), we examined judges' RoS sensitivity in a series of simple lottery experiments. Although comparative ratings of lotteries experienced in small and large samples seemed to reflect some intuitive understanding under most simplifying task conditions, they violated normative principles when complicating conditions (e.g., loss aversion, uncertainty aversion, ratio biases, Bayesian priors) overshadowed the RoS. The final discussion revolves around the difficulty, or impossibility, to conduct empirical tests of rationality.

The present article is concerned with a central issue of decision research, which is however often neglected in the pertinent literature, namely the distinction between rationality proper and people's preparedness to emulate responses that resemble or mimic rationality. Here we are not using the word "rationality" as a synonym of such criteria as individual payoff maximization, correspondence with some accuracy criterion, coherence with a normative law, or hedonic measures of satisfaction and consumption. There are good reasons to contend that individually maximized payoffs may be collectively mal-adaptive and conflict-prone (as in social dilemmas, see van Lange et al., 2013); correspondence can reflect a spurious correlation (Fiedler, 2000; Kutzner & Fiedler, 2015); coherence is equivocal when different normative rules are in conflict; and what is hedonically superior today may be inferior tomorrow or in a year.

If we had one of these meanings in mind, we could use precisely these well-defined terms. Let "rationality" instead refer to thoughtful cognitive inferences that rely on good reasons, that is, on explicable and testable mental models of reality, models anchored in normative rules that impose logical constraints on judgments and decisions, which need not be supposed however to provide ultimate and incontestable solutions. Rationality need not entail the claim of optimality or normative uniqueness. Even when a mental model does not offer an ultimate solution but only a viable or satisficing perspective – as is typical of science – the reasoning process it informs is rational in the sense that (a) the mental model is stated transparently and can thus be validated critically, and (b) inferred problem solutions follow from a deliberate set of clearly articulated rules (Kruglanski, 1990; Sarbin, Taft & Bailey, 1960), much like syllogistic conclusions follow from minor premises (data) subsumed under major premises (model).

However, the vast majority of references in the literature on judgment and decision making (JDM) do not treat rationality as well-reasoned, transparent inferences guided by distinct constraints. It is rather common to define rationality in terms of uniquely correct, optimal solutions (Le Mens & Denrell, 2011; Lieder & Griffiths, 2019), fitting a normative criterion of coherence (e.g., Bayesian calculus) or a correspondence criterion of maximized profit.

To be sure, nominal definitions are a matter of convention and nobody is entitled to determine which definition of rationality is right or wrong.[4] Yet, in cumulative science, there are good reasons why definitions are not arbitrary but vary a lot in usefulness, conceptual clarity, parsimony and in the quality of research they inspire. For new definitions to be useful, they should fit the conceptual framework of other already established definitions and paradigms rather than causing conceptual confusion. They should not be theoretically empty but lead to distinct and testable implications, aligned with major empirical distinctions. And, it is important to keep Occam's razor in mind. The principle of parsimony prohibits scientists from founding two or more definitions for the same referent. With regard to all these pragmatic criteria, it seems justified to ask whether rationality is a parsimonious concept if its meaning is the same as profit maximization; if rationality refers to an empty set of allegedly optimal behavior that is unlikely to be ever reached; or if different normative criteria render optimal solutions equivocal.

Apart from these conceptual vicissitudes, particularly serious problems arise when the term rationality serves to interpret empirical findings. One common source of confusion is that absence of (certain sources of) irrationality is not the same as rationality. An increasing number of experiments and simulation studies demonstrate that biased judgments and decisions can arise in the absence of biased processes (Costello & Watts, 2019; Denrell & Le Mens, 2008). Even in the absence of wishful thinking, one-sided motives, conflicts, vested interested or resentments, when all stimuli are equally likely to be noticed, perceived, remembered and retrieved, correct assessment and calculation and Bayesian updating may nevertheless produce countless biases and illusions, reflecting measurement sampling error, regressive shrinkage, and other pitfalls of the probabilistic world, such as selective feedback. Yet, when experiments on illusory correlations (Costello & Watts, 2019; Fiedler, 2000), confirmation bias (Fiedler, Walther & Nickel, 1999), overconfidence (Moore & Healy, 2008) or preference reversals (Walasek & Stewart, 2015) rule

[4] Max Weber alone referred to four different types of rationality (Kalberg, 1980)

out motivational and affective biases and mental insufficiencies, they do not provide cogent evidence for rationality proper. Controlling for distinct sources of irrationality on selected tasks cannot establish the ideal of rationality. After all, a truly rational agent might recognize and correct for a bias – normally and innocently as it may have arisen – in a metacognitive quality check (Fiedler, Ackerman & Scarampi, 2019).

How could a single empirical investigation ever establish an unequivocal proof of a principle as unique as rationality? Virtually all empirical tests are contingent on the control of internal and external validity (Campbell, 1957), which is never perfect. Even impressive empirical demonstrations of seemingly miraculous rationality, such as optimal reduction of uncertainly in visual perception (Trommershäuser, Maloney & Landy, 2008), may be peculiar to selective stimuli and task settings (see Marcus & Davis', 2013, critique of overstated rationality claims). Science is a pluralistic endeavor that always allows for Panglossian, Meliorist, and Pessimist perspectives on rationality (Stanovich, 2011), making an existence proof of a non-irrational decision much more likely than a universality proof of rationality.

Genuine versus incidental rationality. Let us illustrate the crucial difference between genuine rationality and incidental findings that happen to match normative predictions without providing unequivocal diagnostic evidence for rationality, with respect to the illusory-correlation example focused in Costello and Watts (2019). Given the same (75%) trend toward more positive than negative behaviors in a large group (18 out of 24) as in a smaller group (6 out of 8), a more favorable impression of the majority than the minority need not reflect an irrational "illusion" or memory bias favoring negative minority behaviors (Hamilton & Gifford, 1976; Kutzner & Fiedler, 2015). Such an illusory correlation may result from a non-selective, unbiased induction algorithm (Fiedler, 1991, 2000). Bayesian calculus or lesser regressive shrinkage in large than in small samples offer sufficient accounts for 18 out of 24 providing stronger evidence for prevailing positivity than 6 out of 8. Yet, this one-time correspondence with an unbiased algorithm by no means proves rational impression judgment. Rational judges ought to engage in metacognitive monitoring and control, taking all parameters and boundary conditions of the algorithm into account.

Rule of Succession. A rational participant in an illusory-correlation experiment should in particular live up to the rule of succession (RoS; de Finetti, 1937), the topic of the present investigation. As clearly explained by Costello and Watts (2019), an illusory correlation favoring majorities can originate in an unbiased, normatively justified "… inference from samples with equal feature rates but different category sizes and no additional information" [p. 446]. Yet, mere coherence with an unbiased algorithm in one inference task can hardly prove rationality.

Let us briefly introduce the RoS. Assume that the same proportion (P=.75) of focal outcomes is observed in a small (3 out of 4) and a larger sample (18 out of 24). It is easy to see that the expected p of focal outcomes in the universe from which samples are drawn is lower (i.e., less extreme) than the observed sample proportion P. The principle of insufficient reason (Savage, 1954) implies that in the absence of other prior knowledge, every possible value of p could have produced P=.75. Given P=.75, there are clearly more possibilities for p to be below rather than above .75; p can be expected to be less extreme than P. Flat priors render inferences of p from P regressive (Campbell & Kenny, 1999; Fiedler & Krueger, 2012). The degree of regressive shrinkage (of $\hat{p}$ relative to P=.75) increases with decreasing sample size. The expected p underlying an observed P=.75 is lower when sample size is small rather than large.

Indeed, in a very small sample of n=2 or n=3, we may observe the most extreme proportion P=1.0 when the true p is only moderate; this is hardly possible for a large sample of, say, n=100. Conversely, a small sample more likely exaggerates a dominant outcome (p > 0.5) than in a large sample. Specifically, the RoS holds that $\hat{p} = \frac{(k+1)}{(n+2)}$, where $\hat{p}$ is the estimated p, n is sample size, and $k = n \cdot P$ is the number of focal outcomes. Table 1 (left part) gives a quantitative impression of the regressive shrinkage of p relative to P, at sample size n = 10 versus n = 30.

Apparently, given RoS as normative benchmark, observing the same P = .75 ratio of positive behaviors in a small and a large sample justifies the assumption of a higher prevalence p

in the majority than in the minority. Such a normatively justified inference alone can explain an illusory-correlation bias; it is not necessary to postulate any biased cognitive mechanisms.

Yet, cognitive inference that matches the RoS in sign or even in size does not afford cogent evidence for "rationality" proper. It might be a spurious inference that only happens incidentally to coincide with RoS. Given only three possibilities (i.e., the majority is more negative, equal, or more positive), a bias in the correct direction might reflect chance or superficial reasoning. Participants may have heard in a statistics course that larger samples are superior, that reliability increases with sample size, or favoring a majority may reflect a mere-exposure effect (Zajonc, 2001). Treating responses congruent with the RoS as rational requires that they do not originate in a spurious advantage or coincidence. It implies that correct judgments or preferences reflect a well-reasoned, defendable inference process, grounded in a mental RoS model, from which sound inference can be derived systematically, not haphazardly. To be sure, a mental model need not consist of RoS in formal algebraic notation. Rational agents may mentally emulate the underlying Bayesian logic. Yet, an empirical proof of rationality takes more than group means matching the ordinal trend of a normative formula. It must demonstrate the ability to apply the RoS systematically, not incidentally. Indeed, empirical proofs of rationality are intricate.

RoS premises and limiting conditions. The RoS has clearly defined boundary conditions under which its application can be considered rational. Individuals who have acquired rational mastery should understand when to use or not to use the RoS. Most evident is the exclusive application to inferential tasks, where subjects make inferences about non-observed events by estimating a latent parameter *p*. This condition is not met in a task that just requires estimating an observed sample proportion P. Besides, the RoS implies flat priors; it is therefore not applicable when explicit prior knowledge about *p* overrides RoS predictions. The introduction of sample noise (i.e., imperfect reliability of the observed information) adds another challenge to the naïve application of the RoS, which simply ignores the impact of noise.

To test participants' sensitivity to these boundary conditions, beyond the truism that sample size matters, we conducted a series of simple lottery experiments. If correct choices reflect participants' rational competence, they should not utilize the RoS blindly but understand when it does and when it does not apply. Correct responses need not reflect rationality proper. Other, non-rational tendencies might prepare judges to produce responses that mimic seemingly rational rules incidentally. When sensitivity to irrelevant distractors overshadows the correct but shortsighted use of the RoS, one should not praise the resulting mix as proof of rationality.

Understanding RoS Boundary Conditions: Empirical Challenges for Rationality

In the next section, we first present computer simulations of p related to P under distinct boundary conditions. Then we report a series of simple lottery-betting experiments designed to test the participant' understanding of these limiting conditions. In the light of the empirical results, we will finally return to a discussion of the ambitious rationality assumption and the mundane alternative assumption that correct response tendencies on experimental tasks of limited diagnosticity may merely mimic rational trends, which do not provide cogent evidence for rationality proper.

Simulation Results

To simulate the relationship of p and P under different boundary conditions, in which the RoS was either applicable or not, we generated one million binary samples for two different sample sizes (n=10 vs. 30); p varying randomly in the range of $0 \leq p \leq 1$. Because the algorithm remembered the p corresponding to each sampled P, it was possible to determine, separately for n=10 and n=30, the expected value of the p distribution giving rise to the sample proportion P.

Table 1 presents the function relating P to $\hat{p}$. The regressive nature of RoS is apparent from the first three table columns. Given a frequent outcome (P>.5), the corresponding population $\hat{p}$ is lower; if the focal outcome is infrequent (P<.5), $\hat{p}$ is higher than P. This regressive pattern is stronger for small (n=10) than for large samples (n=30). Note that, depending on n, $\hat{p}$ deviates noticeably from P but the magnitude of these RoS effects is very modest. At P=.8, $\hat{p}$ becomes .750

for n=10 as compared to .781 for n=30, respectively, which amounts to a contingency of $\Delta\hat{p}$ = .781 – .750 = .031 in favor of the majority (i.e., the larger sample). A rare sample outcome (P=.2) implies a contingency of $\Delta\hat{p}$ = .219 – .250 = –.031 favoring the minority. Thus, sensitivity for RoS calls for a notably high mental resolution level.

Table 4.2./1: Simulated estimations of population parameters p given sample proportions P, sample size n, and Bayesian priors. (The experiments below used the shaded proportion P=.8 and P=.2).

Observed P	Flat priors: all p equally likely			Distinct priors: $p \geq .50$	
	$\hat{p}$ given n = 10	$\hat{p}$ given n = 30		$\hat{p}$ given n = 10	$\hat{p}$ given n = 30
0.0	0.083	0.031		0.542	0.516
0.1	0.167	0.125		0.549	0.519
0.2	0.250	0.219		0.558	0.524
0.3	0.333	0.312		0.570	0.532
0.4	0.417	0.406		0.588	0.545
0.5	0.500	0.500		0.613	0.570
0.6	0.583	0.594		0.648	0.616
0.7	0.667	0.688		0.697	0.691
0.8	0.750	0.781		0.760	0.781
0.9	0.833	0.875		0.836	0.875
1.0	0.917	0.969		0.917	0.969

The two right-most columns of Table 1 reflect a condensed but similar pattern under restrictive priors, ensuring a dominant rate of focal outcomes ($p \geq .50$). To simulate this case, we estimated $\hat{p}$ from a restricted integral (assuming flat priors in the range of $.5 < p < 1.0$). The function relating $\hat{p}$ to P continues to be more regressive for n=10 than for n=30. The contingency Δp is still negative for low P (reflecting enhanced regression for small samples) but positive (reverse) for (P=.7) or above. Thus, both lotteries should gain in attractiveness when (p≥.50).

Now let us consider the role of noise. Assume that subsequent communication of sampled outcomes is subject to noise (e.g., memory decay), such that a proportion e of all sampled outcomes are replaced by random binary outcomes.[5] Table 2 displays the $\hat{p}$ estimates resulting mingling a proportion of e=.2 or e=.4 of all sampled outcomes with noise. Now, non-regressive $\hat{p}$ is possible. Observing an extreme proportion of P=.8 focal outcomes in spite of noise implies that the true p must have been relatively high. Table 2 shows more inaccurate $\hat{p}$ estimates due to counter-regressive overestimation of for intermediate P values (for more details, see Appendix).

Table 4.2./2: Simulated estimations of population parameters p as a function of sample proportions P, sample size n, Bayesian priors, and amount of noise e

Observed P	Flat priors; measurement error e = .2			Flat priors; measurement error e = .4	
	$\hat{p}$ given n = 10	$\hat{p}$ given n = 30		$\hat{p}$ given n = 10	$\hat{p}$ given n = 30
0.0	0.094	0.035		0.111	0.042
0.1	0.141	0.072		0.148	0.065
0.2	0.209	0.154		0.203	0.112
0.3	0.298	0.266		0.281	0.206
0.4	0.397	0.383		0.382	0.345
0.5	0.500	0.500		0.500	0.500
0.6	0.603	0.617		0.618	0.655
0.7	0.702	0.734		0.719	0.794
0.8	0.791	0.846		0.797	0.888
0.9	0.859	0.928		0.852	0.935
1.0	0.906	0.965		0.889	0.958

[5] Note that this operationalisation of noise refers to P rather than to p.

Further Limiting Conditions

If rational individuals possess an adaptive RoS tool, they should not naively apply it when appropriate conditions are not met, for instance, when sample-estimation tasks are exhaustive, noisy, or when priors are not flat. Moreover, sensitivity to RoS should not be overshadowed by common biases, such as uncertainty aversion (Ellsberg, 1961) or gain versus loss framing. If RoS performance actually deserves the predicate “rational”, it would be essential to demonstrate that judges do not administer the rule uncritically but only when appropriate. When RoS inferences must compete with unwarranted (irrational) inferences, the former should dominate the latter. Such quality control is the least one would expect of a superlative competence called rationality.

Empirical tests of Sensitivity to RoS Boundary Conditions

We tested these considerations in a series of short experiments. We examined participants’ willingness to play one of two lotteries that varied in sample size, both under standard conditions that call for RoS use and under RoS-inappropriate conditions. A simple lottery task minimized the demand for prior knowledge and memory capacity and rendered RoS boundary conditions obvious. In all experiments, participants were given a choice between two Lotteries, A and B, each described by a sample of randomly drawn binary outcomes. Although the observed sample proportion of P = .8 positive outcomes was the same for A and B, sample sizes n_A=10 and n_B=30 varied. As the probability p is higher for the larger (p_B = .781) than for the smaller sample (p_A = .750; see Table 1), an adaptive agent who understands the RoS will choose Lottery B.

Yet, the RoS only applies to inferences from sample proportions P to latent probabilities p. If samples contain the full universe (i.e., p = P), there is no reason to favor B over A. An initial research goal was to examine *Basic RoS Understanding* in a Lotteries (A vs. B) × Task Framing (Inference vs. Estimation) design (see upper-most section of Table 3). The preference of Lottery B over A only holds for inferences of p from P, not for estimates of p=P. Experiments 1, 2, and 3 test this prediction.

Table 4.2./3. Mapping simulation results onto lottery experiments designed to examine RoS reasoning

Experiment	RoS implication	Experimentally tested hypothesis	Empirical result
Basic RoS Unterstanding			
1, 2, and 3	Less regressive *p* inferences (*p* closer to P) with increasing *n*	If constant winning rate is high (.5<P=.80), Lottery B should be preferred to Lottery A if $n_B > n_A$	Basic RoS under-standing: B preferred to A given $n_B > n_A$
1 and 3	This prediction only holds for inferences, not for estimations	No preference of B over A when sample = population	Indeed, B preference confined to inferences, not for estimations
RoS Immunity to Competing Judgment Biases			
1, 3, and 4	RoS is immune to uncertainty (vagueness) aversion	RoS-congruent lottery choices should not be overshadowed by uncertainty aversion (less attractive inference than estimations task)	Both lotteries, A and B, less attractive when task calls for inference, rather than estimation
3 and 4	RoS implies reversed lottery preferences when P=.2 rather than P=.8	Neither loss aversion nor ratio biases should prevent reversal to A preference when P=.2	No systematic reversal found, reflecting ratio bias dominating RoS
1	RoS applicable when priors guarantee high winning rate (*p*≥.5)	Relative advantage of B over A does not reverse for *p*≥.5	Complicating priors seem to overshadow basic sense of RoS
RoS with Noise or Non-Flat Priors			
1 and 2	Noise (due to memory decay of measurement error) is irrelevant to RoS	Neither relative A vs. B preference nor absolute lottery attractiveness should depend on noise	Noise maintains or or even increases B-preference, but renders both A,B less attractive
Deception and Linguistic Pragmatics			
4	Reversal of RoS when deception inverts win and loss information	Lottery preferences should be invariant to truth value reflection after deception	Deceptive high P=.8 anchor overshadows basic sense of RoS

The second section in Table 3, *Immunity of RoS to competing judgment biases*, draws on task settings (in Experiments 1, 3 and 4) when RoS must compete with, and must not be overshadowed by common biases like uncertainty aversion (Ellsberg, 1961), loss aversion (Camerer, 2005), or ratio bias (Denes-Raj, Epstein & Cole, 1995), or when flat-priors are replaced by distinct priors. Because priors and other judgment biases are ubiquitous, interference of other biases or priors with the RoS would strongly compromise the rationality assumption.

Granting that measurement error (noise) is ubiquitous in everyday life, the third table section is devoted to Experiments 1 and 2 on *RoS with Noise or Non-Flat Priors*, based on simulations showing that both high noise and auspicious priors ($p \geq .5$) ought to render both lotteries more attractive. Devalued lotteries under such conditions (due to risk or uncertainty aversion or complexity aversion) would be incompatible with the simulation results. In a final section, Experiment 4 deals with *Deception and Linguistic Pragmatics*, testing whether the RoS is immune to deceptive negations of wins and losses.

Methods

Note that our aim here was to clarify the non-viability of empirical tests of the generic superlative of rationality. For this purpose, it is sufficient and logically appropriate to consider existence proofs of the susceptibility of RoS to various boundary conditions. We neither want to propagate a unique algorithmic model covering all conditions, under which decisions adhere to or violate the normative rules (see Budescu, Weinberg & Wallsten, 1988). Nor do we aim at a representative design to maximize the external validity of RoS reasoning. Both ambitious goals would call for an extended research program way beyond the scope of the present article.

Regarding external validity, it would be necessary to run multiple parametric experiments of the Cartesian product of all possible stimuli, task settings, and boundary conditions varying jointly in a representative design (Brunswik, 1947; Dhami, Hertwig & Hoffrage, 2004). Such a monstrous hyper-design would have to manipulate all relevant boundary conditions like payoff structures, the intricate role of prior knowledge, or presenting P in experienced versus descriptive format (Hertwig, Hogarth & Lejarraga, 2018). We deliberately refrain from such meta-analytic research goals and confine ourselves to presenting existence proofs of violations of rationality.

Basic Materials and Procedures. In a questionnaire, participants were told to "imagine you are offered a choice between two lotteries, A and B. Both lotteries consist of an urn containing blue and green balls (that do not differ in any other aspect). When you blindly grasp into the urn and draw a ball, you win 200 Euro if the ball is blue; you lose 50 Euro if the ball is green." They were then told that prior to drawing from a chosen urn, "you can observe how many blue and green balls are drawn in a sample from both urns."

In the *sample-inference task*, participants read, "In the sample from Lottery A, 10 balls were randomly drawn from the total of 500 balls in the urn, of which 8 were blue and 2 were green. In the sample from Lottery B, 30 balls were randomly drawn from the total of 500 balls in the urn, of which 24 were blue and 6 were green." In the *sample-estimation task*, instead, samples and populations were identical: "In the sample from Lottery A, all 10 balls in the urn were drawn, of which 8 were blue and 2 were green. In the sample from Lottery B, all 30 balls in the urn were drawn, of which 24 were blue and 6 were green".

On either task, participants then rated both lotteries on seven-point scales: "How much would you like to play Lottery A(B)? [Not at all □ □ □ □ □ □ □ very much]". A comparative bipolar rating was included in Experiments 2, 3, and 4: "If you have a choice between Lottery A and Lottery B, which one do you prefer? [Lottery A □ □ □ □ □ □ □ Lottery B]". Modified task settings in specific experimental conditions allowed us to test distinct simulation results.

Task framing. Experiments 1 and 3 included the task framing manipulation (see first section of Table 3). Each participant completed both the *p*-inference and the P-estimation task. Since $\hat{p}$=P for the estimation task (using samples identical to populations), whereas *p*-inferences are lower than P=.8 for all sample sizes, the RoS implies that $\hat{p}$ is less extreme than P for all inferences. Yet, because inferences entail uncertainty whereas estimations only rely on sample assessment, lotteries in the estimation task may be generally more attractive than in the inference task, due to uncertainty aversion. Thus, the RoS-consistent framing × lotteries interaction must compete with and may be overshadowed by a task framing main effect of uncertainty aversion.

Reward-poor lotteries. Whereas the basic task involves two reward-rich lotteries (P=.8), Experiments 3 and 4 also included a reward-poor condition (P=.2), in which samples included 8 green and 2 blue balls (Lottery A) and 24 green and 6 blue balls (Lottery B). Simulation (Table 2) and reflection show that reward-poor lotteries should produce a mirror image of reward-rich lotteries (second section in Table 3). Just as p_B is higher than p_A for P=.8, p_B is lower (more extreme) than p_A for P=.2. In a reward-poor setting A should be more attractive than B.

Note however that reward-rich and reward-poor lotteries differ in another respect, irrelevant to the RoS. Recall that in the basic task, the absolute payoff for wins (+200) is four times as high as for losses (-50); the expected value is $EV_{P=.8} = .8 \cdot 200 + .2 \cdot -50 = 135$. In the reward-poor condition, the negative payoff for a wrong decision is set to -150, reducing the expected value to $EV_{P=.2} = .2 \cdot 200 + .8 \cdot -150 = -80$. Yet, because the payoff difference for wins and losses increases from P=.8 to P=.2, EV considerations cannot explain a reduced A preference at P=.2, which might rather be an elusive result of loss-aversion. Although a lower winning rate could render the entire choice task less attractive for reward-poor tasks, this should not reduce the preference of A over B. If anything, then, the impact of loss aversion might exaggerate the preference for the smaller lottery (A) in the reward-poor compared to the reward-rich condition.

Another non-rational bias that may overshadow the RoS in the reward-poor condition (P=.2) is the well-known ratio bias (Denes-Raj et al., 1995). While the RoS implies a preference for larger samples when P=.8, but a preference for smaller samples when P=.2, a ratio bias predicts a constant preference for n=30 over n=10. For both P=.8 or P=.2, a larger sample contains more winning balls than a smaller sample. Thus, a reward-poor setting in Experiment 3 allows us to examine the immunity of RoS to both loss aversion and ratio biases.

Noisy samples. We tested the impact of noise both within (Experiment 1) and between participants (Experiment 2). In a second stage of Experiment 1, after completing the basic lottery task in the first stage, a modified instruction introduced noise (see third section of Table 3). Participants were told that due to restricted viewing conditions, the color of sampled balls could only be determined at 80% certainty. As the noisy version always followed the basic version, participants were sensitized to the instruction shift (Fischhoff, Slovic & Lichtenstein, 1979). Yet, observing a sample proportion as extreme as P=.8, despite 20% noise, implies a relatively high population *p* (prior to noise intrusion), which should still be more extreme for the larger lottery B than for the smaller lottery A ($\hat{p}$ = .846 vs. .791; see Table 2). Yet, while noise should conserve the attractiveness advantage of B over A, the question is whether uncertainty aversion resulting from noise causes a general decrease in the attractiveness of both lotteries, A and B.

Impact of non-flat priors. In a third stage of Experiment 1, after sensitization to the basic and the noisy lottery task, we introduced a new modification. Explicit priors ($p \geq .5$) saying that "the possibility can be excluded that there are less blue than green balls in the urn" ought to render lotteries more pleasant and less uncertain. Again, non-flat priors did not affect the relative preference of B ($\hat{p}$ =.781) over A ($\hat{p}$ =.760), as Table 1 shows for (P=.80). There is no good reason for non-flat priors to render both lotteries less attractive.

Pragmatics of Deception. Finally, another judgment bias that might overshadow the RoS is numerical anchoring (Harris & Speekenbrink, 2016). In Experiment 4 (last section in Table 3), we operationalized anchoring through deception. Instructions announced that the lottery proprietor inverted wins and losses (to increase the attractiveness of reward-poor lotteries). Thus, a deceptively high winning rate of P=.8 indicated a true winning rate of P=.2; "rational" agents ought to draw equivalent inferences as in a non-deceptive control condition with honest P=.2. To the extent that judgments are biased toward a high anchor of P=.8, a deceptive anchoring might persist. Resisting such an anchoring bias also constitutes a precondition of rational inferences.

Participants. Of 71 participants in Experiment 1, 40 received the sample-inference task and 31 the sample-estimation task. In Experiment 2, 46 participants were in the reward-rich and 44 in the reward-poor condition (*N*= 90). Experiment 3 presented 36 participants with the basic and 41 participants with the noisy (80% certainty) version (N= 77). In Experiment 4, all 35 participants received all conditions. Because all experiments recruited participants from same cohort of the same university pool and used the same lottery tasks presented in the same questionnaire format, the reported findings not only refer to distinct within- or between-participants contrasts within experiments. We also report meta-analytic results pooling across experiments with equivalent modules. In addition to simple t-tests for distinct a-priori contrasts, we provide Rosenthal and Rubin's (2003) $r_{\text{equivalent}}$ as a simple and straightforward effect-size indicator. All data are available under https://heibox.uni-heidelberg.de/d/4b0e1ff7d0a5438ba55b/.

Results and Discussion

Basic RoS Understanding

Let us first consider the participants' understanding of the basic RoS implication concerning the same high winning proportion P=.8 observed in a small sample (Lottery A) and in a larger sample (Lottery B). Recall that the RoS implies a superior winning probability p of the larger lottery, but only for an inference task, not for a sample estimation task.

Inference task. Across all experiments (1, 2, and 3, see upper-most section in Table 3) that included the basic inference task in a reward-rich condition (P=.8), the large-sample Lottery B indeed received higher ratings, M_B=4.60 on a scale from 1 to 7, SD=1.69, than the small-sample Lottery A, M_A=4.06, SD=1.75, $t(126) = 4.77$, $p < .001$, $r_{\text{equiv}} = .394$.

The same ordinal preference of B over A was observed separately in Experiment 1, M_B=4.75, SD=1.53, versus M_A=4.32; SD=1.56, $t(39)$=1.88, p=.068, r_{equiv} = .295; in Experiment 2, M_B=4.85, SD=1.64, versus M_A=4.37; SD=1.74), $t(40)$=2.79, p=.008, r_{equiv} = .412, and in Experiment 3, M_B=4.24, SD=1.84, versus M_A=3.56; SD=1.85, $t(45)$=3.66, p<.001, r_{equiv} = .488.

Estimation task. Also consistent with the RoS logic (Table 3), the superiority of Lottery B over A was not evident in sample estimations of p=P. Pooling across experiments, the respective statistics were M_B =4.95, SD=1.65, for Lottery B and M_A =5.05, SD=1.57, for Lottery A, $t(76)$=-

0.80, p=.426, r_{equiv} = .092. Again, the same pattern held within singular experiments. Lottery B was neither more attractive than Lottery A in Experiment 1 (M_B =5.23, SD=1.71 vs. M_A =5.42; SD=1.45, t(30)=-0.67, p=.508, r_{equiv} = .126, nor in Experiment 3 (M_B=4.76; SD=1.59 vs. M_A=4.80; SD=1.61;, t(45)=0.42, p=.676, r_{equiv} = .064 (see top-row charts in Figure 1).

An intuitive sense of the basic RoS prediction seems to be even more apparent in bipolar ratings of Lottery A (low end) versus B (high end) preferences. Pooling across experiments, the mean bipolar rating, $M_{bipolar}$ = 4.79, SD=1.53, clearly exceeds the rating-scale midpoint of 4, t(86) = 4.82, p<.001; r_{equiv} = .466. Bipolar inference and estimation ratings only exceeded the scale midpoint for the inference task, $M_{bipolar}$ = 4.80, SD=1.51, t(45) = 3.60, p<.001; r_{equiv} = .481, but not for the estimation task, $M_{bipolar}$ = 3.83, SD=1.06, t(45) = 1.11, p=.273; r_{equiv} = .167 (see second and third row charts in Figure 1). An inference-estimation contrast was clearly significant, t(45) = 4.83, p<.001; r_{equiv} = .593, reflecting a stronger B preference only for the inference task.

This initial evidence suggests that lottery preferences indeed somehow resemble RoS predictions. The same high prevalence of 80% winning outcomes tended to be valued more if observed in a large than in a small sample. The size of this effect was remarkably strong (r_{equiv} in the range of .30 to .50) despite the tiny differences in RoS predictions. Indeed, limited statistical power and a ceiling effect due to generally high expected values of both lotteries ($EV_A = 200 \cdot .750 – 50 \cdot .250 = 138$; $EV_B = 200 \cdot .781 – 50 \cdot .219 = 145$) may have kept the effect small. The preference of Lottery B over A was however peculiar to an inference task. It disappeared when estimates of samples that equal the population made inferences obsolete.

RoS Immunity to Competing Judgment Biases

Yet, even though lottery preferences pointed in the same direction as the RoS, this need not provide cogent evidence for a deeper understanding that deserves the superlative label "rational". Indeed, a preference for the larger of two lotteries may be the incidental result of an irrational heuristic, such as a mere exposure effect (Zajonc, 2001), a ratio bias (Denes-Raj, Epstein & Cole, 1995), or a superficial statistics lesson saying that larger samples provide more reliable evidence. For a closer check on the participants' genuine understanding of the RoS, let us therefore look at whether lottery ratings continue to follow the RoS under less auspicious conditions, when the RoS has to compete with misleading judgment biases (Section 2 in Table 3).

Inferential uncertainty. A framing main effect (i.e., contrast between inference and estimation), averaging over lotteries, affords a first test of common biases that might overshadow the RoS. Both lotteries, A and B, were more attractive when framed as estimation than as inference. Thus, given the same instruction about a sample (with a proportion P) drawn from a universe (with a probability *p*), the willingness to play both lotteries was higher when the task framing focused on P estimation rather than *p* inference. This unwarranted bias, despite the RoS rationale for divergent P and *p*, was evident in separate analyses of Experiment 1, $M_{\text{A,B estimation}} = 5.32$, SD=1.37, $M_{\text{A,B inference}} = 4.54$, SD=1.37, $t(69)$=2.40, p=.019, r_{equiv} = .281; and Experiment 3, $M_{\text{A,B.estimation}} = 4.78$, SD=1.57; $M_{\text{A,B.inference}} = 3.90$, SD=1.73, $t(45)$=3.95, p<.001, r_{equiv} = .516.[6]

Thus, in line with Ellsberg's (1961) paradox conceived as vagueness aversion (Budescu et al., 1988), the uncertainty of an inference task apparently reduced the lotteries' attractiveness compared to an estimation task. Yet, mastery of the RoS should prevent judges from drawing unwarranted conclusions from the apparent superiority of a sample estimation P (with an expected value of $EV_{A,B|P=.8} = 200 \cdot .80 - 50 \cdot .20 = 150$) over an inference *p* ($EV_{A,B|p(P=.8)} = 200 \cdot .781 - 50 \cdot .219 = 145$). Understanding the RoS means to recognize that P is slightly higher than *p* when the lottery is kept constant. In any case, whether the bias reflects a diffuse ambiguity aversion effect or an unwarranted expected-value calculation, the framing main effect indicates that judges' sense of the RoS was overshadowed by an irrational judgment bias. The willingness to play Lotteries A and B (on average) was clearly higher after an estimation than after an inference task (5.32 vs. 4.54 in Experiment 1 and 4.78 vs. 3.90 in Experiment 3).

---

[6] Pooling was impossible because framing in Experiment 1 varied between but within participants in Experiment 3.

Uncertainty from noise. How about uncertainty aversion elicited through distracting noise? As the RoS does not apply to noisy sampling (Section 3 of Table 3), there are two possible answers. Either judges who follow the RoS simply ignore noise or, as in the simulation underlying Table 2, judges might replace a proportion e of randomly chosen outcomes by binary random values. A glance at Table 2 shows that in this case, the RoS is no longer regressive. Observing an extreme P=.8 in spite of 20% noise implies that the corresponding $\hat{p}$ must have been particularly extreme. For n = 30, or when pooling over n = 30 and n = 10, the best estimate is indeed more extreme than the observed proportion ($\hat{p} > P$).

However, the 2nd chart in top row of Figure 1 shows that when a noise proportion of e=.2 was introduced in Experiment 1, the lottery ratings did not follow either of these two reasonable strategies. Judges neither ignored noise, nor did they provide higher ratings of both lotteries under noise. Rather, compared to the no-noise baseline, both lottery ratings decreased under noise, apparently reflecting a kind of uncertainty aversion induced through noise, $M_{\text{A,B noise}}$=3.81; SD=1.59; compared to $M_{\text{A,B no noise}}$=4.88; SD=1.41, $t(75) = 4.24$, p<.001, r_{equiv} = .689. This held for inferences, $M_{\text{inf noise}}$ = 3.59; SD= 1.61, $M_{\text{inf no noise}}$ = 4.54; SD=1.37; $t(39)$=4.69, p<.001, r_{equiv} = .611, as for estimations, $\text{M}_{\text{est noise}}$ = 4.10, SD= 1.54, $\text{M}_{\text{est no noise}}$ = 5.32, SD=1.37; $t(30)$=7.14, p<.001, r_{equiv} = .803. The RoS does not justify such a noise-dependent downward bias.

While in Experiment 1 a similar preference for B over A was obtained under noise $\text{M}_{\text{A.noise}}$ = 3.45, SD=1.74, $\text{M}_{\text{B.noise}}$ = 4.32, SD=1.56, and under no noise, $\text{M}_{\text{A.no noise}}$ = 3.72, SD=1.66, $\text{M}_{\text{B.no noise}}$ = 4.75, SD=1.53, in Experiment 2 the Lottery B advantage tended to be more pronounced under noise, M_{B}=4.854; SD=1.636 vs. M_{A}=4.366; SD=1.743, $t(40)$=2.79, p=.008, r_{equiv} = .41, than without noise, M_{B}=5.611; SD=1.293 vs. M_{A}=5.333; SD=1.604, $t(35)$=1.20, p=.238, r_{equiv} = .0205.

Bipolar ratings corroborate these results in Experiment 2 (second row charts in Figure 1). B preferences (M_{bipolar}=4.442; SD=1.543 vs. scale midpoint 4), $t(76)$=2.511, p=.014, r_{equiv} = .280, were stronger under noise, $M_{\text{bipolar noise}}$ = 4.780; SD=1.573, r_{equiv} = .458, than without noise, $M_{\text{bipolar no noise}}$ = 4.056, SD=1.43, $t(35)$=0.23, p=.817, r_{equiv} = .040. The difference between noise conditions was significant, $t(75)=2.10$, $p=.039$, $r_{equiv} = .040$. Bipolar ratings correlated strongly with unipolar B-A differences, $r=.70$, testifying to the reliability of judgment data.

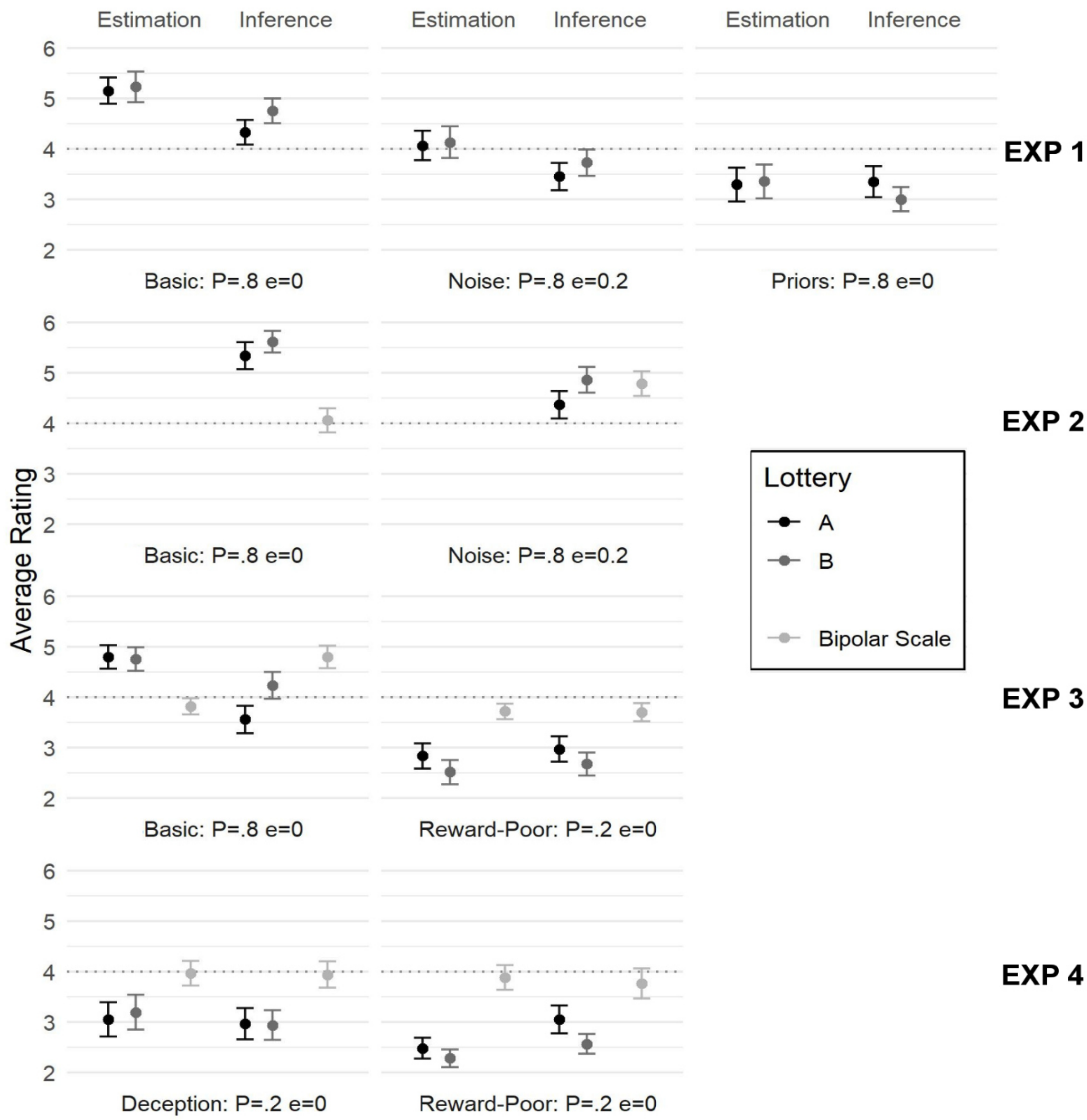


Figure 4.2./1. Mean ratings of Lottery A (n_A = 10) and B (n_B = 30) as a function of Experiments (Exp. 1–4), sample proportions P, noise *e*, priors, and framing (inference vs. estimation). Bipolar ratings > 4 reflect B preferences. Error bars: standard error of the mean.

Although accentuated lottery differences under noise are not surprising theoretically (Costello & Watts, 2019; Kutzner & Fiedler, 2015) and are particularly in line with the simulation results in Tables 1 and 2, lottery devaluation under noise is incompatible with the logic of noise correction, which should render lotteries more attractive if 20% noise did not prevent a sample proportion of P=.8. Instead, noise served to make lotteries less attractive.

Difficulty of mental operations. Analogous to the ambiguity aversion resulting from noise, introducing non-zero priors might render the lottery choice more complicated. Devaluation of lotteries in response to such complications, even when non-zero priors increase the winning chances, might be called a complication aversion bias (Section 3 in Table 3).

Indeed, when explicit priors guaranteed high winning rates ($p \geq .5$) for both lotteries, they were rated less attractive than without priors (see 3rd chart in top row of Figure 1). Despite an auspicious high-winning guarantee, lottery ratings in the inference condition ($M_{\text{A,B priors}}$=3.175, SD = 1.631) decreased markedly relative to the basic condition ($M_{\text{A,B basic}}$= 4.538, SD = 1.370), $t(39)$ = −4.89, p < .001, r_{equiv} = .627. The same held for sample estimates ($M_{\text{A,B}}$=3.323 vs. 5.323; SD = 1.744 vs. 1.370), $t(30)$ = −6.777; r_{equiv} = .788. Again these downward biases cannot be justified on logical grounds (see Table 1).

While $p \geq .5$ priors should render both lotteries more attractive and conserve the advantage of B over A (Table 1), the extra difficulty of complicating priors apparently served to reduce the lotteries' attractiveness, overshadowing the RoS advantage of B over A. Conspicuous difficulties to integrate priors with sample data are hardly compatible with Bayesian rationality.

Reward-rich vs. reward-poor lotteries. A straightforward RoS implication is that the large sample advantage in a reward-rich task setting (P=.8) switches into a large-sample disadvantage in a reward-poor setting (P=.2), as evident from Section 2 in Table 3. However, several well-established biases might work against such a symmetrical RoS effect. Ample research on the ratio bias (Denes-Raj et al., 1995), the denominator neglect (Reyna & Brainerd, 2008), or absolute sample-size biases (Fiedler et al., 2016; Price, Smith & Lench, 2006) predict preferences for the larger sample for both high and low P ratios, because the ratio's numerator receives more weight than the denominator. Once more, an intuitive sense of RoS obtained under simplifying conditions may have been overshadowed by counteracting cognitive illusions.

Recall that the shift from P=.8 to P=.2 in Experiment 3 came along with higher losses for wrong choices (-150 rather than -50). As a consequence of this confound, the comparison of reward-rich (P=.8) and reward-poor lotteries (P=.2) involves a mix of two factors, a marked shift in expected value and unequal RoS sensitivity to high versus low P levels. However, while this confound renders generally lower (unipolar) lottery ratings (averaging across A and B) equivocal, it does not prevent us from analyzing the relative preference for B versus A. Analogous to the Lottery-B advantage at P=.8, the RoS implies a reverse Lottery-A advantage at P=.2. Indeed, the reversal at P=.2 should not be weakened because, if anything, the enhanced win-loss difference (from +200 vs. -50 to +200 vs. -150) increasing the advantage of the lottery with a higher *p*.

Yet, the findings of Experiment 3 did not support the RoS' predicted reversal from a Lottery B advantage at P=.8 to an A advantage at P=.2 (see third-row charts in Figure 1). For convenience, for P=.2, we exchanged ratings of A and B and subtracted seven-point bipolar ratings from 8, making B always the label for the superior lottery.

Whereas Lottery B as usual received better unipolar ratings, M_B = 4.239, *SD*=1.840, than Lottery A, M_A = 3.565, *SD*=1.846, in the P=.8 condition, $t(45) = 3.662$, $p < .001$, r_{equiv} = .488, we did not obtain the predicted reversal for P=.2, M_B = 2.682, *SD*=1.537; M_A = 2.977, *SD*=1.649, $t(88) = -1.28$, $p = .204$, r_{equiv} = .-28. The negative sign indicates an unexpected result, pointing in the direction opposite to the RoS.

Bipolar ratings reflect the same asymmetry. Whereas bipolar inference ratings exceeded the scale midpoint of 4 in the P=.8 condition, $M_{bipolar}$ =4.804, *SD* = 1.515, $t(45) = 3.60$, $p < .001$, r_{equiv} = .481, no reversal was obtained for P=.2, $M_{bipolar}$ = 3.71, *SD* = 1.21, $t(43) = -1.62$, $p = .113$, r_{equiv} = -.245. A significant difference between conditions, $t(88) = 3.793$, $p < .001$, r_{equiv} = .379, reflects a systematic RoS neglect for the reward-poor P=.2 setting. Again, Experiment 3 shows that any rudimentary RoS understanding was overshadowed by cognitive biases of the ratio bias type.

Deception and linguistic pragmatics. When a deceptive report of P=.8 actually implies P=.2, lottery inferences ought to support the RoS prediction that in reward-poor tasks, Lottery A should be favored over B. Experiment 4 (see last section in Table 3) was devoted to testing the immunity of RoS reasoning to such vicissitudes. Regardless of a superficial anchoring bias in favor of Lottery B (induced by the deceptive P=.8), participants should prefer Lottery A over B. This was neither the case for unipolar ratings, M_A=2.848, SD=1.578 vs. M_B=2.797, SD=1.462, $t(78) = 0.33$, $p = .741$, $r_{equiv} = 0.04$, nor on a bipolar scale, $M_{bipolar}$=4.063, SD=1.505, $t(78) = 0.37$, $p = .710$, $r_{equiv} = 0.04$ (see bottom-row charts in Figure 1).

Inferences in a non-deceptive P=.2 condition showed a weak and non-significant trend to prefer Lottery A (M_A=3.057, SD=1.626 vs. M_B=2.571, SD=1.170, $t(34) = 1.82$, $p = .078$, $r_{equiv} = 0.30$). A similar trend for estimations compromised this finding, however (M_A =2.486, SD=1.245 vs. M_B=2.286, SD=1.017, $t(34) = 1.49$, $p = .147$, $r_{equiv} = 0.25$).

Ratings in the deception condition, with a misleading P=.8 anchor, reflected a questionable compromise between naïve, anchor-driven RoS use and partial sensitivity to pragmatic negation. Lottery differences in unipolar ratings were negligible after deception, for inferences, M_A=2.971, SD=1.823 vs. M_B=2.943, SD=1.748, $t(34) = 0.09$, $p = .928$, $r_{equiv} = 0.02$, and estimations, M_A=3.057, SD=1.984 vs. M_B=3.200, SD=2.069, $t(34) = 0.50$, $p = .619$, $r_{equiv} = 0.09$.

In bipolar ratings, no lottery preference was found for either task-framing condition, in the non-deceptive, $M_{bipolar\text{-}Est}$ =3.886, SD=1.430, $t(34) = 0.47$, $p = .639$, $r_{equiv} = 0.08$; $M_{bipolar\text{-}Inf}$ =3.771, SD=1.784, $t(34) = 0.76$, $p = .454$, $r_{equiv} = 0.13$) or in the deceptive condition, $M_{bipolar\text{-}Est}$ =3.981, SD=1.424, $t(34) = 0.12$, $p = .906$, $r_{equiv} = 0.02$; $M_{bipolar\text{-}Inf}$ =3.943, SD=1.552, $t(34) = 0.22$, $p = .829$, $r_{equiv} = 0.04$).

A Framing (inference vs. estimation) × Deception (present vs. absent) ANOVA showed that bipolar ratings were also unaffected by Framing, $F(1,102)=0.10$, $p=.749$, $d=0.02$), Deception, $F(1,102)=0.33$, $p=.565$, $d=0.06$, and their interaction, $F(1,102)=0.04$, $p=.848$, $d=0.06$. A three-factorial ANOVA of unipolar ratings, though, yielded a Deception main effect, $F(1,238)=7.69$, $p=.006$, $d=0.35$, reflecting naive enhancement of both lotteries (across inference and estimation) after Deception ($M_{\text{AB-Deception}}=3.348$, $SD=1.814$ vs. $M_{\text{AB-NoDeception}}=3.010$, $SD = 1.522$). Apparently, the elusive P=.8 anchor served to enhance all lotteries. Main effects of Lotteries, $F(1,238)=0.80$, $p=.372$, $d=0.088$, and Framing, $F(1,238)=0.65$, $p=.422$, $d=0.05$, were as negligible as their interaction $F(1,238)=0.51$, $p=.475$, $d=0.11$. The three-way interaction was also negligible, $F(1,238)=0.03$, $p=.858$, $d=0.07$.

Again, the deception manipulation of Experiment 4 corroborates the notion that, although partially sensitive to deception, participants were far away from using the RoS in a sovereign, systematic, and rational manner.

General Discussion

The reported simulations and lottery experiments converge in several plausible conclusions about the subjective understanding of the RoS in particular and about empirical tests of rationality in general.

Regarding the psychophysics of RoS, our experiments showed that under ideal conditions, when stimulus observations sampled from two lottery urns differed in nothing but sample size and no other complicating or distracting influences were present, many participants seemed to have internalized the gist of the rule. They intuitively understood that the same high winning proportion is worth more in a large ($n=30$) than in a smaller ($n=10$) sample. They also seemed to understand that the large-sample advantage only holds for sample-based inferences, but not for estimations of the proportions observed in the sample.

However, while these findings are apparently in line with ordinal implications of Bayesian statistics and of the RoS in particular, they hardly provided cogent evidence for rationality conceived as “thoughtful cognitive inferences that rely on explicable and testable mental models of reality". One need not attribute stronger trust in larger than in smaller samples to profound mastery of rational reasoning. For a more mundane and more realistic explanation, one might simply assume that enhanced trust in larger samples reflects a learned statistical rule or a crude gut feeling that more information is superior, in line mere exposure or a repetition bias (Unkelbach, Koch, Silva & Garcia-Marques, 2019; Zajonc, 2001).

To infer an ultimate mental competence that deserves to be called rationality proper, it would be necessary to demonstrate judges’ sensitivity to the RoS boundary conditions and the ability to abstract from misleading extraneous biases. According to such more demanding rationality standards, our lottery experiments demonstrated that most people did not apply the RoS systematically. Their ratings violated normative standards as soon as a rudimentary understanding of the RoS was overshadowed by fallacious influences. For instance, extra uncertainty resulting from sampling error or measurement error distracted from the RoS, which ought to be unaffected by such distracters. Ambiguity aversion (i.e., devaluation of inferences relative to estimations) was stronger than the accuracy gain through RoS. Likewise, loss aversion (i.e., devaluation of lotteries allowing for negative payoff) also overshadowed RoS effects. Moreover, a conspicuous reluctance to integrate sample inferencing with prior knowledge is fundamentally incompatible with a Bayesian sense of rationality.

On a more fundamental level, one might question the viability of simplifying lottery tasks as experimental models of rational inference. What is commonly considered a major asset of this research tool, stochastic independence, may be simply unrealistic for real-world observations that are hardly ever stochastically independent. Why should serially experienced binary outcomes – correct versus incorrect student responses, positive versus negative social behaviors, cooperation versus competition – in reality be stochastically independent? In other words, why should we uncritically adopt RoS, or any other normative principle, as arbitrary measure of rationality simply because it can be expressed in a mathematical notation, regardless of its unworldly and unrealistic assumptions?

More generally, we believe that any blatant attempt to assess rationality inductively is condemned to be inconclusive. Even an instance of perfect fit between participants' responses and an arbitrarily selected norm must remain equivocal and need not reflect truly cogent and diagnostic evidence. Rationality as an extreme and universal claim is impossible to prove inductively. A single existence proof of a violation is enough to falsify, whereas numerous confirmations cannot logically verify a universal claim. Scientists should thus refrain from useless attempts to operationalize the idealized competence of rationality as incidental performance on arbitrarily selected tasks (see Marcus & Davis, 2013).

Experimental designs with only two ordinal outcomes – stronger inferences from either small or large samples, possibly moderated by another dichotomy, inference versus estimation – are insufficient to establish the validity of a high-resolution rule like the RoS. Beyond merely ordinal predictions concerning the crude comparison of small versus large samples, the RoS calls for precise discrimination of such subtle quantities as $\Delta\hat{p}$ = .250 – .219 = .031 for P = .2 or $\Delta\hat{p}$= .417 – .406 = .011 for P = .4 (assuming n=30 vs. n=10; see Table 1). These subtle differences have to compete with implications of incisive priors – based on solid evidence or on invalid stereotypes – that can change probabilistic inferences by magnitudes. We therefore pose it would be wise to let rationality be what it is, namely, an idealistic criterion for adaptive judgment and decision making, which is unlikely to be ever reached by even the smartest individuals, groups, and organizations.

5. Illusory Correlations and Semiotics

So far, the presented empirical work has illustrated how pragma-linguistics, the cooperative and competitive use of language in a social context, offers a valuable meta-theoretical lens to theorize about judgments and (ir)rationality of choice. This chapter aims to demonstrate that apart from pragma-linguistics, focusing on the signs we use to communicate is another valuable meta-theoretical lens for JDM theorizing. Like light shining through a stained-glass window, meaning is shaped by the structure it passes through repeatedly in the communicative process. At the same time, this structure is also shaped by the meaning we strive to share and the differences that make a difference (Bateson, 1972) in the world. This mutual interdependence makes it worthwhile to investigate the process of meaning-making through signs (*semiosis*) in JDM research, particularly for the formation of judgments.

To correctly infer correlations in the environment, which is to judge, for instance, how likely it is that a stranger in uniform who carries a weapon will behave friendly towards us, or if the clouds in the sky on our hiking trip herald an upcoming storm, can be central to our survival and wellbeing. The prior article has already introduced the concept of illusory correlations (Chapman, 1967), the phenomenon of perceiving an exaggerated or even inexistent relationship between variables such as social group and behavior, events and their consequences, or attributes of stimuli. Prior theoretical accounts of illusory correlations have focused on cognitive processes (e.g., expectancies, weighting of present and absent information) and unequal sample sizes in observation (for an overview, see Fiedler et al., 2022). This chapter outlines a radically different meta-theoretical lens, the semiotic approach. The presented article then demonstrates how the structure of the sign system used to communicate healthiness and tastiness explains a perceived illusory correlation between these variables.

5.1. Constructing Judgments Using Signs

What is a sign? Semiotics, the study of linguistic and non-linguistic signs (from greek “semeion”, sign), provides numerous answers that can be broadly classified into classic, functionalist, and structuralist views (see, e.g., Bechmann, 2018) that this chapter will give a short overview over.

In the most classic understanding of what a sign is, one refers to something (*representamen*) that points towards a referent (*object*). Peirce (1931) divides representamen into icons, indices, and symbols. Icons resemble their object, as the drawing of a pipe signifies a pipe as an icon. Indices are formed from experience with the environment – the ringing of the bell signifies food for Pavlov's dog. Symbols bear an arbitrary relationship to their objects but are placed within a system of shared sign-object relationships such as language – as the words dog and pipe on this page.

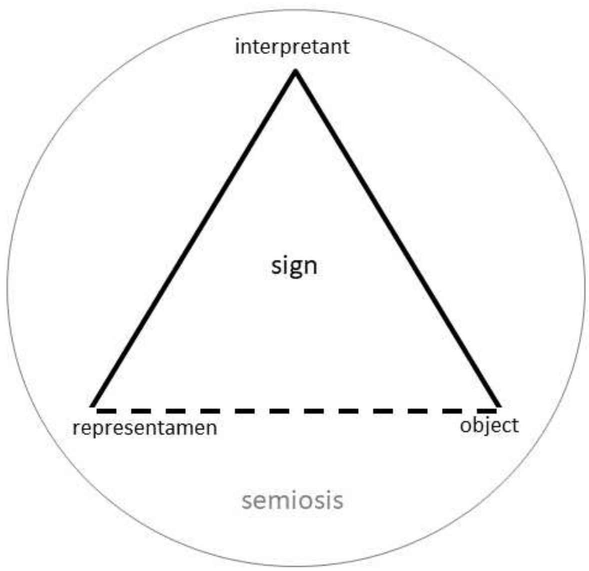


Figure 5.1./1. Peirce’s (1931) semiotic triangle, a triadic sign interpretation model.

Crucially, Peirce added a third concept to form a triangle of semiotic relationships that characterizes a more modern, functionalist understanding of signs (see Figure 5.1./1). To understand semiosis, the active process of meaning-making using signs, Peirce deemed it necessary to consider the representation of the sign-objects relationship and its effect on affect, cognition, and behavior, to use the psychological terms. Such an *interpretant*, which represents that a ringing bell signifies food in one context and will lead to salivation, but the beginning of a break leading us to leave a room in another, cannot be substituted by a lexicon entry. It lies within the individual and their knowledge of the context – which is acquired from culture, but also experience, and therefore both socially situated and, within bounds, individual.

In one of the earliest attempts to transfer the notion of semiosis to social psychology, Bühler (1934) combined Peircean semiotics and Searle's (1969) version of speech act theory. Today, the organon model (which emphasizes a functionalist view, even going so far as loaning the term "organon," tool, from Platon) as the pragma-linguistic part of Bühler's monography is still taught, while his reflections on semiotic foundations, "die Zeichennatur der Sprache," are rarely discussed within social psychology. Indeed, at first, it appears like semiotics barely influenced theorizing in contemporary social psychology, let alone JDM research.

Taking a closer look, there is a theoretical approach central to JDM research that while not concerned with communication or language, still offers a very similar perspective. In the *lens model* of visual perception (Brunswik, 1952), which was immediately extended to the analysis of clinical judgments (Hammond, 1955), objects within the environment are called distal variables. Individuals perceive them through more proximal cues (representamen in semiotic terms) that correlate imperfectly with the distal variable (ecological validity). The cues are combined and weighted (cue utilization) to form a judgment within the individual (what could be considered an interpretant). The article in Chapter 5.2. explains the lens model in more detail.

Both Brunswik's probabilistic functionalism and semiotics share highly similar concepts and have a fundamentally constructivist core: they hold that we do not perceive reality directly but construct our representation of reality mediated by cues (lens model) or signs (semiotics).

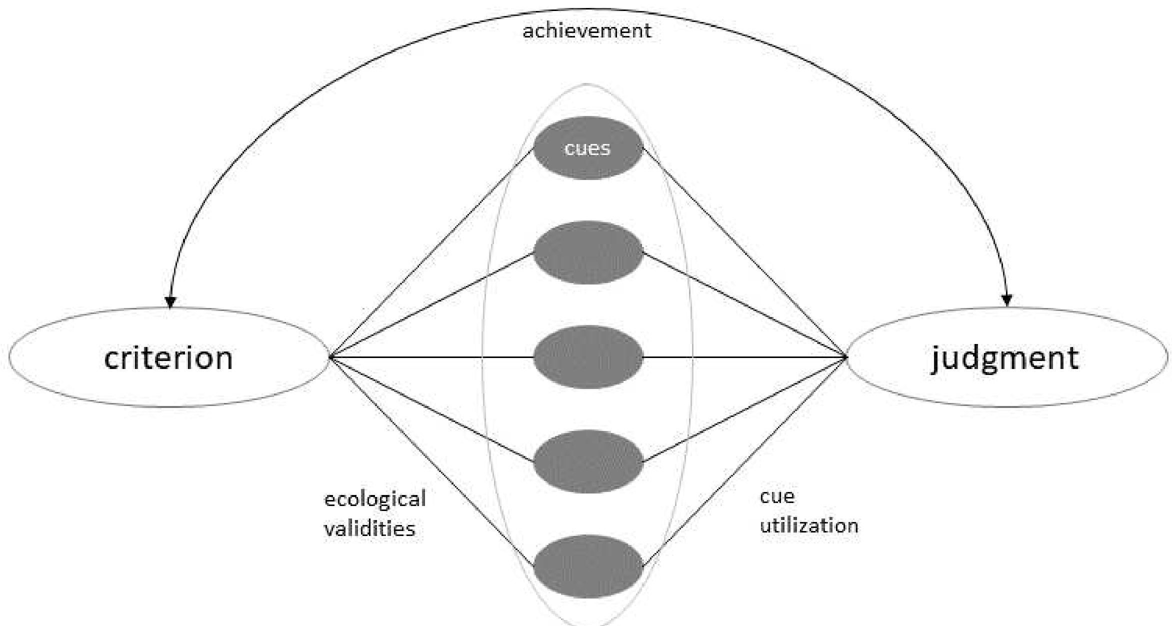


Figure 5.1./2. Brunswik's lens model, in the version by Hammond (1955), for judgments of a distal criterion mediated by proximal cues.

If one follows a very extensive definition of semiotics, including any relationship in which something stands for something else (e.g., Morris, 1946; Eco, 1979), one may even see the lens model as a semiotic model per se. Meanwhile, even if any pointing relationship may be taken as a sign, this does not mean that everything and anything "is" a sign. It needs an individual to take something as a sign (or cue) and engage in the active process of meaning construction that characterizes semiosis and communication.

The classic and functionalist approaches to semiotics mainly aimed at classifying signs according to their relationship to an object within "reality." Modern semiotics frequently abandon this positivist framework and are in closer tune with constructivism. Instead of focusing on single signs and their relation to objects, they investigate the relationship of signs to each other – the *sign system* and its properties (see, e.g., de Saussure, 1967). In this structuralist view, with the use of one sign (e.g., word), the whole system of related signs is implicated – and what is absent but related within the sign system becomes as important for meaning as what is present.

The shift towards a focus on the signs we use for meaning-making is already a far distance from reducing decision-making to individual cognitive processes. The lens model mainly focuses on object-representamen relationships and the achievement of individuals in making judgments that correspond with an "objective reality," which puts it into meta-theoretical proximity with classic and functionalist semiotics. The following article builds on this work but also goes beyond the classic lens model by employing modern, structuralist semiotics as a meta-theory, which leads us to focus on the relationship between representamen. Instead of investigating how the chosen exemplars of distal variables, healthiness and tastiness of food, are represented individually and whether this satisfies a normative standard, we consider them in their joint relation to a web of proximal cues, such as color or types of ingredients.

In this work, we also go beyond the foundations of the lens model by firmly putting verbal communication into the spotlight. We recognize and directly investigate the role of joint meaning-making and language as the primary sign system. Particularly in the employed paradigm, the theoretical focus on communication becomes clearly visible by starting from freely produced descriptions of meals to generate experimental material and then introducing a new form of serial reproduction (Bartlett, 1932). This novel paradigm demonstrates how what is (easily) communicable not only survives but emerges out of the structure of the sign system in the diffusion process of communication. Thus, that the healthiness and tastiness of meals share verbal cues results in a pronounced illusory correlation between them after several rounds of communication.

5.2. From Observation to Social Transmission: How Communication through Cues Shapes the Construction of Perceived Healthiness and Tastiness of Food (Salmen, Haasova, Florack, & Fiedler, submitted)

Perceived healthiness and tastiness constitute key attributes of meals that motivate consumption choices. This paper introduces a novel theoretical perspective, the semiotic approach, to explain a perceived positive relationship between healthiness and tastiness of food, contrary to the prevailing stereotype of a generally negative relationship. The distal concepts of tastiness and healthiness are not amenable to direct perception but must be inferred from proximal cues (e.g., color or freshness). The central tenet of the semiotic approach is that similar cues inform healthiness and tastiness judgments, resulting in a positive illusory correlation. Experiments 1 and 2 established that participants predominantly use shared cues to communicate healthiness and tastiness; distinctive cues that only refer to one concept are hardly found. In Experiment 3, participants completed a serial backtranslation task where they first received a series of food items described by their levels of healthiness and tastiness and then translated them into descriptions using proximal cues. Subsequent participants received these translated descriptions and translated them back into healthiness and tastiness levels. After another cycle of backtranslation, the descriptions of food items that initially contained uncorrelated levels of healthiness and tastiness showed a pronounced positive correlation. This illusory correlation was stronger the more shared rather than distinctive cues participants used for their translations.

What should I eat? This question represents a decision every person makes several times each day. As mundane as these decisions are, food choices are central to the quality of people's lives. For our body, food provides nutrition and sustenance; for our mind, food consumption offers experiences of pleasure and satisfaction. Perceived healthiness, the extent to which we think food is "good for our body," and tastiness, the extent to which we think food will taste pleasant, constitute two key attributes that motivate people's consumption choices (Johansen et al., 2011; Mai et al., 2014; Steptoe et al., 1995). Notably, our beliefs about how these two attributes interlink – whether healthier food would be less tasty or more so – play an essential and unique role in our food choices (Mai & Hoffmann, 2015; Raghunathan et al., 2006; Tuorila & Cardello, 2002).

Typically, consumers cannot directly perceive healthiness, tastiness, and their relationship before consumption but base their judgments on proximal cues. Examples of such cues are how fresh the meal looks or how much fat it contains. Apart from direct experience, healthiness and tastiness judgments often rely on verbal cues communicated, for example, via packaging, advertisement, and personal conversations. Food is a frequent topic of everyday discourse, and recommendations about what to eat can travel far across social networks – increasingly facilitated by social media (e.g., Coates et al., 2019; Pham et al., 2019). For instance, a friend might describe the meals at a new restaurant as "fast, seasonal, family-style favorites" based on another friend's dining experience. How will someone translate this description into judgments of healthiness and tastiness?

The answer reached in the present research is that similar verbal cues convey healthiness and tastiness, resulting in confounded judgments of these two distal concepts. When the cues persons use to infer healthiness overlap considerably with the cues they use to infer tastiness, every observation of tastiness is, to some extent, an observation of healthiness, and vice versa (Haasova & Florack, 2019a). Through this mechanism of cue overlap, uncorrelated or even negatively correlated healthiness and tastiness levels can appear positively related to the perceiver.

This semiotic account can be juxtaposed against traditional accounts that emphasize stereotypical beliefs, schemata, and expectancies as determinants of illusory correlations (e.g., Bartlett, 1932; Kalish et al., 2007; Kashima, 2000; Moussaïd et al., 2015). Such a traditional approach may suggest a negative relationship between perceived healthiness and tastiness, in accordance with an *unhealthy-tasty intuition* (Raghunathan et al., 2006), which presupposes that what is tasty will not be healthy.

In contrast, the semiotic approach presented in this article, which is rooted in Brunswik's (1952) probabilistic functionalism, arrives at an opposite prediction. Due to the hypothesized overlap of healthiness and tastiness cues, we expect a positive relationship to emerge during cue-based inference, as observed in research on the *healthy-tasty illusion* (Haasova & Florack, 2019a; Jo & Lusk, 2018; Kunz et al., 2020a). We expect that on a perception-driven level of describing and choosing a single meal or food item, the influence of cue overlap will be stronger than the stereotypical expectations. Thus, a perceived positive relationship between healthiness and tastiness can be observed overall.

Perceiving and Communicating Healthiness and Tastiness: The Semiotic Approach

To explain the perceived relationship between healthiness and tastiness, it is necessary to understand the psychological process that underlies the cognitive representation and the verbal communication of tastiness and healthiness impressions. To this aim, we adopt Brunswik's (1952) lens model as a conceptual framework (see also Orquin, 2014). We already introduced cues, which are perceptible manifestations of concepts inaccessible via direct perception (distal constructs), as a theoretical framework. Frequently, healthiness and tastiness are distal constructs, as neither of these characteristics is amenable to direct sensory perception before making food choices.

One assumption of the lens-model framework essential to fully understanding our semiotic theory approach is *vicarious functioning*. Not all valid cues need to be available for a specific food item or meal choice; some cues may be missing or substituted by other available cues. This cue substitution principle is a blessing feature of the semiotic environment. To infer a distal entity under uncertainty, it is not necessary to identify the one and only set of valid cues. Adaptive agents make do with any available cues. A cue like fiber content may be readily available when given on the packaging but unattainable for a restaurant meal. Thus, it is sufficient to consider available cues in the judgment context for a comprehensive theory of real-life judgments.

Note also that the cues that determine tastiness and healthiness judgments need not be objectively valid; cues are subjective weights given by the perceiver when construing tastiness and healthiness (i.e., in Brunswik's terms, the cue utilization coefficients). Although laypeople's healthiness judgments are often quite well-calibrated (e.g., Perkovic et al., 2022), not all cues that participants use may conform to objective standards.

As our current topic's fundamental conundrum is not healthiness and tastiness judgments in isolation but their perceived relationship, we do not investigate cues for healthiness and tastiness separately. Our semiotic approach relies on the structural properties of a sign system's joint implications for healthiness and tastiness inferences. When two constructs share perceptual cues – analogous to two personality scales, say, extraversion and leadership, sharing the same questionnaire items (Shweder, 1977) – inferences of both constructs are inextricably confounded (Fiedler et al., 2008; Plessner et al., 2000). We use the term *cue overlap* to denote the semiotic confound of constructs inferred from *shared cues* (Burke et al., 1993; Fiedler et al., 2008; Plessner et al., 2000). In contrast, it would be necessary to map tastiness and healthiness onto sets of distinctive cues to represent these constructs as uncorrelated. Our approach focuses on the degree to which cue overlap, the presence of shared cues, constrains the independent assessment of tastiness and healthiness and leads to illusory correlations.

Therefore, the first question to be tackled is which cues used to communicate tastiness or healthiness tend to be shared or distinctive. Prior research has identified some shared cues that commonly increase healthiness judgments and decrease tastiness judgments, such as “good” fat, package coloring, or perceived sugar levels (Hallez et al., 2023; Huang & Wu, 2016; Irmak et al., 2011; Mai et al., 2016; Prada et al., 2022; Raghunathan et al., 2006). However, the list of shared cues that affect perceived healthiness and tastiness in the same direction is more extensive, including freshness (Fenko et al., 2009), naturalness (Dubé et al., 2016; Lunardo & Saintives, 2013; Magnusson et al., 2003; Rozin et al., 2004), “organic” labeling (Nadricka et al., 2020), ingredient images (Lancelot Miltgen et al., 2016; Rebollar et al., 2017; Thomas & Capelli, 2018), saturation of packaging color (Kunz et al., 2020a), color variety (König & Renner, 2018, 2019), attractiveness (Ares et al., 2014; Becker et al., 2011; Karnal et al., 2016; Mizutani et al., 2010; Velasco et al., 2014; Visschers et al., 2010), and familiarity (Cavanagh & Forestell, 2013; Underwood & Klein, 2002).

As for distinctive cues, the literature reveals that nutritional labels, which provide exact numerical information about the product, tend to be utilized by consumers to judge only healthiness but not tastiness (Haasova & Florack, 2019b; Kunz et al., 2020b; Wang et al., 2016), while a review of prior research reveals no examples of purely distinctive cues for tastiness. One primary implication of our semiotic approach is that inferences based on the known cues for tastiness and healthiness should induce a positive illusory correlation between perceived tastiness and healthiness of meals, as these cues are predominantly shared.

Healthy-Tasty Illusion or Unhealthy-Tasty Intuition?

While perceived healthiness and tastiness are relevant in their own right, most interest in research and practice is sparked by their perceived relationship. How consumers see the relationship between these food attributes is a major determinant of food choices and health hazards. The issues of overconsumption, unhealthy food consumption, and struggles to adopt a healthier diet depend on the perceived relationship between healthiness and tastiness (e.g., Briers et al., 2020; Mai & Hoffmann, 2015; Raghunathan et al., 2006; Russell et al., 2015).

The scientific literature offers a mixed and equivocal picture of the perceived relationship between tastiness and healthiness. Perhaps the most famous line of research focuses on the *unhealthy-tasty intuition*, the lay belief that healthier food is less tasty, and tastier food is less healthy (Hallez et al., 2023; Huang & Wu, 2016; Mai & Hoffmann, 2015; Prada et al., 2022; Raghunathan et al., 2006; van der Heijden et al., 2020; Werle et al., 2013). This intuition may be at the root of the notorious self-control dilemma, a conflict between the short-term hedonic goal of tasty eating and the long-term goal of healthy nutrition (e.g., Fishbach & Zhang, 2008). The goal conflict model of eating (Stroebe et al., 2013), which extends this perspective, is commonly applied to examine restrained eaters. Consumers who assume a negative relationship between healthiness and tastiness are more likely to show unhealthy eating patterns; a positive relationship may support more healthy food choices (e.g., Briers et al., 2020; Mai & Hoffmann, 2015; Raghunathan et al., 2006; Russell et al., 2015).

In contrast to this unhealthy-tasty intuition, we refer to a positive correlation between healthiness and tastiness judgments as the *healthy-tasty illusion* (Haasova & Florack, 2019a; Jo & Lusk, 2018; Kunz et al., 2020a). Research on the unhealthy-tasty intuition and the healthy-tasty illusion represents two different lines of theorizing on how consumers construct their healthiness and tastiness judgments. In studies that find a healthy-tasty illusion (Haasova & Florack, 2019a; Jo & Lusk, 2018; Kunz et al., 2020a), participants encounter a wide range of specific exemplars of foods from many categories, usually without additional ascribed labels, contexts, or highlighted attributes, often in formats such as pictures and descriptions containing a plethora of freely chosen cues to healthiness and tastiness. In such a task, when consumers choose a specific meal or a food product and infer its attributes, they will use all cues available through perception and communication rather than sticking to a simplifying stereotype that may emphasize a negative relationship. According to our semiotic approach, a positive illusory correlation, the healthy-tasty illusion, should emerge in such a task to the extent that the cues used are shared rather than distinctive.

Aims of the Present Research

In the present research, we aim to show that due to the hypothesized overlap of healthiness and tastiness cues, on a cue-driven level of describing and choosing a single meal or food item, people perceive and communicate a positive relationship between healthiness and tastiness. As already outlined, the cue overlap of typical food descriptions should generally be high, inducing a positive illusory relationship between healthiness and tastiness ratings of food items or meals. Nevertheless, because the cues identified by prior literature may not be representative but selectively biased towards plausible and easy-to-communicate stereotypes, we started our investigation with an experiment to establish a basic set of shared and distinctive cues used for tastiness and healthiness judgments. We particularly wanted to test the premise that shared cues dominate people's verbal sign system used to describe and infer healthiness and tastiness. Thus, the primary aim of Experiment 1 is to assess commonly used cues to communicate healthiness and tastiness verbally. The cues identified in Experiment 1 form the basis for Experiment 2, which introduces the receiver perspective and addresses how the uncovered cues influence recipients' judgments of tastiness and healthiness. Finally, Experiment 3 combines the perspectives of communicator and receiver to represent the full circle of communication. Here, in a novel serial back-translation paradigm, we let participants pass on meal descriptions, translating from their tastiness and healthiness ratings to cue patterns and back to tastiness and healthiness ratings.

To the extent that Experiment 1 reveals and Experiment 2 confirms that participants communicate tastiness and healthiness primarily through shared cues, back-translation should render their perceived correlation more and more positive. In our major Experiment 3, we investigate this mechanism in a longitudinal serial-reproduction design across two successive cycles (serial back-translation). In this refined experiment, we manipulate the amount and direction of overlap in different cue sets across four experimental conditions, gathering cogent evidence for the crucial assumption that positive illusory correlations should increase with increasing cue overlap. When participants communicate a series of meal descriptions with initially uncorrelated tastiness and healthiness ratings, their use of shared cues with a common direction should produce a positive illusory correlation. Using truly distinctive cues instead should conserve the initial zero correlation. Shared cues with an opposite direction should increase one judgment while decreasing another, thus inducing a negative relationship between perceived healthiness and tastiness.

Notably, Experiment 3 deliberately tests the hypothesis that the positive relationship conveyed by semiotic characteristics will dominate the negative relationship suggested by the unhealthy-tasty intuition. We assume that the perception-driven task of describing a variety of specific meals from many categories invites usage of the available cues rather than reliance on a simplistic stereotype (e.g., Oakes & Slotterback, 2005; Rozin et al., 1996).

Experiment 1: Cues of Healthiness and Tastiness

Which cues do people use to communicate healthiness and tastiness of food? Are these cues *shared* (used to communicate both constructs) or *distinctive* cues (used to communicate only one construct, not the other)? In the first experiment, we focused on how participants communicate the healthiness and tastiness of various meals in a free-response format. This experiment addressed two main objectives. The first objective was to collect spontaneously used verbal cues that allow participants to explain how they judge the tastiness and healthiness of meals and classify these cues as shared or distinctive. The second objective was to assess the relative frequency of shared and distinctive cues in spontaneous communications about meals.

Method

We used R 3.6.1 (R Core Team, 2019) for all quantitative analyses reported in this paper.

Participants. Forty-five participants (33 identified as women, M_{age} = 26.02, age range: 19 – 52) recruited from a participant pool at Heidelberg University completed the survey. No participants were excluded. Participants received a 4€ reimbursement. In all reported experiments, only native-level German speakers could participate. All participants provided informed consent.

Material and Procedure. The questionnaires of all reported experiments were prepared and presented with SoSci Survey (Leiner, 2020). Fifty-nine pictures that advertised canteen meals, offering savory and sweet options (GMS Gourmet GmbH), served as stimuli. After a short introduction to the task, each participant rated and described ten meals randomly drawn from the pool of 59 pictorial stimuli. Each meal appeared on a new page with questions about healthiness and tastiness. Participants rated healthiness ("How healthy does this meal look?", *extremely unhealthy* to *extremely healthy*; 100-point horizontal slider). They then provided cues in an open response format, prompted by the question, "Why do you rate the healthiness of this meal like this? Which features of the meal would you use to explain your rating to another person?" Further instructions clarified the task: "It is possible that the same feature indicates both healthiness and tastiness. You are allowed to repeat your answers. However, also try to name features that only indicate healthiness." Tastiness judgments ("How tasty does this meal look?", *extremely disgusting* to *extremely tasty*) and cues ("Why do you rate the tastiness of this meal like this? Which features of the meal would you use to explain your rating to another person?") followed the same format. Whether questions about healthiness or tastiness appeared first on the page was randomized across meals. Afterward, participants indicated how they perceived the relationship between healthiness and tastiness by answering the questions "How would the meal's healthiness change if it were made tastier" (*a lot less healthy* to *a lot healthier*, 100-point slider) and "How would the meal's tastiness change if it were made healthier" (*a lot less tasty* to *a lot tastier*, 100-point horizontal slider). Responses to these questions were recoded to a scale from -50 (*a lot less healthy [tasty]*) to 50 (*a lot more healthy [tasty]*). The survey concluded with demographic questions. Participation took 22 minutes on average (range 12 to 34).

Results and Discussion

Content Analysis. Two independent coders unaware of the research objectives were introduced to the concept of cues and received examples of healthiness and tastiness cues from prior literature (e.g., colorful, fresh). An initial screening of the open responses revealed that in this domain, most cues do not take the form of a present or absent feature but graded dimension values (a feature present to a varying degree). For instance, "greasy" frequently appeared in different sentence contexts (e.g., "not at all greasy," "too greasy," "good amount of greasiness") to express varying degrees of healthiness and tastiness. Therefore, we decided to code responses like "looks very greasy" and "does not look greasy" as one cue variable ("The meal is [quantifier] greasy."), with different verbal quantifiers (How much? To what extent?) indicating healthiness and tastiness levels.

Altogether, in 1,768 valid responses (aggregated over participants and items), we found 32 different cue variables that were used to describe the tastiness and healthiness of meals (see Table 1). Four-hundred-five responses could not be coded (see supplementary material).

Classification of shared and distinctive cues. For each of the 32 cue variables identified in the content analysis, we first decided whether the cue indicates only healthiness (at least 95% of responses made to justify healthiness judgments, at most 5% made to justify tastiness judgments), only tastiness, or both. We found only thirteen distinctive cues, with six items classified as distinctive cues for tastiness and seven for healthiness. The remaining 19 cues were shared between tastiness and healthiness. Most frequently used cues were shared (nine out of 15 statements used 50 times or more, see Appendix A). Overall, the preponderance of shared cues reflected a built-in property of the verbal sign system, cue overlap, which makes similar healthiness and tastiness levels easier to communicate than divergent levels.

Table 1: Cue Variables with Adverbial Quantifiers (Left) and Noun Quantifiers (Right) Used to Describe Levels of Healthiness and Tastiness in Experiment 1

The meal is [not at all, a little, averagely, very, extremely] …	The meal contains [no, very few, an average amount of, many, extremely many] …
… greasy.	… vitamins.
… balanced.	… nutrients.
… pleasant texture.	… sugar.
… appetizing.	… processed ingredients.
… pleasant color.	… well-matched ingredients.
… well-seasoned.	… carbohydrates.
… fresh.	… proteins.
… juicy.	… different ingredients.
… aromatic.	… calories.
… salty.	… healthy fats.
… filling.	… vegetables.
… hearty.	… ready-made products.
… spicy.	… high-quality ingredients.
… pleasant temperature.	… water.
… hard to digest.	… meat.
… crispy.	… grilled ingredients.

Correlation of healthiness and tastiness ratings for individual meals. Across all participants and stimuli, healthiness and tastiness ratings correlated positively, $r(453) = .48$, *95%, CI* = .40 – .54, $t(453) = 11.50$, $p < .001$, $d = 0.52$. Moreover, changes in one aspect aligned with corresponding changes in the other. When participants imagined that the meal was made tastier, they expected it to become healthier, $M = 9.01$, $SD = 20.92$, $t(454) = 9.19$, $p < .001$, $d = .43$. Similarly, when imagining a meal becoming healthier, they expected it to become tastier, $M = 12.03$, $SD = 20.40$, $t(454) = 12.58$, $p < .001$, $d = .59$. Both measures converged in implying a positive relationship between food item's healthiness and tastiness ratings, in accordance with a tasty-healthy illusion (Haasova & Florack, 2019a; Jo & Lusk, 2018; Kunz et al., 2020a)[7].

[7] Although the true correlation was not controlled in Experiment 1, this will be the case in the main Experiment 3 below.

To summarize, this preliminary experiment revealed 32 cue variables that can be combined with different verbal quantifiers. Most of these verbal cues, particularly the most frequently used ones, were shared cues with similar implications for healthiness and tastiness. This preponderance of shared cues suggests that language as a semiotic system renders a positive correlation between distal attributes easier to communicate than a negative correlation.

Experiment 2: Cue Utilization

Experiment 1 represented the communicator's perspective: someone sees a meal, forms a judgment, and communicates it to others using verbal cues. Experiment 2 takes a recipient perspective, by complement. When communication partners receive a cue-based description, the question is whether they can translate it back into accurate levels of healthiness and tastiness. Specifically, Experiment 2 explores the impact of communicated cues on receivers' healthiness and tastiness judgments, aiming to validate the classification of cues identified in Experiment 1 as either distinctive or shared. We also sought to characterize the relationship of cues to concepts further. Shared cues may either influence healthiness and tastiness in the same direction or affect these judgments in opposite directions (implying a high judgment of one but a low judgment of the other distal concept).

This step is pivotal to achieving the overall aim of this series of studies. Shared cues with a common direction should force a positive correlation between perceived healthiness and tastiness. Distinctive cues should retain the original relationship. Shared cues with opposite directions should produce a negative correlation. In Experiment 3, we put this prediction to a rigorous test, which is only possible by achieving the objectives of Experiment 2: extracting cues with known semiotic properties (distinctive *vs*. shared; common *vs*. opposite direction).

Method

Participants. After excluding ten participants who failed an initial instruction check, as Abbey and Meloy (2017) suggested, the final sample consisted of 62 participants (46 identified as women, M_{age} = 20.53, age range: 18 – 26). They received course credit for the experiment conducted in a computer lab at the University of Vienna.

Material and Procedure. We supplemented the 32 cue variables extracted in Experiment 1 with four more cue variables from other prior investigations of healthiness and tastiness (Luomala et al., 2015; Roininen et al., 1999, 2001). Participants were randomly assigned to two subsets of 18 cues; n = 34 received subset 1, and n = 28 received subset 2 (see Appendix A). As an attention check, we added the items "The meal is [quantifier] healthy" and "The meal is [quantifier] tasty" (e.g., after "This meal is extremely healthy," they were expected to choose the option "This statement tells me that the meal is healthy").

After providing informed consent and reading a short instruction, participants responded to the 20 items (18 cue variables and two attention checks) in random order, each presented on a new page. Participants received each cue variable, including the attention check, with the instruction, "Somebody describes their food with the following statement: ". The cue variable appeared three times on the same page, first with the quantifier "not at all" or "none," then "average" or "average amount of," and then "extremely" or "many." Each cue-quantifier-combination was followed by "This statement tells me…" and the choice options (1) *nothing about whether the meal is healthy*, (2) *that the meal is not healthy*, (3) *that the meal is average in healthiness*, and (4) *that the meal is healthy*. The same cue-quantifier-combinations were then repeated with "tasty" instead of "healthy." The order of questions about healthiness and tastiness was randomized. After answering these questions for all 18 cue variables and the two attention check items, participants indicated on two questions which healthy-tasty relationship they stereotypically expected ("In general, how would a meal's tastiness [healthiness] change if it were made healthier [tastier]" (*a lot less tasty [healthy]* to *a lot tastier [healthier]*, 100-point slider).

Because Experiment 3 involved translations in both directions, from cue descriptions to judgments of distal attributes and vice versa, we wanted to construct a scale of five graded quantifiers applicable to all cue variables and varying degrees of healthiness and tastiness. A spatial arrangement procedure was used to scale five verbal quantifiers for nouns (e.g., none, many) and five corresponding adverbial quantifiers (e.g., not at all, extremely). Participants could freely drag five rectangular text boxes containing the quantifiers and place them on a horizontal line graph so that the box positions on the line reflected the perceived distances between verbal quantifiers (see Hout et al., 2013). After completing this task, participants provided demographic information. The entire session took 25 (14 – 38) minutes on average.

Results and Discussion

Healthiness and tastiness judgments for cue-quantifier combinations. We analyzed to what extent the participants inferred healthiness and tastiness from each combination of cue variable and quantifier (e.g., "extremely greasy"). We operationalized this as the percentage of participants who chose a response other than "no relationship." For instance, we computed the percentage of all participants who saw "extremely greasy" and chose the response "tells me that the meal is not healthy," "tells me that the meal is average in healthiness, "or "tells me that the meal is healthy." We computed the same cue utilization measure for tastiness. Using these two coefficients, we then classified the cue variables as shared or distinctive.

None of the 36 cue variables were classified as distinctive (see Appendix A) using the cut-off value from Experiment 1 (less than five percent of participants use the cue to infer one of the constructs). Thirteen cues were linearly related to both healthiness and tastiness. Increasing the quantifier value (i.e., from "average" to "extremely") implied higher levels of both healthiness and tastiness for nine cue variables (shared ++) and lower levels for two cue variables (shared --). Only two cue variables implied an opposite relationship: one construct increases while the other decreases (shared +-). All remaining cue variables implied a non-linear relation between both concepts. Altogether, participants tend to infer healthiness and tastiness simultaneously from all of the cue variables extracted in Experiment 1, with some cue variables being more distinctive than others but no purely distinctive cues.

Interrater agreement. We analyzed the agreement between respondents concerning the meaning of cue-quantifier combinations, using the percentage of participants who chose the response option selected by the majority. For instance, we computed the percentage of all participants who answered "extremely greasy" with "is unhealthy," the most frequent response for all participants. Participants generally agreed about the level of healthiness that a cue-quantifier combination indicated. On average, across all cue variables, 76.29% of participants chose the level of healthiness the majority chose, SD = 18.69, range = 33 – 100. A similar consensus was obtained for tastiness inferences (M = 77.59%, SD = 14.62, range = 52 – 100; see Appendix A).

Stereotypical relationship of healthiness and tastiness. Despite the clear-cut preponderance of shared cues fostering the healthy-tasty illusion, Experiment 2 also provided some evidence for the unhealthy-tasty intuition (Huang & Wu, 2016; Mai & Hoffmann, 2015; Raghunathan et al., 2006; Werle et al., 2013) when participants were asked about meals in general. Indeed, participants expected a meal that is made tastier to become less healthy, M = -6.87, SD = 18.98, range = -40 – 33, *95%CI* = -12.69 – -2.05, $t(61) = 2.85$, $p = .006$, $d = 0.36$. In the reverse direction, participants did not expect a meal that is made healthier to change in tastiness, M = 0.23, SD = 18.65, range = -31 – 50, *95%CI* = -4.51 – 4.96, $t(61) = 0.10$, $p = .92$, $d = 0.01$. Note that individual participants varied considerably in the extent to which they believed that healthiness and tastiness generally correlate positively, negatively, or at zero level.

Refinement of the quantifier scale. We used the spatial arrangement procedure to scale five noun quantifiers (e.g., none, many) and five corresponding adverbial quantifiers (e.g., not at all, extremely) in order to construct scales of five graded quantifiers applicable not only to all cue variables but also to varying degrees of healthiness and tastiness. The spatial arrangement of the quantifiers was approximately equidistant between neighboring quantifiers in both the set of noun quantifiers and adverbs, with mean distances ranging between 121 px and 139 px for noun quantifiers and 123 px and 137 px for adverbs. Absolute positioning on the visual scale was comparable between noun and adverbial quantifiers, with differences in mean absolute positions ranging between 0.5 and 24 px on a line of 800 px width. All participants positioned the quantifiers in the same order along the line for both noun quantifiers and adverbs.

Experiment 3: Communicating Healthiness and Tastiness

The first two experiments uncovered verbal cue variables that mediate distal inferences about the healthiness and tastiness of meals. We found strong evidence for cue overlap. All cues collected in Experiment 1 resulted in inferences about both constructs for at least some participants in Experiment 2. However, although we found no purely distinctive cues, we can contrast shared cues against relatively more distinctive cues, for which only very few participants made inferences on both constructs.

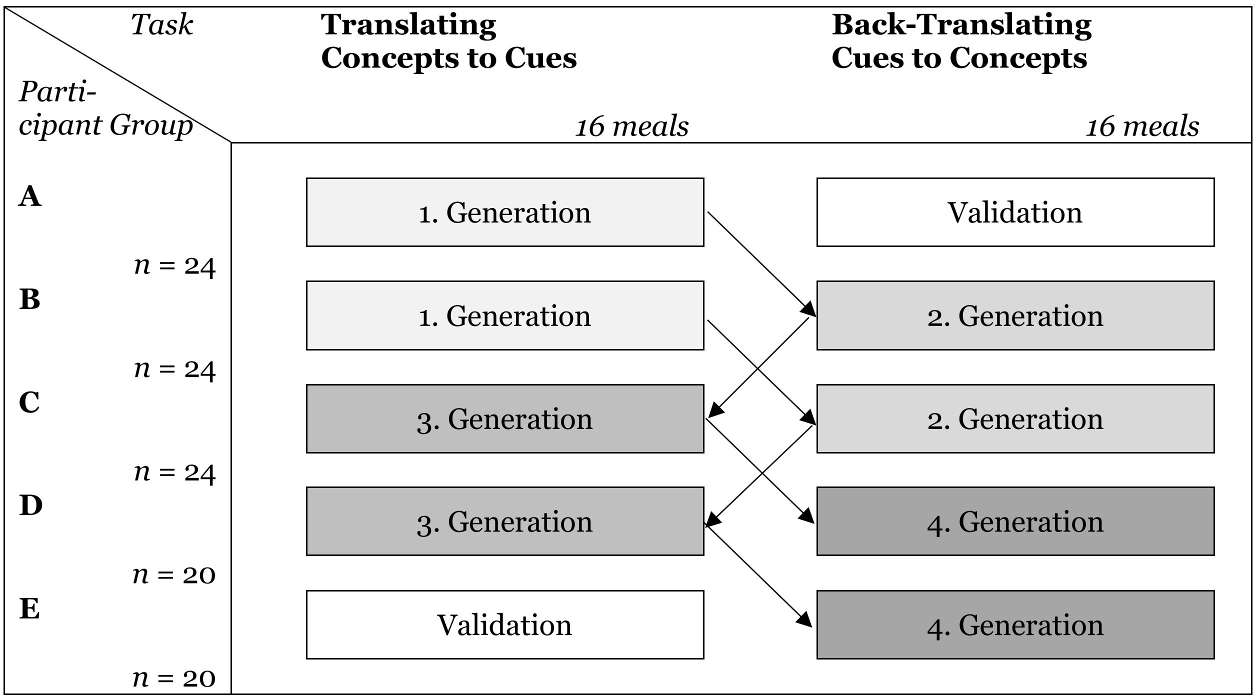
In Experiment 3, we put our predictions from the semiotic approach to a systematic test. Suppose initially uncorrelated healthiness and tastiness judgments of a series of meals are repeatedly communicated. Will the correlation between healthiness and tastiness take on the properties of the linguistic sign system? Will the healthiness-tastiness correlation become increasingly positive if the sign system consists of shared cues with similar implications for both attributes? Conversely, will the correlation be systematically lower when more distinctive cues are used for communication? To find answers to these questions, Experiment 3 consists of a serial back-translation task, where participants translated given healthiness and tastiness levels into cues. Crucially, the available cue sets consisted of only shared cues, increasing numbers of more distinctive cues, or even shared cues with opposite implications for tastiness and healthiness (see Figure 2 below). Other participants then received these cue patterns and back-translated them into healthiness and tastiness levels.

Method

Participants. Of 139 participants, 112 entered the analyses (86 identified as women, M_{age} = 23.02, age range: 18 – 66). The remaining participants were excluded due to failing an attention check in the initial instruction (n = 8), random responding (n = 12), and technical failure (n = 7). Only native-level German speakers could participate. The experiment was the first in several experiments conducted within the same session at an experimental laboratory at Heidelberg University, for a reimbursement of 4€.

Overview. The *serial back-translation* paradigm constitutes a new variant of *serial reproduction* (Bartlett, 1932; Kashima, 2000; Moussaïd et al., 2015), in which a participant receives a message they then tell another participant, who tells it to the next participant and so forth, in a chain of reproductions reminiscent of the children's game Telephone. In this variant, the task for participants changes between generations of social transmission (see Figure 1). In Generation 1 (see Figure 1), participants received descriptions of how healthy and tasty a meal was, but then translated this message into cues (i.e., five cue variables, each with one of five graded quantifiers). In Generation 2, participants received the cue-based message from Generation 1 and back-translated them into levels of healthiness and tastiness. This alternation of tasks is called *back-translation*, a method successfully used to induce and measure semiotic influences in other social cognition domains (Fiedler et al., 2008; Plessner et al., 2000). Within this experiment, the semiotic influence was amplified by repeating the back-translation cycle twice. Generation 3 again translated the concepts into cues, and Generation 4 back-translated these cues into concepts (see Figure 1).

Figure 5.2./1: Shifted Serial Back-Translation Design (Experiment 3). All participants translated healthiness and tastiness (concepts) to cues for 16 meals and back-translated from cues to the concepts for further 16 meals. These translations and back-translations (messages) are passed from one generation to another along the arrows. The validation sets used to check the assignment of cues to concepts are reported in the supplementary material.



Within this serial back-translation paradigm, we gave participants access to different cue sets to do their translations. Across all meals, each participant encountered all four cue sets, while for each individual meal they described, they had access to only one of the four sets. Each set contained five cue variables from Experiment 2 and varied in cue overlap (see Figure 2). The *Overlap Maximum cue set* consists of five shared cues influencing healthiness and tastiness in a common direction (see the upper right part in Figure 2). By comparison, in the *Overlap ++ cue set* (lower right part of Figure 2), only one cue is shared (i.e., nutrients), along with four distinctive cues, two for each concept. The *Separable cue set* (upper left part in Figure 2) also consists of two pairs of distinctive cues, but the fifth cue is neutral (i.e., unrelated to both distal attributes)[8]. In the *Overlap +- cue set* (see lower left part in Figure 2), the fifth cue (crispy) has opposite implications for tastiness (crispy = tasty) and healthiness (crispy = unhealthy).

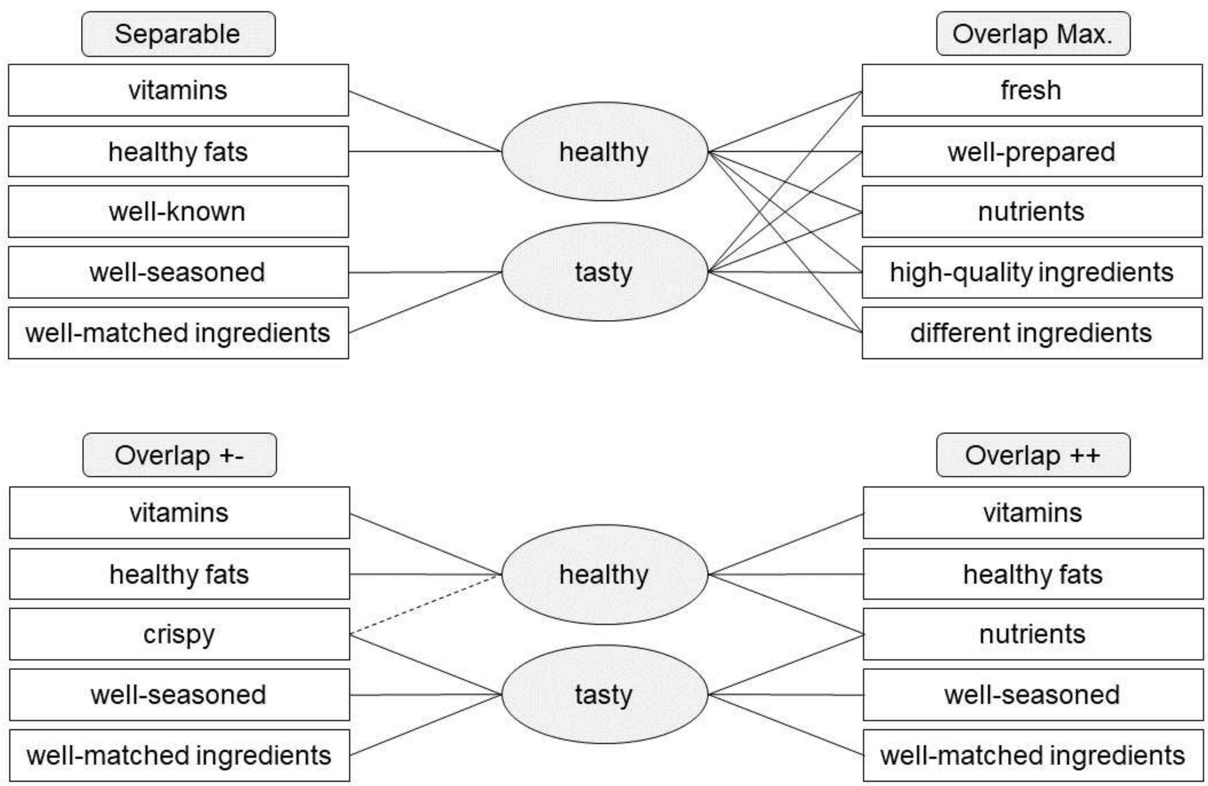
The trajectory of correlations between healthiness and tastiness across generations and cue sets allowed us to assess the process through which communicated information takes on the properties of the semiotic system. In the first generation (first position in the communication chain), tastiness and healthiness levels of the described meals were entirely uncorrelated across meals. The messages consisted of the statements “The meal is … healthy.” and “The meal is … tasty”, with adverbial quantifiers randomly drawn to yield a nonsignificant correlation between healthiness and tastiness across participants and items. We then traced the relationship between tastiness and healthiness levels as meal descriptions passed through four generations of social transmission and four cue sets that varied in cue overlap. According to the cue overlap hypothesis, the Overlap Maximum cue set should produce the strongest (positive) illusory correlation, followed by the Overlap ++ cue set, the Separable cue set, and the Overlap +- cue set.

Design. We manipulated the factors *direction of the task* (translate concepts to cues, back-translate cues to concepts) and available *cue set* (Overlap +-, Separable, Overlap ++, Overlap Maximum; see Figure 2) within participants. The *generation of received messages* (position in the transmission chain) varied within and between participants in the five experimental groups of the shifted serial back-translation design (see Figure 1). For instance, a participant in group B received first-generation messages (initial material) about healthiness and tastiness levels that they translated to cues, and second-generation messages (from group A participants) that consisted of cues they back-translated to healthiness and tastiness levels. The resulting 5×2×4 mixed design included five groups (between-participants) along with two

[8] We pre-tested the neutrality of three candidate cues chosen by the experimenters by asking 18 additional participants (not participating in the main study) if the cue variable told them whether a meal was healthy or tasty.

within-participant factors, the direction of the task and the cue set. Each combination of task and cue set was repeated four times, leading to 32 items per participant.

Figure 5.2./2: Schematic overview of the four cue sets used in Experiment 3. Cues are shown as rectangular boxes, concepts as ovals. Solid lines indicate positive relationships (an increase in cue leads to an increase in concept and vice versa), and the dotted line a negative relationship (an increase in cue leads to a decrease in concept and vice versa).



Procedure. All 32 descriptions of meals were presented in random order. Each new meal was introduced on a new page using the following prompt, "The preceding participants were asked to describe a meal. This is a description from one participant". Below, participants received a message about a meal. For the task of translating concepts to cues, the message read, "The meal is [quantifier] healthy. The meal is [quantifier] tasty.", followed by the instruction:

“Please use the following questions to put together your message for the next participant.” Below, participants saw one of the four cue sets (see Figure 2). For each of the five cue variables in the set, they could choose one of five ascending quantifier values (see Table 1) presented on a horizontal 5-point scale. They also had the option, “I do not want to use this sentence in my message.”

For the task of back-translating cues to concepts, participants received the cue variables with the quantifier values filled in, as the preceding participant had chosen. To pass on this message, they could choose one of five ascending quantifier values (the same as for cue variables) for “The meal is … healthy” and “The meal is … tasty.” They were also allowed to omit each sentence.

When participants had completed this task for one meal, the following page informed them that the message was sent to the next participant and that they would be asked some additional questions. These judgments about the tastiness, healthiness, price, and quality of the meal and whether they would recommend it would not be transmitted to the next participant. Participants then provided these judgments on 100-point horizontal slider scales with labeled endpoints (healthiness: *extremely unhealthy* to *extremely healthy*; tastiness: *disgusting* to *extremely tasty*; quality and price: *extremely low* to *extremely high*; recommendation: *absolutely advise against* to *absolutely recommend*).

Finally, when (back-)translation and subjective judgments of all meals were complete, participants were asked to judge the entire series of 32 items regarding how tasty and healthy the meals were on average, as well as “When a meal was tasty, how healthy was it on average?” (*extremely unhealthy* to *extremely healthy*, 100-point slider) and “When a meal was healthy, how tasty was it on average?” (*disgusting* to *extremely tasty*, 100-point slider).

To complete the experiment, participants provided standard demographical data. The average duration was 27 minutes (range: 23 - 32); a one-minute relaxation break was inserted after the first half of all items.

Analyses. To account for the mixed design, we used the repeated measures correlation coefficient (package rmcorr, function rmcorr) as proposed by Bakdash and Marusich (2017) as well as linear mixed-effects models with the package lme4 (Bates et al., 2015) to analyze messages and judgments. Graphs use ggplot2 (Wickham, 2016). Detailed information on the mixed-effects modeling is available in the supplementary material.

Our investigation focused on the relationship between healthiness and tastiness across generations of serial back-translation. Our design involved messages that either consisted of two quantifier values representing healthiness and tastiness levels (Generation 0, 2, and 4) or consisted of five values using the same quantifiers on cue variables that are a joint scale of healthiness and tastiness (Generation 1 and 3). Thus, only the results of back-translation (Generations 2 and 4 in Figure 1) and the initial material (Generation 0) yielded a correlation r_{ht} between healthiness and tastiness aggregating across participants and meals. For a more fine-grained analysis, we developed a flexible index of the tastiness-healthiness relation that can be calculated for every single message across all generations. It is defined as $interrelation = -\sqrt{\frac{1}{n}\sum(x_i - \bar{x})} + 1$, with n for the number of statements used in the message (two for messages that used concepts, generation 0, 2, and 4, and five for messages that used cues, generation 1 and 3), x the mean quantifier value used in the message and x_i the value of the quantifier chosen for each concept variable or cue variable in the message.

For the current context, this dispersion-based measure ranges from -1 (maximum difference between the values, negative correlation between healthiness and tastiness) to 1 (no difference between the values, positive correlation between healthiness and tastiness). It should be noted that the interrelation score imposes stricter assumptions on the relation of healthiness and tastiness than correlation. It requires not only similar change but similar absolute values to obtain a high interrelation score.

Results and Discussion

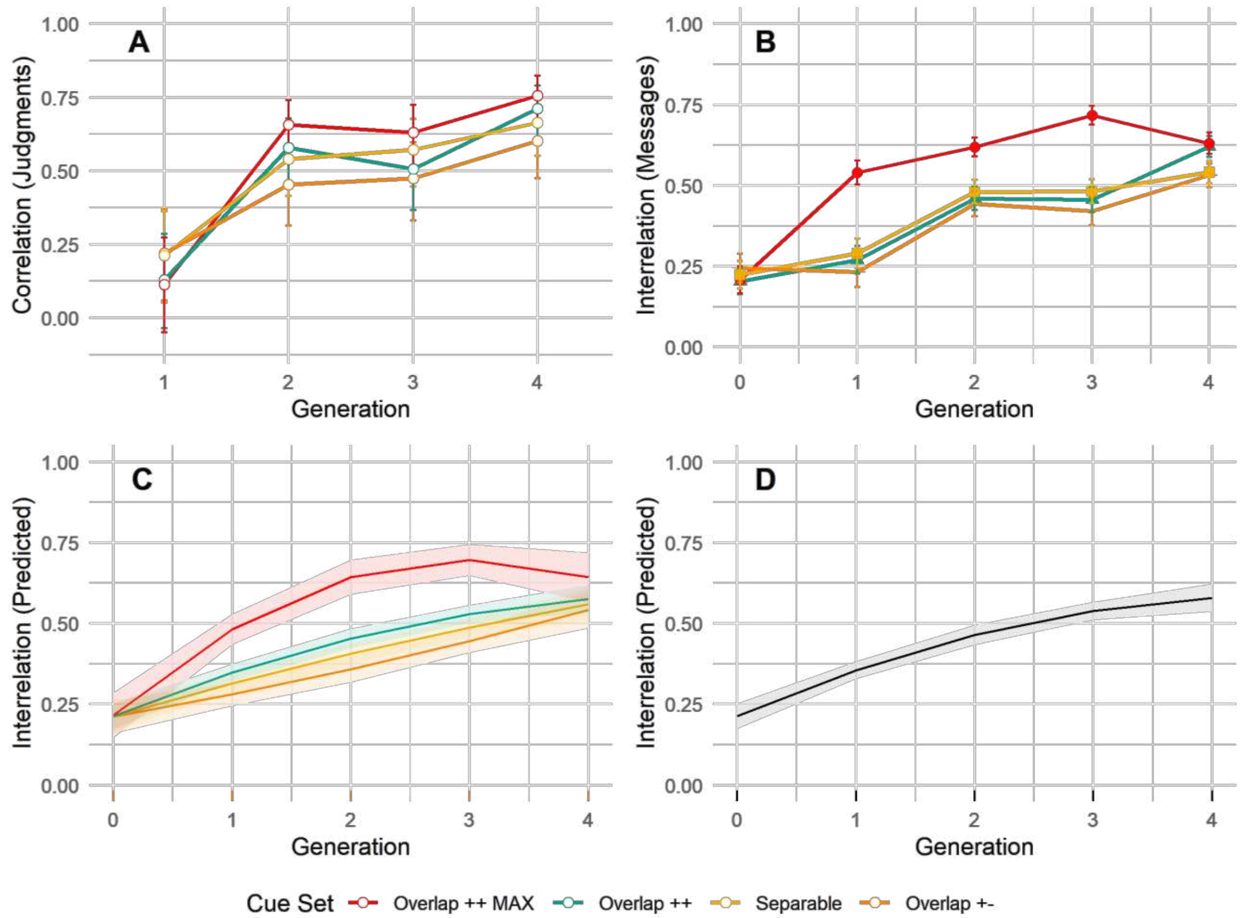
We expected a positive relationship between healthiness and tastiness to emerge and increase across generations of social transmission, with a steeper increase when more shared cues were used compared to more distinctive cues.

Messages. Let us first consider the correlation between healthiness and tastiness levels in the messages, which is only available for generations 0, 2, and 4. On aggregate, across cue-sets, the correlation increased from roughly zero in the initial material (r_{ht} (719) = .05, *95%CI* = -.03 – .12, p = .22) to r_{ht}(698) = .43 (*95%CI* =.37 – .49, $p < .001$) after the first back-translation (Generation 2) to r_{ht}(587) = .58 (*95%CI* =.53 – .64, $p < .001$) in Generation 4.

The interrelation score also reflects this trajectory, as is shown in the upper right chart B in Figure 3. It allows for a much more fine-grained picture, as it is defined for every single message across all generations. Across all participants and cue sets, interrelation increased with generations, b = 7.32, *95%CI* = 6.02 – 8.61, t(322.76) = 11.10, p <.001, d = 14.02, with a decelerating slope in later generations (as apparent in Figure 3, Plot D), b = -0.51, *95%CI* = -1.77 – 0.74, t(682.89) = -0.80, p = .42, d = 0.98.

Most importantly, the strength of the illusory positive correlation depended, as predicted, on the prevalence of shared cues in the cue set: interrelation scores increased faster when more shared cues with a common direction were in the set (b = 0.39, *95%CI* = -0.06 – 0.84, t(3424.15) = 1.70, p = .089, d = 0.75), with a negative curvilinear trend when a high degree of interrelation was reached (see Figure 3, Plot C), b = -0.97, *95%CI* = -1.42 – -0.52, t(3420.23) = -4.21, p <.001, d = 1.86. Thus, an increasing number of shared cues rendered the perceived correlation between healthiness and tastiness more positive. A distinct health-tasty illusion emerged even though these attributes were initially uncorrelated.

Figure 5.2./3: Relationship between tastiness and healthiness in judgments (correlation, A) and messages (interrelation, B) and predictions from the main mixed-effects model between (C) and pooled across cue sets (D) across generations of serial reproduction (initial messages shown as generation 0) and cue sets. Error bars indicate standard error of the mean.



Judgments about individual meals. These findings are mirrored in participants' subjective judgments about the meals, as distinguished from the messages they passed on to other participants. As shown in the upper left chart of Figure 3, the correlation between healthiness and tastiness judgments increased across generations (positions in the transmission chain). From $r(719) = .22$ in the first generation ($p < .001$, range between cue sets $r_{sets} = .15 – .29$), it increases over $r(719) = .55$ ($p < .001$, $r_{sets} = .49 – .60$,) and $r(659) = .57$ ($p < .001$, $r_{sets} =$

.51 – .62) to *r(599)* = .70 (*p* <.001, r_{sets} = .65 – .74) in the last generation. However, as in the messages, there was little change in mean absolute ratings across generations for all general impressions (m_{Tasty} = 54.84 – 58.03, m_{Healthy} = 51.53 – 55.67, m_{Price} = 49.12 – 52.68, m_{Quality} = 51.14 – 53.98, $m_{\text{Recommendation}}$ = 51.95 – 54.72). Thus, the changes affecting the relationship between healthiness and tastiness across generations were symmetrical across low and high values. The healthiness and tastiness ratings correlated highly with judgments of how expensive a meal was, whether it was of high quality, and whether participants would recommend it to others, underscoring the relevance of perceived healthiness and tastiness (see Table 2).

Table 5.2./2: Correlations of item-level judgments pooled across all participants and meals

	Healthy	**Tasty**	**Price**	**Quality**	**Recommendation**
Healthy	1				
Tasty	0.55	1			
Price	0.75	0.62	1		
Quality	0.83	0.72	0.82	1	
Recommendation	0.77	0.83	0.74	0.86	1

Note. Correlations given are repeated measures correlation coefficients. All correlations *p* <.001.

Judgments across all meals. Participants' judgments at the end of the entire sequence of meals again mirrored our findings on messages and judgments of individual meals. After participants described and judged 32 meals, their estimations of the average healthiness and tastiness of all meals were very similar between participant groups, which represent later positions in the transmission chain, for both healthiness (group A: M_A = 51.83, SD_A = 9.73 *vs.* group E: M_E = 55.60, SD_E = 10.93, $t(42)$ = 1.21, p = .23) and tastiness (M_A = 54.92, SD_A = 11.54 *vs.* M_E = 56.20, SD_E = 10.20, $t(42)$ = 0.39, p = .70). More importantly, participants with later positions in the transmission chain perceived a stronger positive relationship between the concepts, with healthiness perceived to predict tastiness (M_A = 54.83, SD_A = 18.21 *vs.* M_E = 68.45, SD_E = 12.83, $t(42) = 2.81$, $p = .01$) and tastiness perceived to predict healthiness (M_A = 48.08, SD_A = 15.92 *vs.* M_E = 58.2, SD_E = 15.89, $t(42) = 2.10$, $p = .04$) in participant groups with a later chain position. Therefore, the emergence of an illusory positive correlation between healthiness and tastiness we reported above for messages and judgments is also perceived by participants across items.

General Discussion

Perceived healthiness and tastiness constitute key attributes of meals that motivate consumption choices. This paper introduced a novel theoretical perspective and methodology, the semiotic approach, to explain how an illusory positive correlation between these attributes, the healthy-tasty illusion, emerges. The presented series of experiments supports both main predictions from this account. First, cues available to communicate healthiness and tastiness verbally are predominantly shared; there is a high amount of cue overlap. Second, communication through cues induces marked positive subjective correlation between the healthiness and tastiness of meals, even for initially uncorrelated material. This illusory correlation emerges stronger and faster the less distinctive the used cues are.

What these findings show is not to-be-expected random noise that any distal perception contains. If participants made mistakes, forgot, or attended to information at random, there would be no reason for a directed and substantial change in the relationship between healthiness and tastiness across described meals. We find an illusory correlation that is already substantial after a single transmission (e.g., one person reading packaging information or asking a friend about their meal), and that increases to a stunning .70 after three more links in the communication chain. The perception of such a strong positive relationship between healthiness and tastiness in meals, while these attributes were originally uncorrelated, clearly shows how cue overlap shapes the perception and communication of food's healthiness and tastiness.

Contrary to the stereotype – prior serial reproduction literature and this research

It is essential to highlight that our findings are congruent with prior observations that perceived healthiness and tastiness are predominantly positively correlated across individual food items (healthy-tasty illusion; Haasova & Florack, 2019a; Jo & Lusk, 2018; Kunz et al., 2020a). Simultaneously, they are incongruent with the stereotype that unhealthy food tastes better than healthy food (unhealthy-tasty intuition, Huang & Wu, 2016; Mai & Hoffmann, 2015; Raghunathan et al., 2006; Werle et al., 2013). While we anticipated the healthy-tasty illusion to emerge from the semiotic perspective, prior theoretical accounts of serial reproduction would have made the opposite prediction. It is a widely held assumption that when individuals retell information, it changes towards their prior schemata, beliefs, and stereotypes (e.g., Bartlett, 1932; Kalish et al., 2007; Kashima, 2000; Moussaïd et al., 2015).

In one of the most influential studies on serial reproduction, Kashima (2000) demonstrated how participants transform a story about a couple towards expected gender norms after repeated retelling. Similarly, Moussaïd and colleagues (2015) found that participants changed messages about risk according to their prior expectancies. The underlying notion that individuals integrate expectancies (priors in Bayesian terminology) with the messages they receive to pass them on and expectancies thus shape the transmitted information is formalized in the Bayesian iterated learning account (Kirby, 2001). In the present case, such a theory would predict a negative relationship between healthiness and tastiness, the *unhealthy-tasty intuition* (Raghunathan et al., 2006), to emerge when information travels through the social network.

In our theoretical account, we clearly distinguish between the role of stereotypes, which suggest inferences about any member of a given category, and the perception of a single meal, person, or other items through cues and particularly verbal descriptions. We do not want to diminish the role of stereotypes – which are profoundly influential, as Kashima (2000) demonstrated so impressively – but point out that whenever individuating information is given through cues, persons will at least supplement their preconceptions to form their judgments. Thus, both the predominant expectancies and the semiotic characteristics of available cues must be considered to form a complete picture.

It may come easy to some readers to reduce our approach to the notion of similarity (as, e.g., Tversky, 1977, presented it), the sharing of attributes between concepts within their mental representation. We expect the cognitive representation to reflect (and uphold) semiotic characteristics like cue overlap. Nevertheless, what our approach highlights, in the spirit of Brunswik, is that what counts for judgment is not which attributes could be known but which cues the environment provides. Consider that we have found, for the same participants, with the same mental representation of healthiness and tastiness, with the same amount of similarity, that the availability of different cues within the conditions of our third experiment profoundly influenced their judgments and the messages they passed on.

It is advisable to distinguish between stereotypical expectancies of concepts' relationships, similarity in cognitive representations of concepts, and semiotic characteristics of cues to the concepts within the environment, such as cue overlap. Nevertheless, they can be expected to align frequently and thus form the powerful distortion effects we can observe in many serial reproduction experiments, which are usually attributed to stereotypes alone.

Constraints on Generality

Given that the materials were validated on samples collected at different universities in two European countries, we expect our results to generalize to other populations of German-speaking adults. Although empirical evidence shows that illusory correlations based on the perceived overlap between constructs are present across languages (e.g., Hamilton & Rose, 1980; Shweder, 1977), the cues and their relation to each concept may differ between languages. While the mechanism of cue overlap, as well as the method of uncovering cues and their semiotic properties, should generalize across languages and samples with different backgrounds, this is not necessarily the case for the specific material and such a strong positive illusory correlation.

It follows from the semiotic approach that through truly distinctive cues, it would be possible to retain the initial uncorrelated relationship between the two constructs. Due to the difficulty of finding genuinely distinctive cues for healthiness and tastiness, we could not sufficiently test this prediction. The constraints of the sign system also prevented us from fully observing the influence of the directionality of shared cues. Theoretically, even forcing a negative relationship between concepts should be possible through shared cues with an opposite direction. However, in our set together with four “distinctive” (mildly shared) cues, the single cue with an opposite direction only slowed the emergence of a positive relationship. Further studies, particularly using samples with different socio-cultural backgrounds and sign systems, may help to shed light on these predictions and the generalizability of our findings.

In addition, participants were forming judgments in rather general situations without the presence of other concrete (self-relevant) goals or decisions. The interplay with other influences on consumer judgments and choices still needs to be investigated and applied to settings outside the laboratory.

How relevant is talking about food?

This research has unequivocally put verbal communication about food’s healthiness and tastiness into focus. How relevant is it for perceiving food’s healthiness and tastiness when no verbal communication is involved? We argue that the present research is not only in theory, through Brunswik’s lens model, connected to direct sensory perception. In this domain, many verbal cues we uncovered directly describe sensory experiences (e.g., colorful, salty, crispy). The grounded theory of desire also supports the close connection between sensory and verbal cues (Papies et al., 2015). It holds that desire, motivation for a specific stimulus such as food, can be sparked by simulations of sensory experiences stored in memory and induced by internal or external, visual, or verbal cues.

Furthermore, tastiness may be open to direct experience upon food consumption through dedicated sensory organs. Nevertheless, additional cues (e.g., sensory or verbal cues) still contribute to the experience (Knöferle & Spence, 2012; Shankar et al., 2009; Spence et al., 2012) – as anybody who tried to enjoy a dish with a congested nose will know. Healthiness, however, is rarely ever open to direct perception, immediate effects of toxicity aside. Sensory cues offer little reliable evidence of healthiness after removing spoiled or moldy food. If the customer cares about the long-term health effects of nutrition, verbal communication may be the only source of reliable information. A recent study by Gandhi and colleagues (2020) showed that words associated with food names show high predictive accuracy (up to 77% for 172 food items) of participants' subjective healthiness judgments – an even better predictor than epistemic knowledge of nutritional values. For example, they found that healthy food items were strongly associated with words related to nature and the cultivation of vegetarian food products (e.g., crop, harvest, and agricultural). This shows again that verbal cues strongly affect people's judgments, even in the presence of interventions and other relevant information. Verbal cues are a unique, effective, and reliable tool to shape subjective healthiness judgments, which in turn influence consumption choices (e.g., König & Renner, 2019). This is clearly visible on the market - many food items now provide cues for healthiness (such as vitamin or fiber content) to aid consumers, and for almost all meals, such cues are readily available online and known to many consumers.

Apart from perceived healthiness and tastiness, the semiotic approach to serial reproduction and the methodology presented in this research is relevant to any other domain in which individuals form judgments about distal constructs through perception and communication. The semiotic approach is as much a theory of perception and communication of food attributes and their relationship as it is a theory of communication and serial reproduction in general. We propose that the sign system is an integral part of the information ecology that shapes cognition based on communication. Therefore, many further areas of psychological research may benefit from a systematic examination of semiotic effects as we have undertaken in the present research.

Implications

Clearly, the ubiquity of shared cues and the pronounced healthy-tasty illusion might enable the consumption of less healthy foods. Tasty items that provide shared cues will appear healthier than they are to consumers, particularly in verbal descriptions that have already been passed on a few times. This might not mean these meals pass as healthy, but maybe just healthy enough for consumption. In addition, consumers often engage in "healthy" image management (Bublitz et al., 2010), a tendency to justify their food choices by boosting subjective healthiness. The abundance of shared cues in the sign system makes this all too easy.

At first sight, this paints a bleak picture for anyone who wants to discourage the consumption of unhealthy food. We find what advertisement and product-development departments have already intuitively understood: it is easy to disguise gummy bears as colorful, fruity, and full of vitamins, thus boosting perceived tastiness and healthiness simultaneously. Shared cues are plentiful, and even if consumers believe that tasty items are unhealthy in general, the influence of cue overlap easily overrules this preconception on the level of single items. As that is where real food choices occur, even following a goal to eat healthy over pursuing tasty food may not be effective.

However, our findings also include some silver linings. The present research shows that we need not fear the unhealthy-tasty intuition when promoting healthy food choices. While labeling food as healthy may work against us as the perceived hedonic quality decreases according to the general stereotype, describing it as fresh, organic, and well-balanced can advertise its health benefits and simultaneously boost perceived tastiness.

This research describes the cues used to communicate the healthiness and tastiness of meals. It shows how cue overlap - that healthiness and tastiness share many cues - leads to a positive illusory correlation between healthiness and tastiness, particularly when meal descriptions are retold. Some parts of this sign system may be deeply rooted in evolution and ecology and are unlikely to change. Nevertheless, the utilization of other cues may shift over time as the cultural and cognitive representations of what signifies healthiness and tastiness change. The predominant study of stereotypes and expectancies must be supplemented with careful examinations of the sign system, both in how communication through cues shapes the construction of the perceived relationship between healthiness and tastiness of food and other areas, such as impression formation, in which perception through cues and communication play a role. As this research demonstrates, the semiotic approach offers the theoretical foundations and methodology to achieve this task.

6. Conclusion

Spanning several phenomena and experimental paradigms central to JDM research, the presented work used the conceptualizations of communication introduced in Chapter 2 as meta-theoretical lenses to inspire theorizing about decision-making. In Chapter 3, seeing the research procedure as communication suggested that participants expect the researcher to communicate cooperatively and give responses corresponding to a hindsight bias due to pragmatic implicatures. However, these implicatures can also be used to de-bias responses, as long as participants' responses are not consolidated into memory. Chapter 4 then focused on competitive uses of communication by subtle means and how they can be used to discourage participants from following a testable, normative mental model for decision-making. Extending beyond the influence of cooperative and competitive pragmatics, Chapter 5 explored how a focus on the structure of the sign system, particularly that concepts may share the same cues, sheds new light on illusory correlations in judgments. In summary, this work furthers a communication perspective within the theoretical landscape of JDM research, in which social cognition, economic theories, and the ecological approach prevail.

Across this thesis, the reader follows a development within the employed conceptualization of communication. It starts with the most basic notion of (imperfect) information transmission from one person to another. Then, a pragmatic perspective adds the importance of context and the cooperative principle of communication. Considering communication as a tool for action shows how speakers use language to further different goals, which also might be competitive and thus to the detriment of the listener. Finally, a semiotic perspective puts signs and the structure of the sign system into focus. Structural semiotics and the language game approach go beyond a naïve positivism that sees "information" as given in either the individual or the environment and take a decidedly constructivist position on joint meaning-making. Indeed, this development also mirrors the discourse in linguistics over time. It is important to note that each of these angles offers a special meta-theoretical focus that can be valuable to JDM research. They are not entirely substitutable, and instead of employing one of them as the (current) best theory of communication, we should ask ourselves which strengths and weaknesses each conceptualization shows for theorizing about a focal phenomenon. However, whichever we choose, they support us in investigating the role of communication processes in decision-making. In the spirit of Bruner (1990), this opens our theorizing to the social, communal nature of our judgments and choices.

The empirical work presented in this thesis exclusively works with verbal communication in a written format. While this is a highly relevant sign system and channel in decision-making, research with a focus on the process of communication needs not (and should not) limit itself to that. Outside the laboratory, decision-making is enmeshed in spoken language and countless paraverbal and nonverbal cues. It remains a task for future research to investigate how these sign systems and their cooperative and competitive pragmatics influence decision-making. While many of the presented meta-theories stem from linguistics, their principles readily apply to nonverbal communication and semiosis.

Considering communication in decision-making outside the laboratory also leads me to note that the presented studies, for enhanced objectivity and experimental control, often reduced communication by participants to a choice paradigm. These designs offered a manageable slice through the fairly chaotic world of meaning-making in the wild as a first test of the presented theories. However, naturalistic communication is much more idiosyncratic, creative and usually defined by the structures of meaning built over larger chunks of texts or conversations. Therefore, a communication approach invites us to explore these theories with a larger variety of methods, for instance, content analysis techniques from computational linguistics to test hypotheses in real-life corpora, or discourse analysis (Potter & Wetherell, 1987).

By necessity, this thesis only presents some particularly striking inspirations from the vast landscape of reflections on communication and some exemplars of how it can enrich JDM theorizing. It leaves aside some fairly widespread concepts, for instance, Watzlawick's (1967) communication axioms or even Freudian iceberg models, that could provide meta-theoretical inspirations for future JDM research. The interested reader might also miss a discussion of another main area of (psycho-)linguistics and social psychology, comparative studies of different language varieties, and the use of native and foreign languages. In the presented work, participants used their native language, either German or English, and while the presented theoretical principles are thought to generalize across languages, that assumption remains to be tested. Going further, this connects the investigation of communication processes within JDM to explorations of sociological and cultural influences with comparative methods (for an overview of comparatist JDM studies, see Li & Kaulius, 2013). Clearly, there are numerous meta-theoretical lenses that future JDM theorizing could explore further.

Behind the desire to uncover mechanisms that produce cognitive illusions and deviations from rationality also lies the aim to aid and improve real-life judgments and decision-making. As the work on the hindsight bias has shown, communication is a valuable tool for action in more than one sense. The competitive force we have shown to hinder goal obtainment in Chapter 4 can also aid rational decision-making and de-bias judgments. It remains to be investigated whether a new set of easily applicable, low-cost interventions based on knowledge about communicative processes in JDM can stand the test of applied scenarios.

This thesis presents communication as one among many useful meta-theoretical lenses in socio-psychological research. It should be noted that communicative processes are often one contributing mechanism among many. The phenomena within JDM research are usually overdetermined. For instance, while we show how pragmatic implicatures might produce a hindsight bias, it can still be expected that motivational and memory processes play an important role in the phenomenon. Within the multi-causal world, communication processes will rarely be the sole cause. Nevertheless, their share in decision-making needs to be investigated and offers new ways of understanding old phenomena.

Bibliography

Chapter 1

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*(5), 752-766. https://doi.org/10.1037/0022-3514.39.5.752

Fischhoff, B., & Broomell, S. B. (2020). Judgment and decision making. *Annual Review of Psychology*, *71*, 331-355. https://doi.org/10.1146/annurev-psych-010419-050747

Gigerenzer, G., & Selten, R. (Eds.). (2002). *Bounded rationality: The adaptive toolbox*. MIT Press. https://doi.org/10.7551/mitpress/1654.001.0001

Hilton, D. J. (1995). The social context of reasoning: Conversational inference and rational judgment. *Psychological Bulletin, 118*(2), 248–271. https://doi.org/10.1037/0033-2909.118.2.248

Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton University Press.

Chapter 2.1.

Austin, J. L. (1962). *How to Do Things with Words: The William James Lectures Delivered in Harvard University in 1955*. Oxford, England: Oxford University Press UK.

Bonaccio, S., & Dalal, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational behavior and human decision processes*, *101*(2), 127-151. https://doi.org/10.1016/j.obhdp.2006.07.001

Chaiken, S., & Trope, Y. (Eds.). (1999). *Dual-process theories in social psychology*. Guilford Press.

Eco, U. (1979). *A theory of semiotics*. Indiana University Press.

Fiedler, K. (2008). Language: A Toolbox for Sharing and Influencing Social Reality. *Perspectives on Psychological Science*, 3(1), 38–47. https://doi.org/10.1111/j.1745-6916.2008.00060.x

Hörmann, H. (1976). *Meinen und Verstehen: Grundzüge einer psychologischen Semantik*. Frankfurt a. M.: Suhrkamp.

Morris, C. (1946). *Signs, language and behavior*. Prentice-Hall.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In *Communication and persuasion* (pp. 1-24). New York: Springer.

Pohl, R. F. (Ed.). (2022). Cognitive illusions: Intriguing phenomena in thinking, judgment, and memory. Routledge.

Saussure, F. de (1964). *Course in general linguistics*. London: Peter Owen.

Schulz-Hardt, S., & Mojzisch, A. (2012). How to achieve synergy in group decision making: Lessons to be learned from the hidden profile paradigm. *European Review of Social Psychology*, *23*(1), 305-343. https://doi.org/10.1080/10463283.2012.744440

Shannon, C. E., & Weaver, W. (1949). The Mathematical Theory of Communication. University of Illinois Press. ISBN 0-252-72548-4.

Trope, Y., & Liberman, N. (2003). Temporal construal. *Psychological Review*, *110*(3), 403–421. https://doi.org/10.1037/0033-295X.110.3.403

Wittgenstein, L. (1953). *Philosophical Investigations*. Blackwell Publishing.

Chapter 2.2.

Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In *Action control* (pp. 11-39). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-69746-3_2

Allport, F. H. (1924). *Social Psychology*. Boston: Houghton Mifflin.

Allport, G. W. (1935). *Attitudes*. In *A Handbook of Social Psychology* (p. 798–844). Clark University Press.

Allport, G. W. (1954). *The nature of prejudice*. Addison-Wesley.

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological bulletin*, *111*(2), 256. https://doi.org/10.1037/0033-2909.111.2.256

Asch, S. E., & Guetzkow, H. (1951). Effects of group pressure upon the modification and distortion of judgments. *Organizational influence processes*, 295-303.

Atkinson, J. W. (1953). The achievement motive and recall of interrupted and completed tasks. *Journal of Experimental Psychology*, *46*(6), 381–390. https://doi.org/10.1037/h0057286

Bandura, A. (1977). *Social learning theory*. Prentice-Hall.

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychological bulletin*, *117*(3), 497. https://doi.org/10.1037/0033-2909.117.3.497

Bartlett, F. C. (1932). *Remembering: a study in experimental and social psychology*. Macmillan.

Berkowitz, L. (1969). *Roots of aggression: A re-examination of the frustration-aggression hypothesis*. New York: Atherton Press.

Billig, M. (1973). Normative communication in a minimal intergroup situation. *European Journal of Social Psychology*, *3*(3), 339–343. https://doi.org/10.1002/ejsp.2420030312

Bischof, N. (1972). The biological foundations of the incest taboo. *Social Science Information/Sur Les Sciences Sociales*, *11*(6), 7–36. https://doi.org/10.1177/053901847201100601

Bonem, M., & Crossman, E. K. (1988). Elucidating the effects of reinforcement magnitude. *Psychological Bulletin, 104*(3), 348–362. https://doi.org/10.1037/0033-2909.104.3.348

Brehm, J. W. (1966). *A theory of psychological reactance*. Academic Press.

Brehm, J. W., & Cohen, A. R. (1962). *Explorations in cognitive dissonance*. John Wiley & Sons Inc. https://doi.org/10.1037/11622-000

Brewer, M. B. (1991). "The social self: On being the same and different at the same time". *Personality and Social Psychology Bulletin*, 17, 475–482. https://doi.org/10.1177/0146167291175001

Bruner, J. S. (1957). On perceptual readiness. *Psychological review*, *64*(2), 123-152. https://doi.org/10.1037/h0043805

Bruner, J. S., & Postman, L. (1948). Symbolic value as an organizing factor in perception. *The Journal of Social Psychology*, *27*, 203–208. https://doi.org/10.1080/00224545.1948.9918925

Brunswik, E. (1952). The conceptual framework of psychology. In O. Neurath, R. Carnap & C. W. Morris (Eds.), *International Encyclopedia of Unified Science, Volume 1, Number 10,* (pp. 567 – 655). University of Chicago Press.

Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological bulletin*, *115*(3), 401-423. https://doi.org/10.1037/0033-2909.115.3.401

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the BIS/BAS scales. *Journal of personality and social psychology*, *67*(2), 319-333. https://doi.org/10.1037/0022-3514.67.2.319

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*(5), 752-766. https://doi.org/10.1037/0022-3514.39.5.752

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

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*(6), 1015-1026. https://doi.org/10.1037/0022-3514.58.6.1015

Crowe, E., & Higgins, E. T. (1997). Regulatory focus and strategic inclinations: Promotion and prevention in decision-making. *Organizational Behavior and Human Decision Processes*, 69, 117-132. https://doi.org/10.1006/obhd.1996.2675

Deci, E. L., & Ryan, R. M. (1987). The support of autonomy and the control of behavior. *Journal of personality and social psychology*, *53*(6), 1024-1037. https://doi.org/10.1037/0022-3514.53.6.1024

Demoulin, S., Pozo, B. C., & Leyens, J.-P. (2009). Infrahumanization: The differential interpretation of primary and secondary emotions. In S. Demoulin, J.-P. Leyens, & J. F. Dovidio (Eds.), *Intergroup misunderstandings: Impact of divergent social realities*. (pp. 153–171). Psychology Press.

Deutsch, M. (1949). A theory of co-operation and competition. *Human relations*, *2*(2), 129-152. https://doi.org/10.1177/001872674900200204

Dweck, C. S. (1999). *Essays in social psychology. Self-theories: Their role in motivation, personality, and development*. Psychology Press.

Eagly, A. H. (1978). Sex differences in influenceability. *Psychological Bulletin*, *85*(1), 86-116. https://doi.org/10.1037/0033-2909.85.1.86

Eiser, J. R. (1971). Enhancement of contrast in the absolute judgment of attitude statements. *Journal of Personality and Social Psychology*, *17*(1), 1–10. https://doi.org/10.1037/h0030455

Eiser, J. R., & Stroebe, W. (1972). *Categorization and social judgment*. Academic Press.

Ellemers, N., Spears, R., & Doosje, B. (2002). Self and social identity. *Annual Review of Psychology*, *53*(1), 161–186. https://doi.org/10.1146/annurev.psych.53.100901.135228

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

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford university press.

Festinger, L. (1962). Cognitive dissonance. *Scientific American*, *207*(4), 93–107. https://doi.org/10.1038/scientificamerican1062-93

Fiedler, K. (2014). From intrapsychic to ecological theories in social psychology: Outlines of a functional theory approach. *European Journal of Social Psychology*, *44*(7), 657–670. https://doi.org/10.1002/ejsp.2069

Fiedler, K., & Bluemke, M. (2005). Faking the IAT: Aided and unaided response control on the Implicit Association Tests. *Basic and Applied Social Psychology*, *27*(4), 307-316. https://doi.org/10.1207/s15324834basp2704_3

Fiedler, K., Jung, J., Wänke, M., Alexopoulos, T., & de Molière, L. (2015). Toward a deeper understanding of the ecological origins of distance construal. *Journal of Experimental Social Psychology*, *57*, 78–86. https://doi.org/10.1016/j.jesp.2014.11.002

Fiedler, K., & Juslin, P. (Eds.). (2006). *Information sampling and adaptive cognition*. Cambridge University Press. https://doi.org/10.1017/cbo9780511614576

Fiedler, K., & Kutzner, F. (2015). Information sampling and reasoning biases: Implications for research in judgment and decision making. In *Blackwell handbook of judgment and decision making: An interdisciplinary perspective*. Blackwell. https://doi.org/10.1002/9781118468333.ch13

Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *Journal of personality and social psychology*, *82*(6), 878-902. https://doi.org/10.1037/0022-3514.82.6.878

Fiske, S. T., & Neuberg, S. (1990). A Continuum of Impression Formation, from Category-Based to Individuating Processes: Influences of Information and Motivation on Attention and Interpretation. *Advances in Experimental Social Psychology*, *23*(C), 1-74. https://doi.org/10.1016/s0065-2601(08)60317-2

Florack, A., Friese, M., & Scarabis, M. (2010). Regulatory focus and reliance on implicit preferences in consumption contexts. *Journal of Consumer Psychology*, *20*(2), 193-204. https://doi.org/10.1016/j.jcps.2010.02.001

Förster, J., Higgins, E. T., & Bianco, A. T. (2003). Speed/accuracy decisions in task performance: Built-in trade-off or separate strategic concerns?. *Organizational behavior and human decision processes*, *90*(1), 148-164. https://doi.org/10.1016/s0749-5978(02)00509-5

Frable, D. E. S. (1997). Gender, racial, ethnic, sexual, and class identities. *Annual Review of Psychology*, *48*, 139–162. https://doi.org/10.1146/annurev.psych.48.1.139

Frey, B. S., & Jegen, R. (2001). Motivation Crowding Theory: A Survey of Empirical Evidence. *Journal of Economic Surveys, 15*(5), 589-611. https://doi.org/10.1111/1467-6419.00150

Friedman, R. S., & Förster, J. (2001). The effects of promotion and prevention cues on creativity. *Journal of Personality and Social Psychology, 81*(6), 1001-1013. https://doi.org/10.1037/0022-3514.81.6.1001

Fujita, K., & Carnevale, J. J. (2012). Transcending temptation through abstraction: The role of construal level in self-control. *Current Directions in Psychological Science*, *21*(4), 248–252. https://doi.org/10.1177/0963721412449169

Gaertner, S. L., Dovidio, J. F., Anastasio, P. A., Bachman, B. A., & Rust, M. C. (1993). The Common Ingroup Identity Model: Recategorization and the reduction of intergroup Bias. *European Review of Social Psychology*, 4, 1-26. https://doi.org/10.1080/14792779343000004

Gigerenzer, G., & Selten, R. (Eds.). (2002). *Bounded rationality: The adaptive toolbox*. MIT Press. https://doi.org/10.7551/mitpress/1654.001.0001

Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of accommodation: Developments in applied sociolinguistics*. (pp. 1–68). Cambridge University Press. https://doi.org/10.1017/CBO9780511663673.001

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist*, 54, 493-503. https://doi.org/10.1037/0003-066x.54.7.493

Gopher, D., & Kimchi, R. (1989). Engineering psychology. *Annual Review of Psychology*, *40*(1), 431-455. https://doi.org/10.1146/annurev.ps.40.020189.002243

Greenberg, J., Pyszczynski, T., & Solomon, S. (1986). The causes and consequences of a need for self-esteem: A terror management theory. In *Public self and private self* (pp. 189-212). New York: Springer. https://doi.org/10.1007/978-1-4613-9564-5_10

Greenwald, A. G. (1968). Cognitive learning, cognitive response to persuasion, and attitude change. In A. Greenwald, T. Brock, & T. Ostrom (Eds.), *Psychological foundations of attitudes* (pp. 148-170). New York: Academic Press. https://doi.org/10.1016/b978-1-4832-3071-9.50012-x

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

Grice, H. P. (1975). Logic and Conversation. In P. Cole & J. Morgan (Eds.) *Syntax and Semantics*. Academic Press. https://doi.org/10.1163/9789004368811_003

Hartwig, M., & Bond, C. F. Jr (2014). Lie detection from multiple cues: A meta-analysis. *Applied Cognitive Psychology*, *28*(5), 661-676. https://doi.org/10.1002/acp.3052

Hardin, C. D., & Higgins, E. T. (1996). Shared reality: How social verification makes the subjective objective. In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition. Handbook of motivation and cognition, Vol. 3. The interpersonal context* (p. 28–84). The Guilford Press.

Haslam, N. (2006). Dehumanization: An Integrative Review. Personality and Social Psychology Review, 10(3), 252–264. https://doi.org/10.1207/s15327957pspr1003_4

Heckhausen, H., & Gollwitzer, P. M. (1987). Thought contents and cognitive functioning in motivational versus volitional states of mind. *Motivation and emotion*, *11*(2), 101-120. http://dx.doi.org/10.1007/BF00992338

Heider, F. (1946). Attitudes and cognitive organization. *The Journal of psychology*, *21*(1), 107-112. https://doi.org/10.1080/00223980.1946.9917275

Heider, F. (1958). *The psychology of interpersonal relations*. New York: John Wiley & Sons. http://dx.doi.org/10.1037/10628-000

Higgins, E. T. (1981). The "communication game": Implications for social cognition and persuasion. In *Social cognition: the Ontario symposium* (Vol. 1, pp. 343-392).

Higgins, E. T. (1998). Promotion and prevention: Regulatory focus as a motivational principle. In M. P. Zanna (Ed.), *Advances in experimental social psychology*, Vol. 30. (pp. 1–46) San Diego, CA: Academic Press. https://doi.org/10.1016/s0065-2601(08)60381-0

Higgins, E. T. (2002). How self-regulation creates distinct values: The case of promotion and prevention decision making. *Journal of Consumer Psychology, 12*, 177–191. https://doi.org/10.1207/s15327663jcp1203_01

Higgins, E., Friedman, R., Harlow, R., Idson, L., Ayduk, O., & Taylor, A. (2001). Achievement orientations from subjective histories of success: Promotion pride versus prevention pride. *European Journal of Social Psychology, 31*(1), 3–23. https://doi.org/10.1002/ejsp.27

Hills, T. T., Todd, P. M., Lazer, D., Redish, A. D., & Couzin, I. D. (2015). Exploration versus exploitation in space, mind, and society. *Trends in Cognitive Sciences*, *19*(1), 46–54. https://doi.org/10.1016/j.tics.2014.10.004

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin, 136*(3), 390–421. https://doi.org/10.1037/a0018916

Hogarth, R. M., & Soyer, E. (2011). Sequentially simulated outcomes: Kind experience versus nontransparent description. *Journal of Experimental Psychology: General*, *140*(3), 434–463. https://doi.org/10.1037/a0023265

Hogg, M. A. (2007). Uncertainty-identity theory. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 39, pp. 69-126). San Diego, CA: Academic Press. https://doi.org/10.1016/S0065-2601(06)39002-8

Hovland, C. I., Janis, I.L., & Kelley, H.H. (1953). *Communication and persuasion*. Yale University Press.

Jost, J. T., & Banaji, M. R. (1994). The role of stereotyping in system-justification and the production of false consciousness. *British journal of social psychology*, *33*(1), 1-27. https://doi.org/10.1111/j.2044-8309.1994.tb01008.x

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American psychologist*, *58*(9), 697-720. https://doi.org/10.1037/0003-066x.58.9.697

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica, 47*(2), 263-291. https://doi.org/10.2307/1914185

Kalish, M. L., Griffiths, T. L., & Lewandowsky, S. (2007). Iterated learning: Intergenerational knowledge transmission reveals inductive biases. *Psychonomic Bulletin & Review*, *14*(2), 288-294. https://doi.org/10.3758/bf03194066

Kashima, Y. (2000). Maintaining cultural stereotypes in the serial reproduction of narratives. *Personality and Social Psychology Bulletin*, *26*(5), 594-604. https://doi.org/10.1177/0146167200267007

Kassin, S. M. (1979). Consensus information, prediction, and causal attribution: A review of the literature and issues. *Journal of Personality and Social Psychology*, *37*(11), 1966–1981. https://doi.org/10.1037/0022-3514.37.11.1966

Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist*, *28*(2), 107–128. https://doi.org/10.1037/h0034225

Kitayama, S., & Miyamoto, Y. (2000). Cultural psychology and macroscopic comparisons between East and West: Significance and empirical findings. *Japanese Psychological Review*, *43*(1), 57–81.

Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, *94*(2), 211–228. https://doi.org/10.1037/0033-295X.94.2.211

Krüger, T., Fiedler, K., Koch, A. S., & Alves, H. (2014). Response category width as a psychophysical manifestation of construal level and distance. *Personality and Social Psychology Bulletin*, *40*(4), 501-512. https://doi.org/10.1177/0146167213517009

Kruglanski, A. W. (1990). Lay epistemic theory in social-cognitive psychology. *Psychological Inquiry*, *1*(3), 181-197. https://doi.org/10.1207/s15327965pli0103_1

Kruglanski, A. W. (2004). *The psychology of closed mindedness*. New York: Psychology Press.

Kruglanski, A. W., & Freund, T. (1983). The freezing and unfreezing of lay-inferences: Effects on impressional primacy, ethnic stereotyping, and numerical anchoring. *Journal of experimental social psychology*, *19*(5), 448-468. https://doi.org/10.1016/0022-1031(83)90022-7

Kruglanski, A. W., & Thompson, E. P. (1999). Persuasion by a single route: A view from the unimodel. *Psychological Inquiry*, *10*(2), 83-109. https://doi.org/10.1207/s15327965pl100201

Kutzner, F., & Fiedler, K. (2017). Stereotypes as pseudocontingencies. *European Review of Social Psychology*, *28*(1), 1-49. https://doi.org/10.1080/10463283.2016.1260238

Lawrence, D. H., & Festinger, L. (1962). *Deterrents and reinforcement: The psychology of insufficient reward*. Stanford Univ. Press.

Lermer, E., Streicher, B., Sachs, R., Raue, M., & Frey, D. (2015). The effect of construal level on risk-taking. *European Journal of Social Psychology*, *45*(1), 99–109. https://doi.org/10.1002/ejsp.2067

Lewin, K. (1943). Psychology and the process of group living. *The Journal of Social Psychology, 17,* 113–131. https://doi.org/10.1080/00224545.1943.9712269

Lewin, K. (1947). Frontiers in Group Dynamics: Concept, Method and Reality in Social Science; Social Equilibria and Social Change. *Human Relations, 1*(1), 5–41. https://doi.org/10.1177/001872674700100103

Lewin, K. (1951). *Field theory in social science: selected theoretical papers (Edited by Dorwin Cartwright.)*. Harpers.

Liberman, N., Sagristano, M. D., & Trope, Y. (2002). The effect of temporal distance on level of mental construal. *Journal of Experimental Social Psychology*, *38*(6), 523–534. https://doi.org/10.1016/S0022-1031(02)00535-8

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting & task performance*. Prentice-Hall, Inc.

Lockwood, P., Jordan, C. H., & Kunda, Z. (2002). Motivation by positive or negative role models: Regulatory focus determines who will best inspire us. *Journal of Personality and Social Psychology, 83,* 854–864. https://doi.org/10.1037/0022-3514.83.4.854

Maass, A. (1999). Linguistic intergroup bias: Stereotype perpetuation through language. In M. P. Zanna (Ed.), *Advances in experimental social psychology, Vol. 31.* (pp. 79–121). Academic Press. https://doi.org/10.1016/S0065-2601(08)60272-5

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological review*, *98*(2), 224-253. https://doi.org/10.1037/0033-295x.98.2.224

Mead, G. H. (1934). *Mind, Self, and Society*. Chicago: University of Chicago Press.

Mehlhorn, K., Newell, B. R., Todd, P. M., Lee, M. D., Morgan, K., Braithwaite, V. A., Hausmann, D., Fiedler, K., & Gonzalez, C. (2015). Unpacking the exploration–exploitation tradeoff: A synthesis of human and animal literatures. *Decision*, *2*(3), 191–215. https://doi.org/10.1037/dec0000033

Melnikoff, D. E., & Bargh, J. A. (2018). The mythical number two. *Trends in Cognitive Sciences, 22*(4), 280–293. https://doi.org/10.1016/j.tics.2018.02.001

Miron, A. M., & Brehm, J. W. (2006). Reactance theory-40 years later. *Zeitschrift für Sozialpsychologie*, *37*(1), 9-18. https://doi.org/10.1024/0044-3514.37.1.9

Moscovici, S. (1976). *Social influence and social change* (Vol. 10). Academic Press.

Mummendey, A., & Wenzel, M. (1999). Social discrimination and tolerance in intergroup relations: Reactions to intergroup difference. *Personality and Social Psychology Review*, 3, 158 -174. https://doi.org/10.1207%2Fs15327957pspr0302_4

Nakassis, C. V. (2005). Formalizing context. *Theory & Psychology*, *15*(1), 128–131. https://doi.org/10.1177/0959354305015001107

Nemeth, C. J. (1986). Differential contributions of majority and minority influence. *Psychological review, 93*(1), 23-32. https://doi.org/10.1037/0033-295X.93.1.23

Nussbaum, S., Trope, Y., & Liberman, N. (2003). Creeping dispositionism: The temporal dynamics of behavior prediction. *Journal of personality and social psychology*, *84*(3), 485-497. https://psycnet.apa.org/doi/10.1037/0022-3514.84.3.485

Ortmann, A., & Rydval, O. (2004). Review of Behavioral Game Theory. *Journal of Economic Psychology*, *25*(5), 671–674. https://doi.org/10.1016/j.joep.2004.06.001

Parducci, A. (1968). The relativism of absolute judgments. *Scientific American*, *219*(6), 84-93.

Pavlov, I. P. (1927). *Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex*. Oxford Univ. Press.

Pechmann, C., & Stewart, D. W. (1988). Advertising repetition: A critical review of wearin and wearout. *Current issues and research in advertising*, *11*(1-2), 285-329. doi:10.1080/01633392.1988.10504936

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In *Communication and persuasion* (pp. 1-24). New York: Springer.

Pham, M. T., & Avnet, T. (2004). Ideals and oughts and the reliance on affect versus substance in persuasion. *Journal of Consumer Research, 30*, 503–518. https://doi.org/10.1086/380285

Roediger, H. L. (1980). Memory metaphors in cognitive psychology. *Memory & Cognition*, *8*(3), 231–246. https://doi.org/10.3758/BF03197611

Rogers, C. R. (1951). *Client-centered therapy: Its current practice, implications, and theory*. Oxford, United Kingdom: Houghton Mifflin.

Rothbart, M., & Lewis, S. H. (1994). Cognitive processes and intergroup relations: A historical perspective. In P. G. Devine, D. L. Hamilton, & T. M. Ostrom (Eds.), *Social cognition: Impact on social psychology*. (pp. 347–382). Academic Press.

Sagristano, M. D., Trope, Y., & Liberman, N. (2002). Time-dependent gambling: Odds now, money later. *Journal of Experimental Psychology: General*, *131*(3), 364–376. https://doi.org/10.1037/0096-3445.131.3.364

Schaller, M., Conway, L. G. III, & Tanchuk, T. L. (2002). Selective pressures on the once and future contents of ethnic stereotypes: Effects of the communicability of traits. *Journal of Personality and Social Psychology*, *82*(6), 861-877. https://doi.org/10.1037/0022-3514.82.6.861

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: informative and directive functions of affective states. *Journal of personality and social psychology*, *45*(3), 513-523. https://doi.org/10.1037/0022-3514.45.3.513

Semin, G. R., & Fiedler, K. (1991). The linguistic category model, its bases, applications and range. *European review of social psychology*, *2*(1), 1-30. https://doi.org/10.1080/14792779143000006

Sherman, J. W. (2006). On building a better process model: It's not only how many, but which ones and by which means? *Psychological Inquiry*, *17*(3), 173-184. https://doi.org/10.1207/s15327965pli1703_3

Skinner, B. F. (1938). *The Behaviour of organisms: An experimental analysis*. New York: Appleton-Century.

Smith, E. R. (1993). Social identity and social emotions: Toward new conceptualizations of prejudice. In D. M. Mackie & D. L. Hamilton (Eds.), *Affect, Cognition, and Stereotyping: Interactive Processes in Group Perception* (pp. 297–315). San Diego, CA: Academic Press. https://doi.org/10.1016/b978-0-08-088579-7.50017-x

Staats, A. W., & Staats, C. K. (1958). Attitudes established by classical conditioning. *The Journal of Abnormal and Social Psychology, 57*(1), 37–40. https://doi.org/10.1037/h0042782

Stephan, W. G., & Renfro, C. L. (2002). The role of threat in intergroup relations. In D. M. Mackie & E. R. Smith (Eds.), *From prejudice to inter-group emotions: Differentiated reactions to social groups* (pp. 191–207). New York, NY: Psychology Press.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and social psychology review*, *8*(3), 220-247. https://doi.org/10.1207/s15327957pspr0803_1

Tajfel, H. (1957). Value and the perceptual judgment of magnitude. *Psychological Review*, *64*(3), 192–204. https://doi.org/10.1037/h0047878

Tajfel, H. (1969). Cognitive aspects of prejudice. *Journal of Social Issues*, *25*(4), 79–97. https://doi.org/10.1111/j.1540-4560.1969.tb00620.x

Tajfel, H. (1978). *Differentiation between social groups*. London: Academic Press.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, *1*(2), 149–178. https://doi.org/10.1002/ejsp.2420010202

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33 – 47). Monterey, CA: Brooks-Cole.

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

Taylor, S. E., Klein, L. C., Lewis, B. P., Gruenewald, T. L., Gurung, R. A., & Updegraff, J. A. (2000). Biobehavioral responses to stress in females: tend-and-befriend, not fight-or-flight. *Psychological review*, *107*(3), 411-429. https://doi.org/10.1037/0033-295x.107.3.411

Thibaut, J. W., & Kelley, H. H. (1959). *Social science classics series. The social psychology of groups*. Transaction Publishers.

Thorndike, E. L. (1898). Animal intelligence: An experimental study of the associative processes in animals. *The Psychological Review: Monograph Supplements, 2*(4), i-109. https://doi.org/10.1037/h0092987

Trope, Y., & Liberman, N. (2003). Temporal construal. *Psychological Review*, *110*(3), 403–421. https://doi.org/10.1037/0033-295X.110.3.403

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review*, *117*(2), 440–463. https://doi.org/10.1037/a0018963

Trueblood, J. S., & Busemeyer, J. R. (2011). A quantum probability account of order effects in inference. *Cognitive Science*, *35*(8), 1518–1552. https://doi.org/10.1111/j.1551-6709.2011.01197.x

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Basil Blackwell.

Unkelbach, C., Fiedler, K., Bayer, M., Stegmüller, M., & Danner, D. (2008). Why positive information is processed faster: The density hypothesis. *Journal of personality and social psychology*, *95*(1), 36-49. https://doi.org/10.1037/0022-3514.95.1.36

Vallacher, R. R., & Wegner, D. M. (1987). What do people think they're doing? Action identification and human behavior. *Psychological review*, *94*(1), 3-15. https://doi.org/10.1037/0033-295x.94.1.3

Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior.* Princeton University Press.

Weber, M. (1947). *Theory of Social and Economic Organization* (Edited with Introduction by T. Parsons). New York: Free Press.

Webster, D., & Kruglanski, A. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology, 67*, 1049–1062. https://doi.org/10.1037/0022-3514.67.6.1049

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological review*, *92*(4), 548-573. https://doi.org/10.1037/0033-295x.92.4.548

Wyer, R. S., & Srull, T. K. (1986). Human cognition in its social context. *Psychological review*, *93*(3), 322-359. https://doi.org/10.1037/0033-295x.93.3.322

Zhang, J. (2004). Residential segregation in an all-integrationist world. *Journal of Economic Behavior & Organization*, *54*(4), 533–550. https://doi.org/10.1016/j.jebo.2003.03.005

Zhang, T.-Y., & Meaney, M. J. (2010). Epigenetics and the environmental regulation of the genome and its function. *Annual Review of Psychology*, *61*, 439–466. https://doi.org/10.1146/annurev.psych.60.110707.163625

Zimbardo, P. G. (2006). The Psychology of Power: To the Person? To the Situation? To the System? In D. L. Rhode (Ed.), *Moral leadership: The theory and practice of power, judgment and policy*. (pp. 129–157). Jossey-Bass.

Chapter 3.1.

Bernstein, D. M., Aßfalg, A., Kumar, R., & Ackerman, R. (2016). *Looking Backward and Forward on Hindsight Bias* (J. Dunlosky, & S. K. Tauber, Eds.) (Vol. 1). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199336746.013.7

Bless, H., Strack, F., & Schwarz, N. (1993). The informative functions of research procedures: Bias and the logic of conversation. *European Journal of Social Psychology, 23*, 149-165. https://doi.org/10.1002/ejsp.2420230204

Grice, H. P. (1975). Logic and Conversation. In P. Cole & J. Morgan (Eds.) *Syntax and Semantics*. Academic Press. https://doi.org/10.1163/9789004368811_003

Hertwig, R., & Gigerenzer, G. (1999). The ‘conjunction fallacy’ revisited: How intelligent inferences look like reasoning errors. *Journal of behavioral decision making*, *12*(4), 275-

305. https://doi.org/10.1002/(SICI)1099-0771(199912)12:4%3C275::AID-BDM323%3E3.0.CO;2-M

Hilton, D. J. (1995). The social context of reasoning: Conversational inference and rational judgment. *Psychological Bulletin, 118*(2), 248–271. https://doi.org/10.1037/0033-2909.118.2.248

Pohl, R. F., & Erdfelder, E. (2022). Hindsight bias. In R. F. Pohl (Ed.), *Cognitive illusions: Intriguing phenomena in thinking, judgment and memory* (pp. 424–445). Routledge/Taylor & Francis Group.

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist, 54*(2), 93–105. https://doi.org/10.1037/0003-066X.54.2.93

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Cambridge, MA: Harvard University Press.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, *185*(4157), 1124-1131. https://doi.org/10.1126/science.185.4157.1124

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90*(4), 293–315. https://doi.org/10.1037/0033-295X.90.4.293

Chapter 3.2.

Bartlett, F. C. (1932). *Remembering: a study in experimental and social psychology*. Macmillan.

Echterhoff, G., Higgins, E. T., & Levine, J. M. (2009). Shared Reality: Experiencing Commonality with others' Inner States about the World. Perspectives on Psychological Science, 4(5), 496–521. https://doi.org/10.1111/j.1745-6924.2009.01161.x

Higgins, E. T., McCann, C. D., & Fondacaro, R. (1982). The “communication game”: Goal-directed encoding and cognitive consequences. *Social Cognition*, *1*(1), 21-37. https://doi.org/10.1521/soco.1982.1.1.21

Higgins, E. T., & Rholes, W. S. (1978). “Saying is believing”: Effects of message modification on memory and liking for the person described. *Journal of Experimental Social Psychology*, *14*(4), 363–378. https://doi.org/10.1016/0022-1031(78)90032-x

Hörmann, H. (1976). *Meinen und Verstehen: Grundzüge einer psychologischen Semantik*. Frankfurt a. M.: Suhrkamp.

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of verbal learning and verbal behavior*, *13*(5), 585-589. https://doi.org/10.1016/S0022-5371(74)80011-3

Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory, 4*(1), 19–31. https://doi.org/10.1037/0278-7393.4.1.19

Chapter 3.3.

Ackerman, R., Bernstein, D. M., & Kumar, R. (2020). Metacognitive hindsight bias. *Memory & Cognition*, *48*(5), 731–744. https://doi.org/10.3758/s13421-020-01012-w

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bernstein, D. M., Aßfalg, A., Kumar, R., & Ackerman, R. (2016). *Looking Backward and Forward on Hindsight Bias* (J. Dunlosky, & S. K. Tauber, Eds.) (Vol. 1). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199336746.013.7

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition*, *35*(2), 201–210. https://doi.org/10.3758/BF03193441

Blank, H., & Nestler, S. (2007). Cognitive Process Models of Hindsight Bias. *Social Cognition*, *25*(1), 132–146. https://doi.org/10.1521/soco.2007.25.1.132

Blank, H., Nestler, S., Collani, G. von, & Fischer, V. (2008). How many hindsight biases are there? *Cognition*, *106*(3), 1408–1440. https://doi.org/10.1016/j.cognition.2007.07.007

Brysbaert, M., & Stevens, M. (2018). Power Analysis and Effect Size in Mixed Effects Models: A Tutorial. *Journal of Cognition*, *1*(1), 9. https://doi.org/10.5334/joc.10

Dietvorst, B. J., & Simonsohn, U. (2019). Intentionally "biased": People purposely use to-be-ignored information, but can be persuaded not to. *Journal of Experimental Psychology. General*, *148*(7), 1228–1238. https://doi.org/10.1037/xge0000541

Echterhoff, G., Higgins, E. T., & Groll, S. (2005). Audience-tuning effects on memory: The role of shared reality. *Journal of Personality and Social Psychology*, *89*(3), 257–276. https://doi.org/10.1037/0022-3514.89.3.257

Echterhoff, G., Higgins, E. T., Kopietz, R., & Groll, S. (2008). How communication goals determine when audience tuning biases memory. *Journal of Experimental Psychology. General*, *137*(1), 3–21. https://doi.org/10.1037/0096-3445.137.1.3

Fischhoff, B. (1975). Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, *1*(3), 288–299. https://doi.org/10.1037/0096-1523.1.3.288

Green, P., & MacLeod, C. J. (2016). simr: an R package for power analysis of generalised linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Grice, H. P. (1975). Logic and Conversation. In P. Cole & J. Morgan (Eds.) *Syntax and Semantics*. Academic Press. https://doi.org/10.1163/9789004368811_003

Guilbault, R. L., Bryant, F. B., Brockway, J. H., & Posavac, E. J. (2004). A Meta-Analysis of Research on Hindsight Bias. *Basic and Applied Social Psychology*, *26*(2-3), 103–117. https://doi.org/10.1080/01973533.2004.9646399

Hardin, C. D., & Higgins, E. T. (1996). Shared reality: How social verification makes the subjective objective. In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition. Handbook of motivation and cognition, Vol. 3. The interpersonal context* (p. 28–84). The Guilford Press.

Hastie, R., & Park, B. (1986). The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. *Psychological Review*, *93*(3), 258–268. https://doi.org/10.1037/0033-295x.93.3.258

Herzog, S. M., & Hertwig, R. (2009). The wisdom of many in one mind: Improving individual judgments with dialectical bootstrapping. *Psychological Science*, *20*(2), 231–237. https://doi.org/10.1111/j.1467-9280.2009.02271.x

Higgins, E. T. (1992). Achieving 'Shared Reality' in the Communication Game: A Social Action That Create; Meaning. *Journal of Language and Social Psychology*, *11*(3), 107–131. https://doi.org/10.1177/0261927X92113001

Higgins, E. T. (1999). Saying is believing" effects: When sharing reality about something biases knowledge and evaluations. In *Shared cognition in organizations: The management of knowledge* (Vol. 1, pp. 33–49).

Higgins, E. T., & McCann, C. D. (1984). Social encoding and subsequent attitudes, impressions, and memory: "Context-driven" and motivational aspects of processing. *Journal of Personality and Social Psychology*, *47*(1), 26–39. https://doi.org/10.1037/0022-3514.47.1.26

Higgins, E. T., & Rholes, W. S. (1978). "Saying is believing": Effects of message modification on memory and liking for the person described. *Journal of Experimental Social Psychology*, *14*(4), 363–378. https://doi.org/10.1016/0022-1031(78)90032-x

Hoffrage, U., Hertwig, R., & Gigerenzer, G. (2000). Hindsight bias: A by-product of knowledge updating? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(3), 566–581. https://doi.org/10.1037//0278-7393.26.3.566

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with More Than One Random Factor: Designs, Analytic Models, and Statistical Power. *Annual Review of Psychology*, *68*, 601–625. https://doi.org/10.1146/annurev-psych-122414-033702

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Leiner, D. J. (2020). *SoSci Survey* (Version 3.2.06) [Computer software]. www.soscisurvey.de

Loftus, E. F. (1975). Leading questions and the eyewitness report. *Cognitive Psychology*, *7*(4), 560–572. https://doi.org/10.1016/0010-0285(75)90023-7

Lüdecke, D. (2019). *sjPlot: Data Visualization for Statistics in Social Science*. https://CRAN.R-project.org/package=sjPlot

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Nestler, S., Blank, H., & Egloff, B. (2010). Hindsight ≠ hindsight: Experimentally induced dissociations between hindsight components. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(6), 1399–1413. https://doi.org/10.1037/a0020449

Pohl, R. F. (2007). Ways to Assess Hindsight Bias. *Social Cognition*, *25*(1), 14–31. https://doi.org/10.1521/soco.2007.25.1.14

Pohl, R. F., Eisenhauer, M., & Hardt, O. (2003). Sara: A cognitive process model to simulate the anchoring effect and hindsight bias. *Memory*, *11*(4-5), 337–356. https://doi.org/10.1080/09658210244000487

Pohl, R. F., & Erdfelder, E. (2016). Hindsight bias. In *Cognitive Illusions: Intriguing Phenomena in Judgement, Thinking and Memory* (2nd ed., pp. 424–445). Psychology Press. https://doi.org/10.4324/9781315696935

Pohl, R. F., & Hell, W. (1996). No Reduction in Hindsight Bias after Complete Information and Repeated Testing. *Organizational Behavior and Human Decision Processes*, *67*(1), 49–58. https://doi.org/10.1006/obhd.1996.0064

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. https://www.R-project.org/

Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition*, *21*(1), 89–102. https://doi.org/10.3758/BF03211168

Roese, N. J., & Vohs, K. D. (2012). Hindsight Bias. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *7*(5), 411–426. https://doi.org/10.1177/1745691612454303

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Schwarz, N. (1996). *Cognition and Communication: Judgmental Biases, Research Methods, and the Logic of Conversation. Distinguished Lecture Series*. Taylor and Francis. https://doi.org/10.4324/9781315805887

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory*, *4*(6), 592–604. https://doi.org/10.1037/0278-7393.4.6.592

Slovic, P., & Fischhoff, B. (1977). On the psychology of experimental surprises. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(4), 544–551. https://doi.org/10.1037/0096-1523.3.4.544

Son, L. K., Hong, S. S., Han, L., Lee, Y., & Kim, T. H. (2021). Taking a naïve other's perspective to debias the hindsight bias: Did it backfire? *New Ideas in Psychology*, *62*, 100867. https://doi.org/10.1016/j.newideapsych.2021.100867

Tetlock, P. E. (1983). Accountability and complexity of thought. *Journal of Personality and Social Psychology*, *45*(1), 74–83. https://doi.org/10.1037/0022-3514.45.1.74

Tulving, E. (1972). Episodic and semantic memory. In *Organization of memory* (xiii, 423-xiii, 423). Academic Press.

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology. General*, *143*(5), 2020–2045. https://doi.org/10.1037/xge0000014

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Chapter 4.1.

Austin, J. L. (1962). *How to Do Things with Words: The William James Lectures Delivered in Harvard University in 1955*. Oxford, England: Oxford University Press UK.

Bless, H., Betsch, T., & Franzen, A. (1998). Framing the framing effect: The impact of context cues on solutions to the 'Asian disease'problem. *European journal of social psychology*, *28*(2), 287-291. https://doi.org/10.1002/(SICI)1099-0992(199803/04)28:2%3C287::AID-EJSP861%3E3.0.CO;2-U

Fiedler, K. (2008). Language: A Toolbox for Sharing and Influencing Social Reality. *Perspectives on Psychological Science*, 3(1), 38–47. https://doi.org/10.1111/j.1745-6916.2008.00060.x

Grice, H. P. (1975). Logic and Conversation. In P. Cole & J. Morgan (Eds.) *Syntax and Semantics*. Academic Press. https://doi.org/10.1163/9789004368811_003

Hörmann, H. (1976). *Meinen und Verstehen: Grundzüge einer psychologischen Semantik*. Frankfurt a. M.: Suhrkamp.

Kühberger, A. (2022). Framing. In *Cognitive illusions* (pp. 61-77). Routledge.

Levine, T. R. (2022). Truth-default theory and the psychology of lying and deception detection. *Current Opinion in Psychology*, *47*, 101380. https://doi.org/10.1016/j.copsyc.2022.101380

Luhmann, N. (1984). *Soziale Systeme: Grundriß einer allgemeinen Theorie*. Suhrkamp.

Maass, A. (1999). Linguistic intergroup bias: Stereotype perpetuation through language. In M. P. Zanna (Ed.), *Advances in experimental social psychology, Vol. 31*. (pp. 79–121). Academic Press. https://doi.org/10.1016/S0065-2601(08)60272-5

Mandel, D. R. (2014). Do framing effects reveal irrational choice? Journal of Experimental Psychology: General, 143, 1185–1198. https://doi.org/10.1037/a0034207

McKenzie, C. R. M., Nelson, J. D. What a speaker's choice of frame reveals: Reference points, frame selection, and framing effects. *Psychonomic Bulletin & Review* **10**, 596–602 (2003). https://doi.org/10.3758/BF03196520

PeConga, E., Pickrell, J. E., Bernstein, D. M., & Loftus, E. F. (2022). Misinformation effect. In *Cognitive Illusions* (pp. 419-435). Routledge.

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge University Press.

Semin, G. R., & Fiedler, K. (1991). The linguistic category model, its bases, applications and range. *European review of social psychology*, *2*(1), 1-30. https://doi.org/10.1080/14792779143000006

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *science*, *211*(4481), 453-458. https://doi.org/10.1126/science.7455683

Chapter 4.2.

Brunswik, E. (1947). *Systematic and representative design of psychological experiments; with* results in physical and social perception. University of California Press.

Budescu, D. V., Weinberg, S., & Wallsten, T.S. (1988). Decisions based on numerically and verbally expressed uncertainties. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 281-294.

Camerer, C. F. (2005). Three cheers—psychological, theoretical, empirical— for loss aversion. *Journal of Marketing Research, 42,* 129–133. http://dx.doi.org/10.1509/jmkr.42.2.129.62286

Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, *54*(4), 297–312. https://doi.org/10.1037/h0040950

Campbell, D. T., & Kenny, D. A. (1999). *A primer on regression artifacts*. Guilford Publications.

Costello, F., & Watts, P. (2019). The rationality of illusory correlation. *Psychological Review*, *126*(3), 437–450. https://doi.org/10.1037/rev0000130

De Finetti, B. (1937). La prévision: Ses lois logiques, ses sources subjectives [Foresight: its logical laws, its subjective sources]. *Annales de l'Institut Henri Poincaré, 17,* 1–68.

Denes-Raj, V., Epstein, S., & Cole, J. (1995). The generality of the ratio-bias phenomenon. *Personality and Social Psychology Bulletin, 21*(10), 1083-1092. https://doi.org/10.1177/01461672952110009

Denrell, J. C., & Le Mens, G. (2008). Illusory correlation as the outcome of experience sampling. In *Proceedings of the Annual Meeting of the Cognitive Science Society, 30*(30), 421-426.

Dhami, M. K., Hertwig, R., & Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. *Psychological Bulletin*, *130*(6), 959–988. https://doi.org/10.1037/0033-2909.130.6.959

Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *The Quarterly Journal of Economics*, 643-669.

Fiedler, K. (1991). The tricky nature of skewed frequency tables: An information loss account of distinctiveness-based illusory correlations. *Journal of Personality and Social Psychology*, *60*(1), 24.

Fiedler, K. (2000). Illusory correlations: A simple associative algorithm provides a convergent account of seemingly divergent paradigms. *Review of General Psychology*, *4*(1), 25-58.

Fiedler, K., Ackerman, R., & Scarampi, C. (2019). Metacognition: monitoring and controlling one's own knowledge, reasoning and decisions. In R. J. Sternberg & J. Funke (Eds.), *The psychology of human thought: An introduction* (pp. 89-103). Heidelberg University Publishing.

Fiedler, K., Kareev, Y., Avrahami, J., Beier, S., Kutzner, F., & Hütter, M. (2016). Anomalies in the detection of change: When changes in sample size are mistaken for changes in proportions. *Memory & Cognition*, *44*(1), 143–161. https://doi.org/10.3758/s13421-015-0537-z

Fiedler, K., & Krueger, J. I. (2012). More than an artifact: Regression as a theoretical construct. In J. I. Krueger (Ed.), Frontiers of social psychology. Social judgment and decision making (p. 171–189). Psychology Press.

Fiedler, K., Walther, E., & Nickel, S. (1999). The auto-verification of social hypotheses: Stereotyping and the power of sample size. *Journal of Personality and Social Psychology*, *77*(1), 5–18. https://doi.org/10.1037/0022-3514.77.1.5

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1979). Subjective sensitivity analysis. *Organizational Behavior and Human Performance*, *23*(3), 339-359.

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

Harris, A. J. L., & Speekenbrink, M. (2016). Semantic cross-scale numerical anchoring. *Judgment and Decision Making*, *11*(6), 572–581.

Hertwig, R., Hogarth, R. M., & Lejarraga, T. (2018). Experience and description: Exploring two paths to knowledge. *Current Directions in Psychological Science*, *27*(2), 123–128. https://doi.org/10.1177/0963721417740645

Kalberg, S. (1980). Max Weber's types of rationality: Cornerstones for the analysis of rationalization processes in history. *American Journal of Sociology*, *85*(5), 1145-1179.

Kruglanski, A. W. (1990). Lay epistemic theory in social-cognitive psychology. *Psychological Inquiry*, *1*(3), 181-197. https://doi.org/10.1207/s15327965pli0103_1

Kutzner, F. L., & Fiedler, K. (2015). No correlation, no evidence for attention shift in category learning: Different mechanisms behind illusory correlations and the inverse base-rate effect. *Journal of Experimental Psychology: General, 144,* 58–75.

Le Mens, G., & Denrell, J. (2011). Rational learning and information sampling: On the "naivety" assumption in sampling explanations of judgment biases. *Psychological Review*, 118(2), 379-392. doi:10.1037/a0023010.supp (Supplemental)

Lieder, F., & Griffiths, T. L. (2019). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*. https://doi.org/10.1017/S0140525X1900061X

Marcus, G. F., & Davis, E. (2013). How robust are probabilistic models of higher-level cognition?. *Psychological Science*, *24*(12), 2351-2360.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, *115*(2), 502–517. https://doi.org/10.1037/0033-295X.115.2.502

Price, P. C., Smith, A. R., & Lench, H. C. (2006). The effect of target group size on risk judgments and comparative optimism: The more, the riskier. *Journal of Personality and Social Psychology*, *90*(3), 382–398. https://doi.org/10.1037/0022-3514.90.3.382

Reyna, V. F., & Brainerd, C. J. (2008). Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. *Learning and Individual Differences*, *18*(1), 89–107. https://doi.org/10.1016/j.lindif.2007.03.011

Rosenthal, R., & Rubin, D. B. (2003). requivalent: A simple effect size indicator. *Psychological Methods*, *8*(4), 492–496. https://doi.org/10.1037/1082-989X.8.4.492

Sarbin, T. R., Taft, R., & Bailey, D. E. (1960). *Clinical inference and cognitive theory*.

Savage, L. J. (1954). *The foundations of statistics*. New York: JohnWiley and Sons.

Stanovich, K. E. (2011). *Rationality and the reflective mind*. New York: Taylor & Francis.

Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2008). Decision making, movement planning and statistical decision theory. *Trends in Cognitive Sciences*, *12*(8), 291–297. https://doi.org/10.1016/j.tics.2008.04.010

Unkelbach, C., Koch, A., Silva, R. R., & Garcia-Marques, T. (2019). Truth by repetition: Explanations and implications. *Current Directions in Psychological Science*, *28*(3), 247-253.

van Lange, P. A., Joireman, J., Parks, C. D., & van Dijk, E. (2013). The psychology of social dilemmas: A review. *Organizational Behavior and Human Decision Processes*, *120*(2), 125-141.

Walasek, L., & Stewart, N. (2015). How to make loss aversion disappear and reverse: Tests of the decision by sampling origin of loss aversion. *Journal of Experimental Psychology: General*, *144*(1), 7–11. https://doi.org/10.1037/xge0000039

Zajonc, R. B. (2001). Mere exposure: A gateway to the subliminal. *Current Directions in Psychological Science, 10*(6), 224-228.

Chapter 5.1.

Bateson, G. (1972). *Steps to an ecology of mind: Collected essays in anthropology, psychiatry, evolution, and epistemology*. Jason Aronson.

Bartlett, F. C. (1932). *Remembering. A study in experimental and social psychology*. Cambridge: University Press.

Bühler, K. (1934). *Sprachtheorie*. Fischer.

Bechmann, S. (2018). Pragmatische Zeichentheorie. In: Liedtke, F., Tuchen, A. (eds) Handbuch Pragmatik. J.B. Metzler, Stuttgart. https://doi.org/10.1007/978-3-476-04624-6_6

Brunswik, E. (1952). The conceptual framework of psychology. In O. Neurath, R. Carnap & C. W. Morris (Eds.), *International Encyclopedia of Unified Science, Volume 1, Number 10,* (pp. 567 – 655). University of Chicago Press.

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

Eco, U. (1979). *A theory of semiotics*. Indiana University Press.

Fiedler, K., Salmen, K., & Ermark, F. (2022). Illusory correlation. In *Cognitive illusions* (pp. 92-107). Routledge.

Hammond, K. R. (1955). Probabilistic functioning and the clinical method. *Psychological Review, 62*(4), 255–262. https://doi.org/10.1037/h0046845

Morris, C. (1946). *Signs, language and behavior.* Prentice-Hall.

Peirce, C. S. (1931). *Collected papers*. Cambridge,: Belknap Press of Harvard University Press.

Saussure, F. de (1964). *Course in general linguistics*. London: Peter Owen.

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge University Press.

Chapter 5.2.

Abbey, J. D., & Meloy, M. G. (2017). Attention by design: Using attention checks to detect inattentive respondents and improve data quality. *Journal of Operations Management*, *53-56, 63-70*. https://doi.org/10.1016/j.jom.2017.06.001

Ares, G., Mawad, F., Giménez, A., & Maiche, A. (2014). Influence of rational and intuitive thinking styles on food choice: Preliminary evidence from an eye-tracking study with yogurt labels. *Food Quality and Preference, 31*, 28-37. https://doi.org/10.1016/j.foodqual.2013.07.005

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated Measures Correlation. *Frontiers in Psychology*, *8*, 456. https://doi.org/10.3389/fpsyg.2017.00456

Bartlett, F. C. (1932). *Remembering. A study in experimental and social psychology*. Cambridge: University Press.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Becker, L., van Rompay, T. J., Schifferstein, H. N., & Galetzka, M. (2011). Tough package, strong taste: The influence of packaging design on taste impressions and product evaluations. *Food Quality and Preference, 22(1),* 17-23. https://doi.org/10.1016/j.foodqual.2010.06.007

Briers, B., Huh, Y. E., Chan, E., & Mukhopadhyay, A. (2020). The unhealthy= tasty belief is associated with BMI through reduced consumption of vegetables: A cross-national and mediational analysis. *Appetite*, *150*, 104639. https://doi.org/10.1016/j.appet.2020.104639

Brunswik, E. (1952). The conceptual framework of psychology. In O. Neurath, R. Carnap & C. W. Morris (Eds.), *International Encyclopedia of Unified Science, Volume 1, Number 10,* (pp. 567 – 655). University of Chicago Press.

Bublitz, M. G., Peracchio, L. A., & Block, L. G. (2010). Why did I eat that? Perspectives on food decision making and dietary restraint. Journal of Consumer Psychology, 20(3), 239-258. https://doi.org/10.1016/j.jcps.2010.06.008

Burke, M. J., Brief, A. P., & George, J. M. (1993). The role of negative affectivity in understanding relations between self-reports of stressors and strains: A comment on the applied psychology literature. *Journal of Applied Psychology, 78*(3), 402–412. https://doi.org/10.1037/0021-9010.78.3.402

Cavanagh, K. V., & Forestell, C. A. (2013). The effect of brand names on flavor perception and consumption in restrained and unrestrained eaters. Food Quality and Preference, 28(2), 505-509. https://doi.org/10.1016/j.foodqual.2012.12.004

Coates, A. E., Hardman, C. A., Halford, J. C. G., Christiansen, P., & Boyland, E. J. (2019). Food and Beverage Cues Featured in YouTube Videos of Social Media Influencers Popular With Children: An Exploratory Study. *Frontiers in Psychology*, *10*, 2142. https://doi.org/10.3389/fpsyg.2019.02142

Dubé, L., Fatemi, H., Lu, J., & Hertzer, C. (2016). The Healthier the Tastier? USA–India Comparison Studies on Consumer Perception of a Nutritious Agricultural Product at Different Food Processing Levels. *Frontiers in Public Health*, *4*. https://doi.org/10.3389/fpubh.2016.00006

Fenko, A., Schifferstein, H. N., Huang, T. C., & Hekkert, P. (2009). What makes products fresh: The smell or the colour?. *Food Quality and Preference, 20*, 372-379. https://doi.org/10.1016/j.foodqual.2009.02.007

Fiedler, K., Bluemke, M., Freytag, P., Unkelbach, C., & Koch, S. (2008). A semiotic approach to understanding the role of communication in stereotyping. In Y. Kashima, K. Fiedler, & P. Freytag (Eds.), *Stereotype dynamics: Language-based approaches to the formation, maintenance, and transformation of stereotypes* (pp. 95–116). Lawrence Erlbaum Associates Publishers. https://doi.org/10.4324/9780203809990

Fishbach, A., & Zhang, Y. (2008). Together or apart: When goals and temptations complement versus compete. *Journal of Personality and Social Psychology, 94*(4), 547-559. https://doi.org/10.1037/0022-3514.94.4.547

Gandhi, N., Zou, W., Meyer, C., Bhatia, S., & Walasek, L. (2020). Knowledge Representations in Health Judgments. In S. Denison., M. Mack, Y. Xu, & B.C. Armstrong (Eds.), Proceedings of the 42nd Annual Conference of the Cognitive Science Society (pp. 280 - 286). Cognitive Science Society.

Haasova, S., & Florack, A. (2019a). Practicing the (un) healthy= tasty intuition: Toward an ecological view of the relationship between health and taste in consumer judgments. *Food Quality and Preference, 75*, 39-53. https://doi.org/10.1016/j.foodqual.2019.01.024

Haasova, S., & Florack, A. (2019b). Sugar labeling: How numerical information of sugar content influences healthiness and tastiness expectations. *Plos One*, *14*(11), e0223510. https://doi.org/10.1371/journal.pone.0223510

Hallez, L., Vansteenbeeck, H., Boen, F., & Smits, T. (2023). Persuasive packaging? The impact of packaging color and claims on young consumers' perceptions of product healthiness, sustainability, and tastiness. *Appetite,* 182, 106433. https://doi.org/10.1016/j.appet.2022.106433

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

Hout, M. C., Goldinger, S. D., & Ferguson, R. W. (2013). The versatility of SpAM: A fast, efficient, spatial method of data collection for multidimensional scaling. *Journal of Experimental Psychology. General*, *142*(1), 256–281. https://doi.org/10.1037/a0028860

Huang, Y., & Wu, J. (2016). Food pleasure orientation diminishes the "healthy= less tasty" intuition. *Food Quality and Preference*, *54*, 75-78. https://doi.org/10.1016/j.foodqual.2016.07.007

Irmak, C., Vallen, B., & Robinson, S. R. (2011). The impact of product name on dieters' and nondieters' food evaluations and consumption. *Journal of Consumer Research, 38*(2), 390-405. https://doi.org/10.1086/660044

Jo, J., & Lusk, J. L. (2018). If it's healthy, it's tasty and expensive: Effects of nutritional labels on price and taste expectations. *Food Quality and Preference, 68*, 332-341. https://doi.org/10.1016/j.foodqual.2018.04.002

Johansen, S. B., Næs, T., & Hersleth, M. (2011). Motivation for choice and healthiness perception of calorie-reduced dairy products. A cross-cultural study. *Appetite, 56*(1), 15-24. https://doi.org/10.1016/j.appet.2010.11.137

Kalish, M. L., Griffiths, T. L., & Lewandowsky, S. (2007). Iterated learning: Intergenerational knowledge transmission reveals inductive biases. *Psychonomic Bulletin & Review, 14*(2), 288–294. https://doi.org/10.3758/BF03194066

Karnal, N., Machiels, C. J., Orth, U. R., & Mai, R. (2016). Healthy by design, but only when in focus: Communicating non-verbal health cues through symbolic meaning in packaging. *Food Quality and Preference, 52*, 106-119. https://doi.org/10.1016/j.foodqual.2016.04.004

Kashima, Y. (2000). Maintaining Cultural Stereotypes in the Serial Reproduction of Narratives. *Personality and Social Psychology Bulletin*, *26*(5), 594–604. https://doi.org/10.1177/0146167200267007

Kirby, S. (2001). Spontaneous evolution of linguistic structure-an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation, 5*(2), 102-110. https://doi.org/10.1109/4235.918430

Knöferle, K., & Spence, C. (2012). Crossmodal correspondences between sounds and tastes. *Psychonomic Bulletin & Review*. 1-15. https://doi.org/10.3758/s13423-012-0321-z

König, L. M., & Renner, B. (2018). Colourful= healthy? Exploring meal colour variety and its relation to food consumption. *Food Quality and Preference*, *64*, 66-71. https://doi.org/10.1016/j.foodqual.2017.10.011

König, L. M., & Renner, B. (2019). Boosting healthy food choices by meal colour variety: results from two experiments and a just-in-time Ecological Momentary Intervention. *BMC Public Health*, *19*(1), 975. https://doi.org/10.1186/s12889-019-7306-z

Kunz, S., Haasova, S., & Florack, A. (2020a). Fifty shades of food: The influence of package color saturation on health and taste in consumer judgments. *Psychology & Marketing*, *37*(7), 900-912. https://doi.org/10.1002/mar.21317

Kunz, S., Haasova, S., Rieß, J., & Florack, A. (2020b). Beyond healthiness: The impact of traffic light labels on taste expectations and purchase intentions. *Foods, 9*(2), 134. https://doi.org/10.3390/foods9020134

Lancelot Miltgen, C., Pantin Sohier, G., & Grohmann, B. (2016). Communicating sensory attributes and innovation through food product labeling. *Journal of Food Products Marketing*, *22*(2), 219-239. https://doi.org/10.1080/10454446.2014.1000435

Leiner, D. J. (2020). *SoSci Survey* (Version 3.2.06) [Computer software]. www.soscisurvey.de

Lunardo, R., & Saintives, C. (2013). The effect of naturalness claims on perceptions of food product naturalness in the point of purchase. *Journal of Retailing and Consumer Services, 20*(6), 529-537. https://doi.org/10.1016/j.jretconser.2013.05.006

Luomala, H., Jokitalo, M., Karhu, H., Hietaranta-Luoma, H. L., Hopia, A., & Hietamäki, S. (2015). Perceived health and taste ambivalence in food consumption. *Journal of Consumer Marketing*, *32*(4), 290-301. https://doi.org/10.1108/JCM-11-2014-1233

Magnusson, M. K., Arvola, A., Hursti, U. K. K., Åberg, L., & Sjöden, P. E. (2003). Choice of organic foods is related to perceived consequences for human health and to environmentally friendly behavior. *Appetite, 40*(2), 109-117. https://doi.org/10.1016/S0195-6663(03)00002-3

Mai, R., & Hoffmann, S. (2015). How to combat the unhealthy= tasty intuition: The influencing role of health consciousness. *Journal of Public Policy & Marketing*, *34*(1), 63-83.

Mai, R., Symmank, C., & Seeberg-Elverfeldt, B. (2016). Light and Pale Colors in Food Packaging: When Does This Package Cue Signal Superior Healthiness or Inferior Tastiness? *Journal of Retailing, 92*(4), 426–444. https://doi.org/10.1016/j.jretai.2016.08.002

Mai, R., Zahn, S., Hoppert, K., Hoffmann, S., & Rohm, H. (2014). Tailoring compensation effects of health-unrelated food properties. *Appetite, 80*, 143–153. https://doi.org/10.1016/j.appet.2014.05.014

Mizutani, N., Okamoto, M., Yamaguchi, Y., Kusakabe, Y., Dan, I., & Yamanaka, T. (2010). Package images modulate flavor perception for orange juice. *Food Quality and Preference, 21*(7), 867-872. https://doi.org/10.1016/j.foodqual.2010.05.010

Moussaïd, M., Brighton, H., & Gaissmaier, W. (2015). The amplification of risk in experimental diffusion chains. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(18), 5631–5636. https://doi.org/10.1073/pnas.1421883112

Nadricka, K., Millet, K., & Verlegh, P. W. (2020). When organic products are tasty: Taste inferences from an Organic= Healthy Association. *Food Quality and Preference*, *83*, 103896. https://doi.org/10.1016/j.foodqual.2020.103896

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Oakes, M. E., & Slotterback, C. S. (2005). Too good to be true: dose insensitivity and stereotypical thinking of foods' capacity to promote weight gain. *Food Quality and Preference, 16*(8), 675-681. https://doi.org/10.1016/j.foodqual.2005.03.010

Orquin, J. L. (2014). A Brunswik lens model of consumer health judgments of packaged foods. *Journal of Consumer Behaviour, 13*(4), 270-281. https://doi.org/10.1002/cb.1465

Papies, E. K., Barsalou, L. W., & Press, G. (2015). Grounding desire and motivated behavior: A theoretical framework and review of empirical evidence. In W. Hofmann, L. F. Nordgren (Eds.), *The psychology of desire* (pp. 36-60). Guilford Press.

Perkovic, S., Otterbring, T., Schärli, C., & Pachur, T. (2022). The perception of food products in adolescents, lay adults, and experts: A psychometric approach. *Journal of Experimental Psychology: Applied, 28*(3), 555–575. https://doi.org/10.1037/xap0000384

Pham, G. V., Shancer, M., & Nelson, M. R. (2019). Only other people post food photos on Facebook: Third-person perception of social media behavior and effects. *Computers in Human Behavior*, *93*, 129–140. https://doi.org/10.1016/j.chb.2018.11.026

Plessner, H., Freytag, P., & Fiedler, K. (2000). Expectancy-effects without expectancies: Illusory correlations based on cue-overlap. *European Journal of Social Psychology, 30(6)*, 837-851. https://doi.org/10.1002/1099-0992(200011/12)30:6<837::AID-EJSP19>3.0.CO;2-2

Prada, M., Saraiva, M., Viegas, C., Cavalheiro, B. P., & Garrido, M. V. (2022). Relationship between objective and perceived sugar content on consumers perceptions about breakfast cereals. *Food Quality and Preference*, 96, 104387. https://doi.org/10.1016/j.foodqual.2021.104387

Raghunathan, R., Naylor, R. W., & Hoyer, W. D. (2006). The unhealthy= tasty intuition and its effects on taste inferences, enjoyment, and choice of food products. *Journal of Marketing*, *70*(4), 170-184.

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. https://www.R-project.org/

Rebollar, R., Gil, I., Lidón, I., Martín, J., Fernández, M. J., & Rivera, S. (2017). How material, visual and verbal cues on packaging influence consumer expectations and willingness to buy: The case of crisps (potato chips) in Spain. *Food Research International*, *99*, 239-246. https://doi.org/10.1016/j.foodres.2017.05.024

Roininen, K., Lähteenmäki, L., & Tuorila, H. (1999). Quantification of consumer attitudes to health and hedonic characteristics of foods. *Appetite, 33*(1), 71-88. https://doi.org/10.1006/appe.1999.0232

Roininen, K., Tuorila, H., Zandstra, E. H., de Graaf, C., Vehkalahti, K., Stubenitsky, K., & Mela, D. J. (2001). Differences in health and taste attitudes and reported behaviour among Finnish, Dutch and British consumers: a cross-national validation of the Health and Taste Attitude Scales (HTAS). *Appetite, 37*(1), 33-45. https://doi.org/10.1006/appe.2001.0414

Rozin, P., Ashmore, M., & Markwith, M. (1996). Lay American conceptions of nutrition: dose insensitivity, categorical thinking, contagion, and the monotonic mind. *Health Psychology, 15*(6), 438. https://doi.org/10.1037/0278-6133.15.6.438

Rozin, P., Spranca, M., Krieger, Z., Neuhaus, R., Surillo, D., Swerdlin, A., & Wood, K. (2004). Preferences for natural: Instrumental and ideational/moral motivations, and the contrast between foods and medicine. *Appetite, 43*(2), 147-154. https://doi.org/10.1016/j.appet.2004.03.005

Russell, C. G., Worsley, A., & Campbell, K. J. (2015). Strategies used by parents to influence their children's food preferences. *Appetite, 90,* 123–130. https://doi.org/10.1016/j.appet.2015.02.038

Shankar, M. U., Levitan, C. A., Prescott, J., & Spence, C. (2009). The Influence of Color and Label Information on Flavor Perception. *Chemosensory Perception*, *2*(2), 53–58. https://doi.org/10.1007/s12078-009-9046-4

Shweder, R. A. (1977). Illusory correlation and the MMPI controversy. *Journal of Consulting and Clinical Psychology, 45*(5), 917–924. https://doi.org/10.1037/0022-006X.45.5.917

Spence, C., Harrar, V., & Piqueras-Fiszman, B. (2012). Assessing the impact of the tableware and other contextual variables on multisensory flavour perception. *Flavour*, *1*(1), 1–12. https://doi.org/10.1186/2044-7248-1-7

Steptoe, A., Pollard, T. M., & Wardle, J. (1995). Development of a measure of the motives underlying the selection of food: the food choice questionnaire. *Appetite*, *25*(3), 267-284. https://doi.org/10.1006/appe.1995.0061

Stroebe, W., Van Koningsbruggen, G. M., Papies, E. K., & Aarts, H. (2013). Why most dieters fail but some succeed: A goal conflict model of eating behavior. *Psychological Review, 120*(1)*,* 110-138. https://doi.org/10.1037/a0030849

Thomas, F., & Capelli, S. (2018). The effect of the number of ingredient images on package evaluation and product choice. *Recherche et Applications en Marketing (English Edition)*, *33*(3), 6-30. https://doi.org/10.1177/2051570718769201

Tuorila, H., & Cardello, A. V. (2002). Consumer responses to an off-flavor in juice in the presence of specific health claims. *Food Quality and Preference, 13*(7-8), 561-569. https://doi.org/10.1016/S0950-3293(01)00076-3

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(4), 327. https://doi.org/10.1037/0033-295X.84.4.327

Underwood, R. L., & Klein, N. M. (2002). Packaging as brand communication: effects of product pictures on consumer responses to the package and brand. *Journal of Marketing Theory and Practice, 10*(4), 58-68. https://doi.org/10.1080/10696679.2002.11501926

van der Heijden, A., te Molder, H., de Graaf, C., & Jager, G. (2020). Healthy is (not) tasty? Implicit and explicit associations between food healthiness and tastiness in primary school-aged children and parents with a lower socioeconomic position. *Food Quality and Preference, 84,* Article 103939. https://doi.org/10.1016/j.foodqual.2020.103939

Velasco, C., Salgado-Montejo, A., Marmolejo-Ramos, F., & Spence, C. (2014). Predictive packaging design: Tasting shapes, typefaces, names, and sounds. *Food Quality and Preference, 34,* 88-95. https://doi.org/10.1016/j.foodqual.2013.12.005

Visschers, V. H., Hess, R., & Siegrist, M. (2010). Health motivation and product design determine consumers' visual attention to nutrition information on food products. *Public Health Nutrition, 13*(7), 1099-1106. https://doi.org/10.1017/S1368980009993235

Wang, Q., Oostindjer, M., Amdam, G. V., & Egelandsdal, B. (2016). Snacks with nutrition labels: tastiness perception, healthiness perception, and willingness to pay by Norwegian adolescents. *Journal of Nutrition Education and Behavior*, *48*(2), 104-111. https://doi.org/10.1016/j.jneb.2015.09.003

Werle, C. O., Trendel, O., & Ardito, G. (2013). Unhealthy food is not tastier for everybody: The "healthy= tasty" French intuition. *Food Quality and Preference*, *28*(1), 116-121. https://doi.org/10.1016/j.foodqual.2012.07.007

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, *143*(5), 2020-2045. https://doi.org/10.1037/xge0000014

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Chapter 6

Bruner, J. S. (1990). *Acts of meaning: Four lectures on mind and culture*. Harvard university press.

Li, L.-J., & Kaulius, M. (2013). Judgement and decision making across cultures. *Advances in Psychological Science*, *21*(3), 381.

Potter, J., & Wetherell, M. (1987). *Discourse and social psychology: Beyond attitudes and behaviour*. Sage Publications, Inc.

Watzlawick, P., Bavelas, J. B., & Jackson, D. D. (1967). Pragmatics of human communication: a study of interactional patterns, pathologies, and paradoxes. New York: Norton.

Supplementary Material

Chapter 3.3.

Power Analysis

To determine the size of the participant sample, we conducted a power analysis by simulation (see, e.g., Brysbaert & Stevens, 2018), based on Slovic and Fischhoff's (1977) results with the same materials, assuming a conservative difference between foresight and hindsight (i.e., $p = .10$ compared to $p = .17$ in the original study). The estimated sample sizes required to reach at least .80 power in the planned mixed-effects models was 57 (probability judgments, both main study and follow-up), 68 (similarity judgments, only main study), and 46 (memory judgments, only follow-up). As we expected substantial attrition from the main study to the follow-up, we planned the main study to exceed these estimates.

Mixed-Effect Models

We used a linear mixed-effects model to analyze probability judgments. This allowed us to adequately represent the longitudinal repeated-measures design, use all collected data within one model despite of attrition at M3 and comprehensively report the multiple contrasts relevant to our hypotheses. Following the experimental design, we used the predictors order of instructions, measurement, and their interaction. Factors were dummy coded, treating hindsight-first order and M1 as the baseline. We indicate effect sizes using *ICC* for random effects and *d* for mixed-effects models (following Brysbaert & Stevens, 2018; Westfall et al., 2014).

Probability judgments. The linear mixed-effects model includes random intercepts for both participants and scenarios, as they are a sample of possible material we seek generalization from (Judd et al., 2017). It was implemented in lmer with the following code:

Judgment ~ Order*Measurement + (1|Participant) + (1| Scenario)

Table 1 summarizes the results of this model. At M1, probability judgments in hindsight (m = 58.85%) were on average 9.02% higher than in the foresight condition (m = 49.83%) replicating the classic between-participant finding. Moreover, the same hindsight effect was obtained regardless of whether the foresight condition was followed or preceded by the hindsight condition (see Figure 2). Given the hindsight-first ordering, the mean outcome probability estimate decreased by 9.63% from 58.85% at M1 to 49.22% at M2. This decline from hindsight to foresight constitutes a well-tuned downward adjustment, despite formerly induced outcome knowledge, matching the strength of the classic between-participants hindsight bias at M1. After two weeks, the adjustments made from M1 to M2 persisted at M3 when the last estimate was given in the hindsight condition (foresight-first order). In the hindsight-first order, the difference between M1 (hindsight condition) and M3 does not reach statistical significance anymore.

Metacognitive comparison to foresight and hindsight condition. We analyzed dichotomized self-classification data of judgments according to the foresight versus hindsight distinction as a function of scaled probability judgments and judgment conditions in the following logistic regression model: Similar To Hindsight ~ scale(Judgment) + Order*Condition. Due to variance in random effects approaching zero, they led to fit issues and were excluded from the model.

Increasing the probability judgment by one standard deviation makes it 1.84 times more likely that the participant responds, "close to hindsight" (see Table 2). However, after accounting for differences in judgments and order, the judgment condition (hindsight *vs*. foresight judgment) has no significant influence. Including the judgment condition into the model does not significantly contribute to model fit, $\chi^2(2) = 2.07$, *n.s*. To quantify this null result further, we computed the Bayes factor for the model comparison (Rouder et al., 2009): once again, the model without condition is preferred, $BF = 28.93 \pm 3\%$.

Meta-memory self-report. We analyzed the dichotomized responses as a function of scaled probability judgments and judgment conditions in a logistic regression model: Judgment ~ Memory Scale + (1|Scenario). The resulting model is summarized in Table 2. The metamemory self-report was dummy-coded, with the "new" response as the baseline. We removed participant random effects due to variance close to zero. Scenario random effects remained ($\tau = 44.94$, residual variance σ^2= 481.37, *ICC* = 0.09). Overall, the model was fit on 256 observations, with marginal / conditional R^2 = 0.124 / 0.199.

Table 1: Results of the Linear Mixed-Effects Model of Probability Judgments

	Probability Judgment			
Predictors	*Estimates*	*95% CI*	*p*	*d*
Intercept: Measurement 1, Order hindsight-first	58.85	51.27 – 66.43	**<.001**	**2.41**
Order: foresight-first	-9.02	-15.10 – -2.95	**.004**	**0.37**
Measurement 2	-9.63	-15.70 – -3.57	**.002**	**0.40**
Measurement 3	-4.44	-11.18 – 2.29	.197	0.18
Order: foresight-first * Measurement 2	22.23	13.65 – 30.81	**<.001**	**0.91**
Order: foresight-first * Measurement 3	17.73	8.31 – 27.15	**<.001**	**0.73**
Random Effects				
σ^2	431.13			
$\tau_{00\ \text{Participant}}$	128.18			
$\tau_{00\ \text{Scenario}}$	34.96			
ICC	0.27			
$N_{\text{Participant}}$	90			
N_{Scenario}	4			
Observations	488			
R^2	Marginal 0.051 / Conditional 0.311			

Note. Single-estimate *p*-values are based on the Satterthwaite approximation. Marginal and conditional R^2 statistics rely on Nakagawa and Schielzeth (2013), *d* on Westfall et al. (2014).

Table 2: Results of the Logistic Regression predicting Ratings on the Metacognitive Insight Scale

	Similar To Hindsight		
Predictors	*Odds Ratios*	*95% CI*	*p*
Intercept: Order hindsight-first, Hindsight Condition	0.88	0.55 – 1.39	.573
scale(Judgment)	1.84	1.43 – 2.37	**<.001**
Order foresight-first	0.46	0.24 – 0.90	**.024**
Foresight Condition	1.11	0.57 – 2.16	.766
Order foresight-first : Foresight Condition	1.48	0.58 – 3.79	.414
Observations	313		
Tjur's R^2	0.087		

Chapter 4.2.

Results of the simulation study can be generated analytically. Calculations start from an observed binomial sample of size n and outcome proportion P. To come up with an estimate of the true underlying proportion p, we consider all possible values of p and weight them by their likelihood, given the observed sample proportion P and sample size n. By default, we allow p to vary in the full range from 0 to 1. Yet in one simulated condition, we introduce a distinct restriction in the possible range of p to $p \geq \frac{1}{2}$. Such priors ensure that values outside this range are ignored when estimating p as a weighted average. Another simulated task condition imposes noise on sampled data. A proportion e of the sampled events is replaced by random outcomes generated by $p = \frac{1}{2}$.

Formally, the estimate $\hat{p}$ of the underlying true proportion p is a weighted average of all possible values of p weighted by their likelihood given sample size n and outcome proportion P:

$$\hat{p} = \frac{\int_{lb}^{1}[B(p'; n, P) * p]dp}{\int_{lb}^{1}[B(p'; n, P)]dp}$$

for $0 \leq p \leq 1$; $0 \leq nP \leq n$; $n \in \mathbb{N}$; $nP \in \mathbb{N}$

When noise is added, the expected probability becomes $p' = p(1-e) + \frac{e}{2}$;

$B(p; n, P) = \binom{n}{nP} p^{nP}(1-p)^{n-nP}$;

$\binom{n}{nP} = \frac{n!}{(nP)!(n-nP)!}$.

The parameter *lb* denotes the lower bound, which is 0 for most experimental conditions, but becomes $\frac{1}{2}$ for the case where the possible range of p is restricted to $\frac{1}{2} \leq p \leq 1$. The parameter e denotes the proportion of noise.

Chapter 5.2.

Cue Variables Used to Justify Levels of Healthiness and Tastiness with Results from Experiments 1 and 2

Cue	**Exp. 1** *Mentions*	% *responses to* *health*	*taste*	**Exp. 2** *Subset*	% used to infer *health*	*taste*	*Agree.*
Is … greasy.	308	**81**	19	1	90	56	74
Is … balanced.	138	**80**	20	1	97	22	88
Has a … pleasant texture.	131	4	**96**	1	12	79	95
Is … appetizing.	129	5	**95**	1	16	74	72
Has a … pleasant color.	122	14	**86**	1	23	48	78
Contains … vitamins.	121	**97**	3	2	**93**	17	86
Contains … nutrients.	96	**99**	1	2	**93**	**24**	82
Is … well-seasoned.	85	8	**92**	1	9	**96**	98
Contains … sugar.	78	**70**	30	2	83	30	71
Contains … processed ingredients.	77	**66**	34	2	66	15	97
Is … fresh.	73	42	**58**	1	**80**	**86**	85
Contains … well-matched ingredients.	72	29	**71**	2	10	**95**	99
Contains … carbohydrates.	66	**97**	3	2	58	21	64
Is … juicy.	59	3	**97**	1	15	81	69
Contains … proteins.	50	**96**	4	2	73	10	74
Is … aromatic.	35	0	**100**	1	9	93	90
Contains … different ingredients.	19	21	**79**	2	**71**	**64**	76
Is … salty.	18	39	**61**	1	64	79	78
Contains … calories.	17	**94**	6	2	58	23	77
Is … filling.	16	**75**	25	1	27	12	65
Is … hearty.	15	0	**100**	1	65	39	70
Contains … healthy fats.	13	**100**	0	2	**88**	13	87
Contains … vegetables.	7	**71**	29	2	87	20	81
Contains … ready-made products.	7	43	**57**	2	93	32	80
Contains … high-quality ingredients.	3	33	**67**	2	**82**	**50**	89
Is … spicy.	3	33	**67**	1	20	55	45
Contains … water.	3	**67**	33	2	24	37	59
Has a … pleasant temperature.	2	0	**100**	1	11	77	61

Cue	**Exp. 1**	% responses to		**Exp. 2**	% used to infer		
	Mentions	*health*	*taste*	*Subset*	*health*	*taste*	*Agree.*
Is … hard to digest.	2	50	50	1	85	16	87
Contains … meat.	1	**100**	0	2	57	31	64
Is … crispy.	1	0	**100**	1	**18**	**63**	70
Contains … grilled ingredients.	1	**100**	0	2	42	29	64
Added after Experiment 1:							
Contains … fresh ingredients.				2	87	69	83
Contains … fried ingredients.				2	92	21	79
Is … cooked through.				1	44	77	63
Is … well-prepared.				1	**22**	**98**	87

Note. The table is ordered by frequency of usage in Experiment 1. Cue variables in bold were used in Experiment 3. The overall number of mentions and the percentage of these mentions made as justifications for healthiness and tastiness judgments (Experiment 1) and the percentage of participants who use the cue to infer healthiness and tastiness (Experiment 2) are given. The percentage of participants who agreed on the level of healthiness or tastiness associated with the cue is given as agreement.

Information on Mixed-Effects Modelling

The fixed-effects structure of the presented mixed-effects model was pre-registered and reported without deletion (full model approach). Random effects were chosen by their contribution, deleting from a full model. Degrees of freedom are obtained via Satterthwaite's approximation. We report *ICC* for random effects and *d* for fixed effects (following Westfall et al., 2014) as effect size measures, in addition to marginal and conditional R^2 (Nakagawa & Schielzeth, 2013).

We analyzed the interrelation score with the linear mixed-effects model $\text{interrelation} = \text{poly}(\text{generation}, 2) * \text{cue set} + (1|\text{ID})$. The random effects contributed only marginally to model fit (*ICC* = 0.01); thus, only random intercepts for participants were kept. For the factor cue set, we pre-registered a contrast vector corresponding to the number and direction of overlapping cues in the set (overlap +- -> -1, separable -> 0, overlap ++ -> 1, overlap max. -> 5). We tested this model against a model with cue set dummy-coded with the Distinctive set as the baseline to validate this contrast choice, as per the pre-registered procedure. The theoretically motivated contrast vector fit the data to a comparable degree as the dummy coding, $\chi^2(4) = 3.45$, $p = .49$, and was kept as it is more parsimonious. We only planned for a linear relationship for the effect of generation, but this does not represent the descriptive data well (see Figure 1, Plot B). Therefore, we introduced orthogonal second-degree polynomials of generation, which significantly improved model fit, $\chi^2(2) = 26.49$, $p < .001$. As this complicates the interpretation of single slope estimates for generation in the model, the predictions from this model are shown in Figure 1, Plot C and D. The full model is presented in Table 3.

Table 3: Linear Mixed-Effects Model Predicting Interrelation Scores in Messages

Linear Mixed-Effects Model of Interrelation						
Predictors	*Estimates*	*95% CI*	*T*	*df*	*P*	*d*
Intercept	0.39	0.37 – 0.41	34.34	184.93	**<.001**	0.75
Generation	7.32	6.02 – 8.61	11.10	322.76	**<.001**	14.02
Generation²	-0.51	-1.77 – 0.74	-0.80	682.89	.424	0.98
Cue Set	0.03	0.02 – 0.04	7.33	3426.91	**<.001**	0.05
Generation * Cue Set	0.39	-0.06 – 0.84	1.70	3424.15	.089	0.75
Generation² * Cue Set	-0.97	-1.42 – -0.52	-4.21	3420.23	**<.001**	1.86
Random Effects						
σ^2	0.27					
$\tau_{00\ ID}$	< 0.01					
ICC	0.01					
N_{ID}	136					
Observations	3546					
Marginal R^2 / Conditional R^2	0.079 / 0.092					

Note. N includes the initial material (messages of 24 fictitious participants) received in generation 1. As these messages were generated randomly while maintaining a nonsignificant relationship between healthiness and tastiness, starting conditions for each chain varied slightly. The significant but extremely small main effect of cue set (b = 0.03, 0.02 – 0.04, $t(3426.91) = 7.33$, p <.001, d = 0.05) reflects these fluctuations, which we control for through this predictor.

Extremity

In addition, we analyzed the extremity of healthiness and tastiness descriptions in the messages. As the middle value of the 5-point response scale is anchored as *average*, we compute this by $extremity = |(value - 3)|$, ranging from 0 (average) to 2 (extremes).

The extremity of healthiness and tastiness values used in the produced messages did not change over generations ($m_{Extremity}$ = 1.06 – 1.19), between different cue sets ($m_{Extremity}$ = 1.08 – 1.10) or their interaction ($m_{Extremity}$ = 1.05 – 1.25). A mixed-effects model that predicts extremity with the same predictors as the main model showed very little explanatory value (marginal R^2 = .006, conditional R^2 = .033) and is therefore not discussed further. We find no evidence for systematic polarization or regression to the mean within the messages.

Attrition

Participants included virtually all available cues from their messages, 95.6%, excluding validation items and neutral cues. As attrition was rare and randomly distributed, we will not present further analyses.

Validation Sets. In participant group A and E (see Figure 1), we also included validation sets, which consisted of sixteen meal descriptions, each designed to validate the results (cue-concept associations) from Experiments 1 and 2. The responses to these messages were not transmitted to the following participants.

In one of the validation sets (in group E), participants received messages that described only the healthiness or the tastiness of a meal, with the other construct missing. Participants then used the four cue sets (one in each message) to translate the message. This allowed us to observe, for instance, whether participants used the cues we classified as tastiness cues when they translated tastiness alone. To test this, we computed the internal consistency between the tastiness cues in messages that translated tastiness levels, and for healthiness cues in messages that translated healthiness levels. Internal consistency was satisfactory overall (Cronbach's α M = .77), both for healthiness cues (α = .68 – .91) and tastiness cues (α = .67 – .83) in the four cue sets. Based on the consistency measure, the assignment of cues to constructs appears valid, while each cue is still treated as a separate aspect.

The results of this validation set (from group E) also support the high cue utilization assigned to cues in Experiment 2 and simultaneously reflect the cue overlap inherent in the sign system. When only the healthiness level was available, 94% of messages contained related cues (99% for distinctive cues and 89% for shared cues), with the same pattern for tastiness (96% overall, 99% distinctive, 93% shared cues). At the same time, 67,5% of messages also used cues not assigned to healthiness when only healthiness was given (65% for tastiness). Even cues we categorized as more distinctive offered very limited separability of healthiness and tastiness.

In the other validation set (group A), participants received messages that contained only one of the five cues available in each cue set. Based on this cue, they transmitted a message about the healthiness and tastiness of the described meal. Cues that we classified as distinctive always led to messages with statements about the concept they were assigned to, but 86% of messages based on cues we believed to be distinctive for healthiness after Experiments 1 and 2 contained statements about tastiness. In comparison, 75% of messages based on cues supposed to be distinctive for tastiness contain statements about healthiness.

Interestingly, this tendency to utilize any available cue even extends to the cue we categorized as neutral (uninformative regarding healthiness and tastiness) before the experiment. Most messages used the neutral cue in the presence of either healthiness (63%) or tastiness information (65%). The neutral cue presented alone leads to statements about healthiness and/or tastiness in 54% of messages within the validation set. In the transmitted messages, participants used the neutral cue in 63% of cases. This can be taken as a demonstration of the principle of vicarious functioning – in the absence of better cues, even the least informative of cues will be used.

List of Publications Included in this Thesis

The following articles are part of this publication-based thesis.

Fiedler, K., & **Salmen, K.**[9] (2021). Major theories in social psychology. In *Oxford Research Encyclopedia of Psychology*. Oxford University Press. https://doi.org/10.1093/acrefore/9780190236557.013.810

Salmen, K., Ermark, F. K. G., & Fiedler, K. (2023). Pragmatic, constructive, and reconstructive memory influences on the hindsight bias. *Psychonomic Bulletin & Review*, 30(1), 331–340. https://doi.org/10.3758/s13423-022-02158-1

Fiedler, K., **Salmen, K.**[10], & Prager, J. (2022). Toward stronger tests of rationality claims: Spotlight on the rule of succession. *Decision*, 9(3), 195–211. https://doi.org/10.1037/dec0000167

Salmen, K., Haasova, S., Florack, A. & Fiedler, K. (submitted). From observation to social transmission: How communication through cues shapes the construction of perceived healthiness and tastiness of food. *JEP: General*.

Contributions to publications without first authorship:

---

[9] K.S. contributed equally to conceptualization, writing and visualization.

[10] K.S. had a supporting role in conceptualization, methodology, data collection, analysis, and writing. K.S. took a lead role in visualization.

Declaration

in accordance to § 8 (1) c) and d) of the doctoral degree regulation of the Faculty

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

FAKULTÄT FÜR VERHALTENS- UND EMPIRISCHE KULTURWISSENSCHAFTEN

Promotionsausschuss der Fakultät für Verhaltens- und Empirische Kulturwissenschaften der Ruprecht-Karls-Universität Heidelberg / **Doctoral Committee of the Faculty of Behavioural and Cultural Studies of Heidelberg University**

Erklärung gemäß § 8 (1) c) der Promotionsordnung der Universität Heidelberg für die Fakultät für Verhaltens- und Empirische Kulturwissenschaften / **Declaration in accordance to § 8 (1) c) of the doctoral degree regulation of Heidelberg University, Faculty of Behavioural and Cultural Studies**

Ich erkläre, dass ich die vorgelegte Dissertation selbstständig angefertigt, nur die angegebenen Hilfsmittel benutzt und die Zitate gekennzeichnet habe. / I declare that I have made the submitted dissertation independently, using only the specified tools and have correctly marked all quotations.

Erklärung gemäß § 8 (1) d) der Promotionsordnung der Universität Heidelberg für die Fakultät für Verhaltens- und Empirische Kulturwissenschaften / **Declaration in accordance to § 8 (1) d) of the doctoral degree regulation of Heidelberg University, Faculty of Behavioural and Cultural Studies**

Ich erkläre, dass ich die vorgelegte Dissertation in dieser oder einer anderen Form nicht anderweitig als Prüfungsarbeit verwendet oder einer anderen Fakultät als Dissertation vorgelegt habe. / I declare that I did not use the submitted dissertation in this or any other form as an examination paper until now and that I did not submit it in another faculty.

Vorname Nachname / First name Family name	Karolin Salmen
Datum / Date	28.07.2023
Unterschrift / Signature	Dem Dekanat der Fakultät für Verhaltens- und Empirische Kulturwissenschaften liegt eine unterschriebene Version dieser Erklärung vom 28.07.2023 vor.