Biases in Sequential Decision Making:

Testing and Extending Coherence-Based Models of Information Integration and Search

Inaugural Dissertation

submitted in partial fulfillment of the requirements

for the degree of

Doctor philosophiae (Dr. phil.)

at the

Faculty of Human Sciences

University of Cologne

in accordance with the doctoral regulations

dated December 18, 2018

by

Carolin Häffner

from Bad Kreuznach

March 2026

Author Note

This dissertation was conducted within the framework of a research project funded by the German Research Foundation (DFG; Project No. 451253565), which also provided financial support for the author's doctoral position. The research within this project was conducted in close collaboration with Dr. Marc Jekel.

Additional support for a subset of studies reported in this dissertation (Chapter 5) was provided by a grant from the Center for Social and Economic Behavior (C-SEB), University of Cologne (Grant ID: Rd18-2024-JSUG-Häffner), awarded to the author.

The dissertation is submitted to Prof. Dr. Andreas Glöckner and Prof. Dr. Angela Dorrough for review.

Diese Dissertation wurde von der Humanwissenschaftlichen Fakultät der Universität zu Köln im Juli 2026 angenommen.

Abstract

Many real-world decisions unfold sequentially: information is acquired step by step while emerging preferences systematically bias the evaluation and selection of evidence. Coherence-based models provide a powerful framework for explaining such biases by proposing that preferences, beliefs, and information search mutually constrain one another. However, their specific implications for sequential decision environments and for integrating biases in information evaluation and search remain insufficiently specified. Across three articles, this dissertation examines and extends a sequential coherence-based account of biased decision making. First, it shows that process models must be translated into statistical predictors that preserve their implied dynamics to ensure that statistical tests are informative about the validity of the underlying process assumptions—a requirement that is typically not met in research using the stepwise evolution-of-preference paradigm. Second, the dissertation advances theory integration by embedding partisan motivated reasoning within the same coherence framework that accounts for predecisional information distortion, showing that ideological priors function as top-down constraints within a dynamic coherence-based updating process. Third, the framework is extended to information search, integrating preference-driven and preference-confirming search within the same sequential architecture. Taken together, the findings demonstrate that clarifying the mapping from coherence-based process models to statistical predictions and integrating biases in information evaluation and search as manifestations of a common coherence-driven process advances theory construction and integration on biased decision making.

Keywords: theory construction, theory integration, sequential decision making, coherence-based models, biases; predecisional information distortion, information search

Zusammenfassung

Viele reale Entscheidungen verlaufen sequenziell: Entscheidungsrelevante Informationen werden schrittweise verarbeitet, während vorläufige Präferenzen systematisch die Bewertung und Suche neuer Informationen beeinflussen. Kohärenzbasierte Modelle bieten einen einflussreichen Ansatz zur Erklärung solcher Verzerrungen, indem sie die wechselseitige Beeinflussung von Präferenzen, Überzeugungen und Informationssuche annehmen. Ihre Implikationen für sequenzielle Entscheidungssituationen sind bislang jedoch nur unzureichend spezifiziert. Die vorliegende Dissertation untersucht und erweitert ein kohärenzbasiertes Modell sequenzieller Entscheidungsprozesse anhand von drei empirischen Artikeln. Erstens zeigt sie, dass Prozessmodelle in statistische Prädiktoren übersetzt werden müssen, die ihre zugrunde liegenden Dynamiken adäquat abbilden. Analysen prädezisionaler Informationsverzerrung legen nahe, dass gängige paradigmabasierte Auswertungen zentrale modellimplizierte Dynamiken verdecken können. Zweitens integriert die Dissertation parteimotiviertes Denken in denselben Kohärenzrahmen, der prädezisionale Informationsverzerrung erklärt, und zeigt, dass ideologische Überzeugungen als Top-down-Einflüsse innerhalb eines dynamischen, kohärenzbasierten Aktualisierungsprozesses wirken. Drittens erweitert sie diesen Ansatz auf den Prozess der Informationssuche, indem präferenzgetriebene und konfirmatorische Suchmechanismen innerhalb desselben kohärenzbasierten Prozesses integriert werden. Insgesamt verdeutlichen die Befunde die Notwendigkeit einer präziseren Ableitung der statistischen Implikationen kohärenzbasierter Prozessmodelle für die Theorieentwicklung. Darüber hinaus zeigen sie, dass Verzerrungen in der Informationsintegration und -suche als kontextabhängige Ausprägungen eines gemeinsamen kohärenzorientierten Prozesses in sequenziellen Entscheidungssituationen verstanden werden können.

Preface

This dissertation is an integrated cumulative thesis comprising three empirical manuscripts that examine coherence-based models of judgment and decision making, with a focus on information integration and information search in sequential decision contexts. All three manuscripts included in this dissertation are currently submitted for publication.

The dissertation is organized such that the three manuscripts are embedded as core chapters (Chapters 4–6) and are framed by an overarching theoretical introduction and a general discussion. These chapters jointly develop, test, and extend coherence-based models of sequential decision making to account for multiple reasoning and search biases within a unified theoretical framework.

Manuscript I (Chapter 4)

Häffner, C., Jekel, M., & Lisovoj, D. (2026). *From conceptual model to statistical test: Paradigm blindness in research on predecisional information distortion.* Manuscript under review.

The theoretical framework was developed jointly by all authors. Experimental programming and data collection were conducted by Daria Lisovoj and me. Data analyses for the main studies were conducted jointly by Daria Lisovoj and me, whereas all re-analyses of published datasets were conducted by me. The manuscript was written jointly by me (leading role) and Marc Jekel (supporting role).

Manuscript II (Chapter 5)

Häffner, C., Sperlich, L., & Lammers, J. (2026). *Top-down intrusions on coherence-driven distortion: Rethinking partisan motivated reasoning.* Manuscript submitted for publication.

The theoretical framework was developed jointly by me (leading role) and Joris Lammers (supporting role). Experimental programming was conducted by Lea Sperlich. I conducted all data analyses and prepared the first full draft of the manuscript, which was subsequently revised in collaboration with the co-authors.

Manuscript III (Chapter 6)

Häffner, C., Jekel, M., & Lisovoj, D. (2025). *Do I search for what I like or what keeps me liking? Disentangling attraction search vs. confirmatory search in the stepwise evolution-of-preference paradigm.* Manuscript in revision.

The theoretical framework was developed jointly by all authors. Experimental programming was conducted by Daria Lisovoj. Data analyses were conducted jointly by Daria Lisovoj and me. I prepared the first full draft of the manuscript, which was revised in collaboration with the co-authors.

All studies reported in this dissertation were preregistered prior to data collection. Materials, anonymized data, and analysis scripts are publicly available via the Open Science Framework (OSF), with study-specific links provided in the appendices of the respective chapters.

Figures and tables within each chapter retain the numbering of the original manuscripts to preserve consistency with the submitted versions.

Acknowledgments

Now that the day has come to finally submit my dissertation, I find myself thinking of Momo, the street sweeper:

You must never think of the whole street at once, understand? You must only concentrate on the next step, the next breath, the next stroke of the broom, and the next, and the next. Nothing else.

— *Momo*

What a road it has been to get here. And I would not be standing at this point without a number of people—far more than I can name here.

Even though I had always enjoyed research, I never seriously considered pursuing a PhD. I simply started working on my Master's thesis with you, Marc—and somehow, one step at a time, it grew into more than five years of working together. First and foremost, thank you, Marc, for your enthusiasm and dedication to research, and for always believing in me—even at times when I was ready to drop the broom. I would not be here without you.

Thank you, Andi, for your critical and insightful feedback throughout the process, and for your steady support—especially in the final months. I am very grateful that you always took the time for my many questions, both professional and personal. Thank you, Angie, for being part of this journey from early on—both during my studies here in Cologne and as part of the team. I still remember the concert we went to together at the very beginning. Thank you for your support along the way, and for taking the time to review my thesis. Thank you as well to Joris and Lea for trusting my research ideas and for such enjoyable collaboration—I truly hope there is more to come. And a very special thank you to Daria, for being the most supportive and invaluable project partner I could have wished for.

I also want to thank the entire SoCCCo team for being the warmest and most supportive group one could wish for. Even if I spent far too much money at the gas station (due to my very poor meal-prepping skills), I will always miss our lunch breaks. Special thanks to Lea, Johanna, and Vanessa for your constant emotional support—and for always reminding me of my strengths when I tended to doubt them.

Another special thank you goes to my closest friends: Thank you, Ronja, for always being there—even while navigating life with a newborn (and yes, one should not compare oneself to you!). Thank you, Lui, for always—truly always—making space for me, and for being there through both the joyful moments and the difficult days. And thank you, Kaddy, Stefan and Daniel for your constant and unwavering support from afar—and Kaddy, for sharing this path with me from school to now our PhDs. Thank you, Rike, Rebecca, and Marie for walking this path alongside me since our very first week at university. And thank you to the Tabata girls for the constant laughter and support, and for making me enjoy a routine for the first time in my life.

I want to thank my family, especially my parents. Thank you, Mum, for your support, and thank you, Dad, for doing everything you could to make my adult life compatible with this project—and for being the greatest supporter throughout my life. I am also deeply grateful to my grandparents—it means so much to me to have them all here to share this moment.

Thank you, Tim, who entered my life right in the middle of this journey and changed it in the best possible way. Thank you for your incredible support in the final months, for grounding me when I needed it most, and for making me feel safe and at home.

And finally, I want to thank my best friend, Marcel—without whom I might never have had the courage to begin this journey. I miss you.

Table of Contents

General Introduction

Imagine deciding which of two proposed policy reforms to support. Both options aim to address the same societal problem but differ in their expected economic efficiency, social consequences, and political feasibility. Initial information may tentatively favor one reform, yet further considerations emerge only gradually—through reports, discussions, or reflection. As additional information becomes available, emerging preferences begin to shape how the options are evaluated and which aspects receive further attention.

Relevant information is rarely encountered simultaneously. Instead, most real-world decisions unfold sequentially as individuals repeatedly integrate new evidence and decide what to consider next, often while a tentative preference is already forming. Understanding how information integration and search unfold under such conditions is therefore central to research on judgment and decision making.

These sequential processes provide the context in which many well-known biases in evaluation and information search are instantiated. Over the past half century, a substantial body of experimental research has documented systematic deviations from normative inference—often described as the "error paradigm" (Funder, 1995; see also Arkes, 1991; Klayman, 1995; Nickerson, 1998). Although this work has generated an extensive catalogue of bias effects, theoretical consolidation has progressed more slowly than empirical accumulation (Eronen & Bringmann, 2021; Gigerenzer & Gaissmaier, 2010; Meehl, 1967). Many biases are characterized at the level of observed outcomes without being embedded in a shared account of the cognitive processes that generate them—a gap that has motivated recent calls for more integrative, process-level frameworks (Glöckner & Betsch, 2011; Simon & Read, 2025).

Coherence-based models of reasoning provide a process-oriented response to this fragmentation. Building on parallel constraint satisfaction frameworks (Glöckner et al., 2014; Glöckner & Betsch, 2008a; Holyoak & Simon, 1999), they conceptualize judgment as a dynamic network in which beliefs, preferences, and pieces of evidence mutually constrain one another until a coherent configuration emerges. Within this perspective, biases are interpreted as emergent consequences of coherence-seeking dynamics rather than as distinct mechanisms. Simon and Read (2025) articulate this implication explicitly by proposing coherence-based reasoning as a general process architecture for biased reasoning. Yet if coherence-based reasoning is to serve as such an architecture, its implications must not only integrate phenomena conceptually but also generate empirically testable predictions that capture the model's core dynamics. This requires translating conceptual process assumptions into statistical predictors that instantiate these dynamics in empirical analyses.

An important bottleneck arises at this stage. Even when conceptual process models specify how preferences, beliefs, and evidence mutually constrain one another, their implications are not always systematically translated into statistical predictors that preserve these dynamics. When hypothesis specification instead follows paradigm-specific conventions, empirical tests may fail to capture the processes implied by the conceptual model.

Sequential decision paradigms make this bottleneck particularly visible. Predecisional information distortion—the tendency to evaluate incoming information in favor of a currently leading option (Russo et al., 1996)—emerges during stepwise information integration and therefore provides a natural setting for examining preference–evaluation coupling. Although coherence-based explanations have been proposed (e.g. DeKay, 2015) standard analyses often operationalize distortion as a function of prior preference alone, leaving key model-implied dynamics only partially examined.

Across three empirical articles, the dissertation examines and extends a sequential coherence model of biased decision making. Chapter 4 addresses a problem of theory construction by systematically deriving model-implied predictors that instantiate the bidirectional updating between preferences and evaluations implied by coherence-based accounts, thereby restoring the diagnostic value of empirical tests for evaluating the model. Chapter 5 advances theory integration by incorporating partisan beliefs as priors within the coherence network, linking belief-based bias to preference construction within a unified architecture. Chapter 6 further extends the framework to information search, examining attraction search—seeking information about the currently preferred option—and confirmatory search—favoring preference-confirming over preference-disconfirming information—within the same sequential decision process.

The remainder of this dissertation is organized as follows. Chapters 1 to 3 develop the theoretical background, reviewing major approaches to judgment and decision making and outlining coherence-based accounts of biased information integration and search. In doing so, they introduce the key phenomena that motivate the empirical investigations. Chapters 4 to 6 present three empirical articles that test and extend this framework. Finally, Chapter 7 provides a general discussion of the main findings, theoretical contributions, and directions for future research.

Chapter 1: Theoretical Foundations of Coherence-Based Decision Making

Research on judgment and decision making has produced a wide range of theoretical accounts that differ in how they conceptualize the formation of preferences and the integration of information during a decision process. Some theoretical approaches assume relatively stable, pre-existing preferences, such that information is evaluated with respect to fixed option values (Brunswik, 1955; Luce, 2000; Luce & Raiffa, 1957). Other approaches instead conceptualize preferences as emerging dynamically during the evaluation of evidence, such that the interpretation of information and the formation of preferences co-evolve during the decision process (Glöckner & Betsch, 2008a; Holyoak & Simon, 1999). These differences become particularly consequential when decisions unfold sequentially, as individuals repeatedly integrate new pieces of information while tentative preferences are already emerging. In such settings, preferences are not simply the outcome of information integration but may themselves influence how subsequent information is interpreted and acquired.

Understanding how preferences and information interact during a decision therefore requires theoretical frameworks that specify the cognitive processes by which information is integrated and preferences evolve over time. However, many influential models focus primarily on predicting decision outcomes rather than the processes that generate them. Such outcome models aim to predict choices or judgments without specifying the cognitive operations through which these outcomes arise (Glöckner & Betsch, 2011). While this perspective has yielded powerful accounts of decision outcomes, it offers limited insight into how information is evaluated, integrated, and searched during the unfolding of a decision.

Process models, in contrast, explicitly articulate assumptions about the cognitive operations through which information is retrieved, evaluated, and integrated during a decision episode. By specifying such mechanisms, these models generate predictions not only for choices but also for variables of the decision process, including response times, confidence judgments (Glöckner et al., 2014; Glöckner & Betsch, 2008a), and patterns of information acquisition (Jekel et al., 2018; Payne et al., 1988). Theoretical work has increasingly emphasized the importance of such process-level explanations, noting that the large number of documented biases in judgment and decision making often remains insufficiently integrated within a common account of the processes that generate them (Simon & Read, 2025). Because the present dissertation examines biases in information evaluation and information search as they unfold during sequential decision making, it focuses on theoretical approaches that explicitly model these underlying decision processes.

Within the broader class of process models, theoretical approaches differ in how they conceptualize the mechanisms underlying decision behavior. A central distinction concerns whether variation in behavior is explained through the selection among qualitatively distinct strategies or through parameter adjustments within a single underlying mechanism of information integration (Glöckner et al., 2014; Kruglanski, 2013). These perspectives correspond to two influential classes of process models: multi-strategy accounts and single-mechanism approaches.

Multi-strategy accounts explain decision behavior by assuming that individuals rely on a repertoire of distinct strategies that are selectively applied depending on environmental structure and task demands (Beach & Mitchell, 1978; Payne et al., 1993). A prominent formulation of this view is the *adaptive toolbox*, according to which decision makers possess a set of simple heuristics tailored to particular classes of problems (Gigerenzer et al., 2000; Gigerenzer & Gaissmaier, 2010). These heuristics typically specify rules for information search, stopping, and choice. For example, lexicographic heuristics such as *Take-the-Best* assume that information is searched in order of cue validity and that search terminates once a discriminating cue is found (Gigerenzer & Goldstein, 1996). Although such models provide parsimonious accounts of behavior in certain environments, critics have noted several limitations. In particular, empirical work has raised questions about whether decision makers reliably rely on distinct heuristics in the way multi-strategy frameworks assume, as individuals often continue to search for and integrate information even in situations where simple heuristics would suffice (Bröder & B. R. Newell, 2008). Beyond these empirical concerns, multi-strategy accounts have also been criticized on theoretical grounds. Without tightly constrained sets of strategies and explicit selection mechanisms, the framework risks limited falsifiability because deviations from predicted behavior can be accommodated by positing additional heuristics or context-dependent selection rules (Glöckner & Betsch, 2011). Although recent Bayesian approaches have improved the formal testability of toolbox models (Scheibehenne, Rieskamp, & Wagenmakers, 2013), the broader question remains whether variation in decision behavior requires multiple distinct mechanisms at all.

In contrast, single-mechanism models assume that decisions arise from a common process of information integration that operates across tasks, with variation in decision behavior emerging through changes in parameters within the same underlying mechanism (Glöckner et al., 2014). Evidence accumulation models represent one influential class of such approaches and have dominated the modeling of decision processes for several decades (Evans & Wagenmakers, 2020; Ratcliff et al., 2016). Across different implementations, these models conceptualize decision making as a stochastic process in which evidence favoring competing options is gradually accumulated over time until a decision threshold is reached, at which point a response is initiated (Busemeyer & Townsend, 1993; Ratcliff, 1978).

A key feature of evidence accumulation models is that variation in behavior is explained through parameter adjustments within the same accumulation process—for example, through differences in drift rates, response thresholds, or starting points (Lee & Cummins, 2004; Ratcliff et al., 2016). This parameter-based flexibility has been characterized as an "adjustable spanner," reflecting the idea that a single cognitive mechanism can account for variation in decision behavior across tasks without invoking qualitatively distinct strategies (B. R. Newell, 2005). As a result, evidence accumulation models have proven highly successful in explaining both choice behavior and response time distributions across a wide range of experimental paradigms (Brown & Heathcote, 2008; Ratcliff & Tuerlinckx, 2002).

However, despite their differences, many established decision models share a common simplifying assumption: although information influences preferences, the evaluative meaning of information itself remains stable throughout the decision process. This assumption of unidirectional reasoning—from information to option evaluations—is characteristic of a wide range of models and implies that emerging preferences do not systematically influence how subsequent information is interpreted (DeKay, 2015; Glöckner et al., 2014). Empirical research on predecisional information distortion challenges this assumption by showing that the evaluation of information can shift during the decision process in ways that support an emerging preference (Carlson & Russo, 2001; Russo et al., 1996). These findings suggest that preferences and evaluations may interact dynamically prior to choice (Glöckner et al., 2010a; Holyoak & Simon, 1999).

Addressing such dynamics requires models that explicitly allow for bidirectional interactions between emerging preferences and the evaluation of incoming information. Coherence-based models of decision making provide a theoretical framework that explicitly captures such dynamics. Building on connectionist theories of parallel constraint satisfaction (McClelland & Rumelhart, 1981; Thagard, 1989, 2000), these models conceptualize judgment as the outcome of a network process in which beliefs, preferences, and pieces of evidence mutually constrain one another until a coherent interpretation of the decision problem emerges (Glöckner & Betsch, 2008b; Holyoak & Simon, 1999). Because this architecture allows emerging preferences to influence ongoing information evaluation and information search, coherence-based models provide the theoretical foundation for the empirical investigations presented in this dissertation.

The following sections therefore introduce coherence-based models of decision making, beginning with their connectionist foundations before turning to parallel constraint satisfaction accounts that formalize coherence maximization as a general mechanism of judgment and decision making.

1.1. Connectionist models

The idea that human cognition strives for internally coherent representations has a long tradition in psychology. Early formulations can be traced back to Gestalt psychology, which emphasized the tendency to organize perceptual and conceptual input into parsimonious and stable wholes (e.g., Kohler & Emery, 1947; Wertheimer, 1928/1938). Related ideas were later taken up in social psychology, most prominently in theories of cognitive balance and dissonance, which conceptualized beliefs and evaluations as interdependent elements within a system seeking internal equilibrium (Festinger, 1957a; Heider, 1946, 1958; Osgood & Tannenbaum, 1955).

A central assumption shared by these approaches is that inconsistency among cognitive elements generates tension, which in turn triggers reorganization until a coherent interpretation is achieved. Coherence is therefore not achieved by passively reading off information from the environment but by actively restructuring how information is interpreted (Festinger, 1957a; Heider, 1958). Classic demonstrations from perception, such as ambiguous figure–ground stimuli, illustrate that identical objective input can give rise to qualitatively different yet internally consistent interpretations, depending on how the elements are organized (Rubin, 1921).

For a long time, however, these ideas remained largely verbal and metaphorical. A major theoretical advance occurred with their formalization in connectionist models in the 1980s (McClelland et al., 1986; Thagard, 1989). Connectionist models represent cognition as a network of interconnected units, where nodes correspond to cognitive elements (e.g., beliefs, cues, options) and weighted connections encode relations of support or conflict. Activation spreads bidirectionally through the network, and processing unfolds as a dynamic, parallel adjustment of activation levels until the system settles into a stable state that maximizes overall coherence (Bröder & Scharf, 2025; McClelland & Rumelhart, 1981).

This formalization made it possible to extend consistency-based reasoning from simple dyadic relations to complex decision situations involving multiple, mutually constraining pieces of information (Read et al., 1997). Parallel constraint satisfaction networks have since been successfully applied across a wide range of domains, including perception, impression formation, analogy, legal reasoning, and preferential choice (Holyoak & Simon, 1999; Holyoak & Thagard, 1989; Kunda & Thagard, 1996; Simon, Krawczyk, et al., 2004).

Crucially for judgment and decision making, connectionist models predict that coherence is constructed during the decision process. As preferences begin to emerge, information supporting the favored option becomes increasingly weighted, whereas inconsistent information is downplayed or reinterpreted. Empirical work has shown that such coherence shifts occur prior to choice, operate outside of awareness, and play an operative role in shaping decisions rather than reflecting post hoc rationalization (Simon, 2004; Simon & Holyoak, 2002).

Taken together, connectionist models provide a formal framework for understanding decision making as a dynamic, bidirectional process in which information and preferences mutually constrain one another. This perspective directly challenges unidirectional accounts of decision making and lays the conceptual foundation for parallel constraint satisfaction models, which explicitly specify coherence maximization as a general mechanism underlying judgment and decision making.

1.2. Parallel Constraint Satisfaction Model of Decision Making (PCS-DM)

Parallel Constraint Satisfaction (PCS) models constitute a formal, process-based instantiation of coherence-based reasoning in judgment and decision making. A central contribution is the PCS model proposed by Glöckner & Betsch (2008a), which specifies how decisions emerge from the dynamic interaction of decision-relevant information within a connectionist network. In this framework, cues, attributes, and decision options are represented as nodes connected by weighted links reflecting relations of support or conflict (Holyoak & Simon, 1999).

Within PCS models, information integration unfolds via an iterative spreading-activation process. Activation propagates bidirectionally through the network, such that mutually consistent elements are strengthened while inconsistent elements are weakened. Competing options are linked by inhibitory connections, whereby increasing activation of one option suppresses activation of alternatives. Over successive iterations, the network converges toward a stable activation pattern that maximizes global coherence. This final state predicts the selected option, the degree of confidence (as a function of activation differences between options), and decision time (as a function of the number of iterations required for convergence; Glöckner & Betsch, 2008a).

Early PCS formulations were criticized for being insufficiently specified. While the dynamics of information integration were formally described, the mapping from task characteristics to network parameters was initially underspecified, rendering the models potentially too flexible (Marewski, 2010; see also Glöckner & Betsch, 2010, 2011). Subsequent work provided increasingly specific tests of model-implied process-level predictions derived from PCS, including systematic effects of coherence on response times and confidence in memory-based decisions (Glöckner & Hodges, 2011) and demonstrations that decision time depends on coherence rather than merely on the amount of available information (Glöckner & Betsch, 2012). These developments culminated in the fully specified Parallel Constraint Satisfaction Model of Decision Making (PCS-DM), which introduces explicit transformation functions from cue validities to network structure and formal links between network dynamics and multiple behavioral measures (Glöckner et al., 2014).

Building on these specifications, PCS-DM has been shown to generate reliable and converging predictions across a range of decision contexts. Empirical work demonstrates that the model accounts for preference-consistent changes in information evaluation in probabilistic inference tasks (Glöckner et al., 2010) and generalizes across different environmental structures, including variations in cue validity and information distribution (Glöckner et al., 2014). Beyond such controlled settings, PCS has also been successfully applied to complex, dynamic decision environments, such as expert decision making in sports, where model predictions can be linked to gaze behavior and option generation (Glöckner et al., 2012). Importantly, these findings extend beyond the prediction of choice outcomes and instead provide process-level constraints that allow PCS-DM to be tested against competing models. Joint tests of model predictions have generally favored PCS-DM over competitors from the adaptive toolbox in controlled environments, particularly when models are evaluated across systematically varied task structures (Glöckner et al., 2014; Glöckner & Betsch, 2012).

A distinctive prediction of PCS-DM is the occurrence of *coherence shifts*, that is, systematic changes in the subjective weighting of information during the decision process (Glöckner & Betsch, 2008a). Closely related phenomena have been studied under labels such as *predecisional information distortion* (Carlson & Russo, 2001; Russo et al., 1996) or *biased assimilation* (Schmittat & Englich, 2016). These effects are typically assumed to operate largely outside of conscious awareness (Simon, 2004) and are theoretically grounded in accounts of explanatory coherence, which posit that cognitive systems actively restructure information to support a coherent interpretation of the decision situation (Thagard, 1989, 2000). Empirically, coherence-driven distortions have been demonstrated repeatedly across judgment and decision-making paradigms (Glöckner et al., 2010; Lee & Holyoak, 2021; Simon et al., 2001). These effects are discussed in greater detail in Chapter 2.1.1.

Despite its strengths, PCS-DM has a limitation that is central for subsequent developments of coherence-based theory. The model specifies information integration only for situations in which all relevant information is assumed to be readily available. As a consequence, it does not generate predictions about information acquisition—namely, how decision makers select which information to sample when information is not fully observable or involves search costs (Marewski, 2010; Söllner et al., 2013). This restriction complicates direct theoretical and empirical comparisons with heuristic models in domains where information acquisition plays a central role.

Subsequent work has addressed this limitation by extending coherence-based modeling to situations in which decision makers must actively acquire information. The integrated coherence-based decision and search framework (iCodes; Jekel et al., 2018; see also Bröder & Scharf ,2025, for a review) builds directly on the PCS architecture but incorporates information search into the same coherence-maximization process that governs information integration.

1.3. The Integrated Coherence-Based Decision and Search Model (iCodes)

Building on the formalization of coherence-based information integration in PCS-DM, the iCodes framework extends coherence-based modeling to environments in which decision makers must actively search for information. While PCS-DM describes how preferences and evaluations co-evolve during information integration, iCodes embeds information search within the same constraint-satisfaction dynamics (Jekel et al., 2018). Emerging preferences therefore influence not only how information is evaluated but also which information is searched for next. As in PCS-DM, processing unfolds through iterative spreading activation within a connectionist network linking options, cues, and available cue values (Glöckner et al., 2014; Jekel et al., 2018). Crucially, only available cue values participate in bidirectional activation. Concealed cue values are connected to the network via unidirectional links: they receive activation from option and cue nodes but do not feed activation back into the network, reflecting that their content is not yet known (Jekel et al., 2018).

Within this architecture, information search is not modeled as a separate decision strategy. Instead, the selection of the next information item emerges from the current coherence state of the network. Because the activation of concealed cue values depends jointly on cue validities and on the current attractiveness of the decision alternatives, the model predicts that information about the currently more attractive option becomes more likely to be searched than information about less attractive alternatives, a pattern referred to as the Attraction Search Effect (Jekel et al., 2018). In this way, information search becomes sensitive to the evolving preference state, in contrast to traditional search models that assume fixed cue-wise or option-wise search orders independent of previously acquired evidence (e.g. Gigerenzer & Goldstein, 1996; Lee & Cummins, 2004; Payne et al., 1988). Because search decisions arise from the same iterative activation process as evaluative integration, the model also generates predictions for the temporal dynamics of information search, using the number of iterations to convergence as a proxy for reaction times (Jekel et al., 2018). A detailed discussion of attraction-based and other search biases is deferred to Chapter 3.

Despite these advances, most coherence-based models—including PCS-DM and iCodes—have primarily been developed and tested in settings in which information is either presented simultaneously or remains accessible throughout the decision process (e.g., Glöckner et al., 2014; Holyoak & Simon, 1999; Jekel et al., 2018; Simon et al., 2001). By contrast, many real-world decisions unfold sequentially as information becomes available step by step.

Understanding how coherence dynamics operate under such sequential conditions therefore provides the point of departure for the remainder of this dissertation.

The following chapters outline key biases in both information evaluation and information search and examine how these phenomena can be understood within coherence-based frameworks when decisions unfold sequentially. Chapter 2 focuses on biased information integration and discusses how coherence-based models account for systematic distortions in the evaluation of subsequently encountered information. Chapter 3 then turns to biases in information search and evaluates coherence-based accounts of how emerging preferences shape the selection of information during sequential decision making.

Chapter 2: A Coherence-based Approach to Biased Information Integration in Sequential Decision Making

This chapter reviews evidence for systematic biases in how information is evaluated during decision making and discusses how these biases can be understood from a coherence-based perspective when information is encountered sequentially. The focus is not on providing an exhaustive review of reasoning biases, but on highlighting those phenomena that are most relevant for the coherence-based framework developed in this dissertation.

The chapter proceeds in three steps. First, it introduces prominent descriptive biases in information integration that have been documented across a wide range of reasoning tasks. Second, it outlines the stepwise evolution-of-preference paradigm as a methodological framework that renders dynamic coherence processes observable over time. Third, it reviews coherence-based models of information integration and discusses their limitations in accounting for sequential integration dynamics.

2.1 Descriptive Biases in Information Integration

A large body of research has documented systematic biases in how individuals evaluate information during reasoning and decision making. Over the past several decades, experimental work has identified numerous deviations from normative models of information processing, often discussed under the broad label of biased reasoning (see Arkes, 1991; Klayman, 1995; Nickerson, 1998). Two phenomena are particularly relevant for the present dissertation. First, research on predecisional information distortion shows that individuals tend to evaluate newly encountered information in a way that favors the currently leading option during the decision process. Second, work on motivated reasoning demonstrates that information evaluation can also be shaped by prior beliefs or attitudes. The following sections briefly review these two lines of research before turning to process-based accounts of biased information integration.

2.1.1 Predecisional Information Distortion

For a long time, systematic bias during the predecisional phase of decision making was widely assumed not to exist. The prevailing assumption was that, prior to reaching a decision, individuals could not yet know which alternative to favor, rendering predecisional distortion conceptually implausible (Frey, 1986; Russo, 2015). Accordingly, biases in information evaluation were primarily attributed to prior beliefs or commitments that shape how evidence is interpreted.

Research on predecisional information distortion (hereafter information distortion) challenged this view by demonstrating that systematic bias can emerge dynamically during the decision process itself as preferences begin to form. Information distortion refers to the tendency to evaluate information in ways that favor the currently leading alternative, such that information about the preferred option is interpreted more positively than information about competing alternatives (Blanchard et al., 2014; Russo, 2015). Crucially, this bias emerges before a final decision has been made and does not require prior beliefs or attitudes toward the alternatives, thereby distinguishing information distortion from belief-driven biases such as motivated reasoning or belief bias (Russo, 2015).

Empirical evidence for information distortion has been documented across a wide range of decision domains and populations. Distortion effects have been observed in professional decision making among auditors and entrepreneurs (Beeler & Hunton, 2002; Russo et al., 2000), in legal contexts where prospective jurors distort the interpretation of evidence in the direction of an emerging verdict (Carlson & Russo, 2001; Engel et al., 2020; Holyoak & Simon, 1999; Simon, Snow, et al., 2004; Thagard, 2003), and in medical decision making, where diagnostic information is interpreted in line with an emerging diagnosis (Kostopoulou et al., 2017; Nurek et al., 2014; Nurek & Kostopoulou, 2016). Similar effects have been demonstrated in person perception and advice taking, where individuals distort behavioral information and advice in line with stereotypes (Dorrough et al., 2017) or initial impressions (Kunda & Thagard, 1996), as well as in risky and preferential choice, where probabilities, outcomes, and product attributes are distorted to support an emerging preference (DeKay et al., 2009, 2011; Miller et al., 2013; Polman & Russo, 2012; Russo et al., 1998).

Across these domains, several robust characteristics of information distortion have been identified. First, the magnitude of distortion is systematically related to commitment to the currently leading alternative. Distortion increases as commitment to the emerging preference strengthens, often operationalized as the subjective likelihood that the leading option will ultimately be chosen once all information has been evaluated (Chaxel et al., 2013; DeKay et al., 2009, 2011; Polman & Russo, 2012).

Second, information distortion is typically not accessible to introspection. Self-reports of biased evaluation are consistently uncorrelated with the observed magnitude of distortion, indicating that individuals are largely unaware of the extent to which their evaluations are biased (Russo et al., 2000, 2006; Russo & Yong, 2011). This lack of introspective access suggests that distortion operates largely outside conscious awareness and may therefore persist even when individuals believe themselves to be evaluating information objectively.

Importantly, information distortion is not limited to qualitative or subjectively interpretable information. A series of studies by DeKay and colleagues shows that decision makers distort even precise numerical values—such as probabilities and monetary outcomes—in the direction of the currently leading alternative (DeKay et al., 2009, 2011, 2012; Miller et al., 2013). The magnitude of distortion for numerical information does not differ from that observed for verbal equivalents (Russo & Yong, 2011). Moreover, distortion mediates the effect of an initial preference on the final choice, underscoring its functional role in shaping decision outcomes rather than merely reflecting a by-product of preference formation (DeKay et al., 2011, 2012; Miller et al., 2013).

Research over the past two decades has further documented the robustness and boundary conditions of information distortion. The effect is attenuated when all relevant information is presented simultaneously rather than sequentially (Carlson et al., 2006), when attribute values are known in advance (Carlson & Pearo, 2004), and when group members disagree about which option is currently leading (Boyle et al., 2012).

A central mechanism proposed to underlie information distortion is the goal of cognitive consistency. As tentative preferences emerge, decision makers appear motivated to interpret subsequently encountered information in ways that maintain consistency with their emerging preference. Experimental work shows that activating the goal of cognitive consistency increases the magnitude of distortion and that individual differences in consistency motivation predict the extent of distortion (Russo et al., 2008). This link between emerging preferences, consistency goals, and biased evaluation provides a natural connection to coherence-based models of decision making, which conceptualize preference formation as a dynamic process in which mutually consistent beliefs gradually stabilize over time.

Information distortion shows that systematic bias can arise dynamically during preference formation, even in the absence of prior beliefs or commitments. However, not all biases in information integration emerge from such developing preferences. Many biases instead reflect the influence of established beliefs, attitudes, or desired conclusions on how information is evaluated. These belief-driven biases are commonly subsumed under the label of motivated reasoning and are reviewed in the following section

2.1.2 Motivated Reasoning

Motivated reasoning refers to a class of reasoning biases in which the evaluation and integration of information are shaped by underlying goals or motivations. A central distinction in this literature concerns the nature of these goals. Specifically, reasoning may be guided by the goal of reaching an accurate conclusion or by the goal of supporting a preferred one (Kunda, 1990), building on earlier work by Kruglanski and colleagues (Kruglanski, 1980; Kruglanski & Ajzen, 1983; Kruglanski & Klar, 1987) and related approaches emphasizing the interplay of cognitive and motivational processes in human reasoning (see also Pyszczynski & Greenberg, 1987).

Both types of goals influence reasoning by shaping which beliefs, strategies, and inferential rules are applied to a given problem. However, they differ fundamentally in their consequences for information integration. Accuracy goals tend to promote more careful and normatively appropriate reasoning strategies, which can attenuate some biases, particularly those associated with superficial or hasty processing (Freund et al., 1985; Kruglanski & Freund, 1983; Tetlock & Kim, 1987). At the same time, incentives or instructions to be accurate often fail to eliminate bias altogether, indicating that accuracy motivation does not guarantee unbiased information integration (Fischhoff, 1977; Lord et al., 1984; Tversky & Kahneman, 1973).

In contrast, directional goals bias reasoning toward conclusions that support desired outcomes. They do so by selectively guiding memory search, belief construction, and the interpretation of evidence in ways that favor preferred conclusions (Kunda, 1990; see also Greenwald, 1980). Importantly, such biases are rarely experienced as intentional. Rather, individuals maintain an illusion of objectivity by constructing justifications for desired conclusions that appear reasonable to an impartial observer (Kruglanski, 1980; Pyszczynski & Greenberg, 1987). As Kunda (1990) notes, people are not free to conclude whatever they wish; instead, desired conclusions are adopted only when supporting evidence can be recruited, often without awareness that this recruitment process itself is biased (see also Darley & Gross, 1983).

Motivated reasoning has been documented across many domains of judgment and decision making. The present dissertation focuses in particular on directional goals in political contexts, where individuals often hold strong prior beliefs tied to partisan or ideological identities. Such beliefs provide a particularly informative case for studying belief-driven biases because they can function as powerful prior commitments that shape how new information is interpreted. Empirical research shows that citizens evaluate political information more favorably when it aligns with their prior commitments and more critically when it originates from opposing groups (Lodge & Taber, 2013; Taber & Lodge, 2006). Classic evidence further demonstrates that partisan source cues can dominate evaluations even when policy content is held constant (Cohen, 2003). These patterns are commonly interpreted as manifestations of motivated reasoning, in which new information is assimilated in ways that support preexisting beliefs (Baron, 2008; Munro & Ditto, 1997). A well-known illustration is the belief polarization effect reported by Lord et al. (1979), where participants holding opposing attitudes toward capital punishment became more confident in their initial positions after reviewing the same mixed evidence. Rather than converging toward a shared conclusion, individuals selectively interpreted the evidence in ways that reinforced their prior beliefs.

In the present dissertation, such belief-driven biases are considered alongside information distortion during preference formation. Understanding whether biases in information evaluation originate from prior beliefs or instead emerge dynamically during preference construction—and how these two sources of bias interact—requires methods that trace how preferences and evaluations evolve over time. The stepwise evolution-of-preference (SEP) paradigm provides a framework for studying these dynamics and is introduced in the following section.

2.2 The Stepwise Evolution of Preference Paradigm

One experimental framework that embodies this process perspective is the SEP paradigm, which has become the standard methodological approach for studying information distortion. The paradigm allows researchers to observe how evaluations, preferences, and confidence evolve as information is encountered sequentially during a decision.

The SEP paradigm was originally introduced by Russo and colleagues (1996, 1998) to study how preferences evolve during multi-attribute choice. In the original implementation, information about competing options was presented attribute by attribute, such that each step of the decision process revealed one attribute describing both alternatives simultaneously (e.g., the economic efficiency of Reform A and Reform B). Participants then evaluated the attribute on a single bipolar scale, indicating the extent to which the information favored one option over the other. After each attribute, participants reported which option they currently preferred, allowing researchers to identify the currently leading and trailing alternative as the decision unfolded.

Subsequent methodological refinements sought to disentangle different sources of distortion that cannot be separated under this bipolar evaluation procedure. Because the original scale captures only how strongly an attribute is perceived to favor one option over the other, it remains unclear whether a shift in evaluation reflects increased support for the leading alternative, decreased support for the trailing alternative, or a combination of both. To address this limitation, Blanchard et al. (2014) introduced a dual-rating approach in which participants evaluate the implications of each attribute separately for each option. Instead of indicating which option an attribute favors on a single scale, participants provide independent evaluations for each alternative based on the same attribute information. This approach allows distortion to be decomposed into pro-leader distortion (more favorable evaluations of information about the leading option) and anti-trailer distortion (less favorable evaluations of information about the trailing option).

A further refinement concerns the presentation structure of information items. Whereas early SEP studies presented attributes describing both alternatives simultaneously, later work (e.g., DeKay et al., 2014) presented information item-wise, such that each step reveals a single piece of information about one option. This item-wise presentation more closely approximates sequential information environments and allows distortion to be examined at the level of individual information items.

The empirical studies in the present dissertation adopt this refined measurement approach. Following the procedure introduced by DeKay et al. (2014), information is presented item-wise and evaluated with respect to the option to which it belongs enabling the separation of pro-leader distortion and anti-trailer distortion. After each evaluation, participants indicate their current preference on a continuous single-scale measure, reflecting the extent to which one option currently leads over the other. This continuous measure replaces the dichotomous preference reports used in earlier SEP studies (e.g., Blanchard et al., 2014; Miller et al., 2013) and provides a more sensitive indicator of the evolving preference state during the decision process.

Information distortion is assessed by comparing evaluations provided during the decision process with baseline evaluations obtained in an independent control condition. In the control condition, participants evaluate the same information items outside of a decision context and without forming a preference between options. Because the information items appear to refer to unrelated options, participants cannot form a coherent preference that might bias their evaluations. The resulting ratings therefore provide an estimate of the baseline attractiveness of each information item.

Distortion is defined as the deviation between a participant's evaluation of an information item in the choice condition and the average evaluation of the same item in the control condition. Positive deviations indicate that an item is evaluated more favorably than would be expected based on its baseline attractiveness, whereas negative deviations indicate less favorable evaluations.

Empirical analyses typically examine whether these deviations systematically depend on the participant's tentative preference indicated in the previous round of the decision process. A common modeling approach predicts distortion as a function of whether the current information item refers to the option that was leading or trailing in the preceding round. If participants tend to assign more favorable evaluations for items associated with the option that was previously leading, this pattern is interpreted as evidence of information distortion, reflecting the influence of emerging preferences on the evaluation of newly encountered information.

2.3 Process-Based Accounts of Sequential Information Integration

The SEP paradigm has provided robust and replicable evidence for information distortion and has clarified important boundary conditions under which such distortions arise. Early work demonstrated that individuals systematically evaluate incoming information in a way that favors the currently leading alternative (Russo et al., 1998), while later reviews have summarized the robustness of this effect across tasks, domains, and variations of the paradigm (Russo, 2015). At the same time, the paradigm is primarily descriptive in nature. While it allows researchers to trace how evaluations and preferences evolve during sequential decision making, it does not by itself specify the cognitive mechanisms that generate these dynamics.

As discussed in Chapter 1, many influential models of judgment and decision making assume unidirectional influence, in which evaluations of information determine preferences, but not vice versa. In light of the findings from SEP research, this assumption becomes a critical limitation, as it precludes feedback from emerging preferences to the evaluation of new information and therefore cannot account for information distortion (DeKay, 2015). This limitation characterizes a wide range of formal frameworks, including prospect theory (Tversky & Kahneman, 1992), decision field theory (Roe et al., 2001), evidence accumulation and sequential sampling models (Ratcliff & Smith, 2004), and heuristic-based approaches such as fast-and-frugal heuristics (Gigerenzer et al., 1999). All of these frameworks assume unidirectional influence from information to preferences and therefore do not allow emerging preferences to feed back into the evaluation of information.

Similar constraints apply to many formal accounts of motivated reasoning. Although recent Bayesian and economic models have offered valuable formalizations of belief updating under motivational influences—often by contrasting observed belief updating with a normative Bayesian benchmark—these approaches typically conceptualize influence as flowing from prior beliefs to the evaluation of new evidence (Drobner & Goerg, 2024; Rigoli, 2021). As a result, they capture top-down belief effects but do not model the reciprocal dynamics between beliefs, preferences, and information evaluation that characterize sequential decision making.

By contrast, coherence-based models constitute a notable exception. By allowing bidirectional links between information, beliefs, and preferences, these models provide a principled framework for explaining how evaluative coherence can give rise to systematic distortions during sequential information integration. The following section therefore reviews coherence-based accounts of biased information integration and highlights their capacity to generate testable predictions within the SEP paradigm.

2.3.1 A Coherence-Based Model of Predecisional Information Distortion

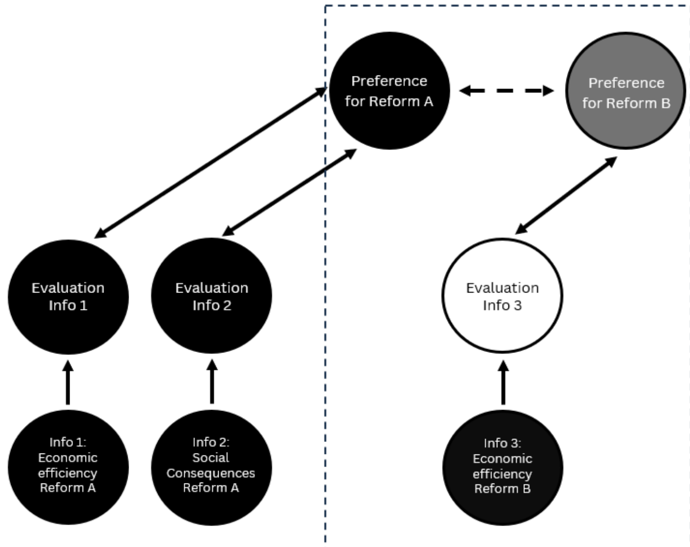
A prominent coherence-based account of information distortion in sequential decision making has been proposed by DeKay (2015). Building on earlier constraint-satisfaction approaches to judgment and decision making (e.g., Holyoak & Simon, 1999), DeKay introduced a simplified connectionist network model that formalizes the dynamic interplay between emerging preferences and the evaluation of sequentially encountered information.

Figure 1 illustrates the core architecture of this conceptual process model. At the lowest level, the information strength of each piece of information—typically measured independently in a control condition—forms the input layer of the network. These inputs feed into a layer of subjective information evaluations. The evaluation nodes are in turn connected to option nodes representing the competing alternatives through bidirectional connections. The option nodes themselves are mutually inhibitory, reflecting the mutual exclusivity of the competing alternatives.

Within this architecture, information evaluation and preference updating unfold in parallel through bidirectional activation dynamics. As each new piece of information is encountered, activation spreads between evaluation nodes and option nodes. Consequently, the evaluation of an information item depends jointly on its objective information strength and on the current preference state. At the same time, the evaluation contributes to updating the relative activation of the competing option nodes.

Figure 1.

Simplified Connectionist Model of Predecisional Information Distortion.



Note. Adapted from DeKay (2015, Fig. 1, p. 406). In the illustrated example, a person has received two positive pieces of information about Reform A in the previous two rounds and has formed an initial relative preference for Reform A over Reform B (black vs. gray option nodes). In the third round of information presentation (dashed box), the person encounters a new piece of information favoring Reform B. The conceptual model predicts that the evaluation of this newly encountered information (white node) depends jointly on its information strength (input layer) and on the already established preference for Reform A over Reform B (option nodes), while simultaneously influencing the relative preference for the competing reforms through the bidirectional coupling between subjective information evaluations and option nodes. Solid arrows represent excitatory associations between information evaluations and option preferences, whereas the dashed connection between the option nodes indicates mutual inhibitory influence between the competing alternatives.

The dynamic implications of this architecture are illustrated in Figure 1. In the example, the decision maker has already encountered two pieces of information favoring Reform A, which has led to a tentative preference for this option (dark node). When a subsequent piece of information favoring Reform B is encountered (dashed box), its evaluation is influenced both by its information strength and by the already established preference state. As activation spreads through the network, the information favoring the nonpreferred option tends to be evaluated less positively than warranted by its objective strength, reflecting anti-trailer distortion. Conversely, if newly encountered information favors the currently leading option, the same mechanism predicts that it will be evaluated more positively than warranted by its objective strength, producing pro-leader distortion (Blanchard et al., 2014; DeKay, 2015).

By explicitly implementing reciprocal influence between preferences and information evaluation, DeKay's model overcomes a key limitation of unidirectional models of decision making. Rather than assuming that preferences simply accumulate evidence, the model captures how dynamic preference–evaluation coupling can give rise to systematic distortions during sequential information integration.

2.3.2 Coherence-Based Accounts of Motivated Reasoning and Confirmation Bias

A broader application of coherence-based reasoning to biased information processing has recently been advanced by Simon and Read (2025). Building on earlier coherence-based models of decision making (Read et al., 1997; Simon, Krawczyk, et al., 2004), they propose a general framework in which a wide range of reasoning biases can be understood as emergent outcomes of coherence-seeking constraint-satisfaction processes.

A central target of this approach is confirmation bias, defined as the tendency to evaluate incoming evidence in a manner that supports preexisting beliefs, expectations, or hypotheses

(Nickerson, 1998). From a normative perspective, such backward influence violates the norm of directionality, which stipulates that inferences should proceed from evidence to conclusions rather than the reverse (Arkes, 1991; Klayman, 1995). Nevertheless, a large body of research demonstrates that belief-consistent information is judged as more convincing than belief-inconsistent information (Lord et al., 1979; Mahoney, 1977).

Simon and Read (2025) illustrate the explanatory power of coherence-based reasoning by revisiting the classic belief-polarization study by Lord et al. (1979). In this study, proponents and opponents of the death penalty were exposed to the same mixed evidence, yet both groups emerged with more extreme initial attitudes. Using Thagard's (1989) ECHO model, Simon and Read show that these findings can be reproduced by a constraint-satisfaction network in which attitudes, beliefs, and evaluations of evidence mutually reinforce one another. As activation spreads through the network, belief-consistent studies are perceived as convincing and well conducted, whereas belief-inconsistent studies are discounted—yielding a coherent but normatively problematic belief configuration.

From this perspective, confirmation bias is not an isolated error but a natural outcome of coherence maximization under informational conflict. Evidence that aligns with existing beliefs increases coherence and is readily assimilated, whereas contradictory evidence induces incoherence. Such incoherence can be resolved normatively by revising the belief or non-normatively by reinterpreting the evidence; coherence-based reasoning predicts that the latter route will often prevail (Simon & Read, 2025).

Simon and Read further argue that the same coherence dynamics can account for politically motivated reasoning. In political contexts, ideological commitments and partisan identities can function as highly central nodes within the belief network. Because these nodes are strongly connected to related beliefs, values, and social identities, information that contradicts them generates substantial incoherence and is therefore more likely to be discounted or critically evaluated, whereas belief-consistent information is more readily accepted.

Importantly, Simon and Read (2025) develop this extension primarily at a theoretical level, arguing that a wide range of reasoning biases may reflect different manifestations of the same underlying coherence dynamics. The implications of this perspective for dynamically unfolding information integration, however, remain largely unexplored. The following section therefore considers several open challenges in process-based accounts of sequential information integration.

2.3.3 Open Challenges in Process-Based Accounts of Sequential Information Integration

Despite their considerable explanatory appeal, existing coherence-based accounts of sequential information integration face important challenges in how their predictions have been operationalized and tested empirically, as well as in their extension to more complex, real-world belief contexts. The following discussion focuses selectively on those challenges that bear directly on the open questions addressed in the subsequent empirical chapters.

First, a central challenge concerns the translation of conceptual process assumptions into statistical models in sequential decision paradigms. Recent debates on theory development have underscored the value of formal and computational modeling for strengthening theoretical precision and constraint (Farrell & Lewandowsky, 2010; Marewski & Olsson, 2009; Oberauer & Lewandowsky, 2019) and have proposed structured approaches to theory construction and formalization (Borsboom et al., 2021; van Dongen et al., 2025; van Rooij & Blokpoel, 2020). However, many theories initially take the form of conceptual process models—verbally articulated accounts specifying a phenomenon’s assumed generative and temporal dynamics

(Borsboom et al.,2021). At this intermediate stage, explanatory progress depends on whether these conceptual commitments are systematically carried forward into hypothesis specification and statistical testing.

In the context of sequential information integration, this translation step is not trivial. Conceptual and formal coherence-based models such as DeKay's (2015) explicitly assume continuous, reciprocal updating between preferences and information evaluation. By contrast, prevailing analytic practices in the SEP literature operationalize information distortion primarily as a function of prior preference states (e.g., Blanchard et al., 2014). Statistical predictors therefore reflect a serial dependence on earlier rounds rather than the parallel, bidirectional dynamics implied by the conceptual model. As a result, empirical tests may approximate a fundamentally parallel process using predictors that do not fully instantiate their core assumptions.

This potential mismatch does not imply a lack of theory. Rather, it reflects a breakdown in the alignment between conceptual process models and their statistical operationalization. When statistical analyses follow paradigm-specific conventions more closely than theory-implied dynamics, model-implied effects may remain only partially examined. Identifying this mismatch, deriving model-implied predictors that preserve reciprocal preference–evaluation coupling, and examining their empirical implications constitute a central contribution of the present dissertation and are taken up explicitly in Chapter 4.

Second, although coherence-based frameworks provide a unifying account of biased reasoning, their application to contexts characterized by strong, real-world belief systems has remained primarily theoretical. Simon and Read (2025) argue that a wide range of reasoning biases—including confirmation bias, belief polarization, and politically motivated reasoning—

can be understood as instantiations of the same underlying constraint-satisfaction dynamics. Central to their account is the inherently temporal nature of coherence formation: networks gradually evolve toward attractor states, such that the impact of incoming information depends on its position in the evolving process and on the strength of prior commitments shaping that trajectory. Yet these dynamic implications have not been subjected to systematic empirical tests in sequential process-tracing paradigms. It remains largely unexamined how stable ideological priors interact with emerging preferences over the course of stepwise information integration.

Complementing this perspective, Oeberst et al. (2023; Oeberst & Imhoff, 2025) integrate coherence-based reasoning with research on belief-consistent information processing and explicitly call for future research to examine when coherence shifts are unidirectional versus bidirectional, particularly under conditions of strongly held prior beliefs. They highlight the need to measure belief change and evaluative distortion jointly and to investigate how different conceptualizations of belief strength shape coherence formation. To date, however, empirical work on belief-consistent processing has largely relied on static pre–post designs or global judgment measures, rather than process-tracing paradigms capable of capturing the temporal interaction between ideological priors, emerging preferences, and information evaluation.

Both frameworks converge in suggesting that stable ideological commitments should intrude into and potentially constrain dynamic coherence-based integration processes. Yet a systematic, trial-level examination of how such priors modulate sequential preference updating and information distortion remains largely absent. Chapter 5 of the present dissertation addresses precisely this open question by embedding partisan ideological priors within a sequential coherence model and testing process-level predictions concerning their interaction with evolving preferences during political information evaluation.

Taken jointly, these challenges delineate two conditions under which coherence-based accounts of sequential information integration can advance beyond their current scope: first, that their dynamic process assumptions be translated into statistically diagnostic predictors; and second, that stable belief structures be explicitly incorporated into models of evolving preference–evaluation coupling. The empirical chapters that follow address these conditions by testing model-implied trial-level dynamics and by examining how ideological priors shape coherence formation over the course of sequential information processing.

Chapter 3: A Coherence-based Approach to Biased Information Search in Sequential Decision Making

Whereas the previous chapter examined how coherence-based processes bias the *evaluation* of information once it is encountered, the present chapter turns to how these processes shape *which* information is selected during sequential decision making. Research on information search shows that decision makers do not sample information randomly but exhibit systematic biases that favor information aligned with emerging preferences or hypotheses.

This chapter reviews evidence for biased information search and examines how such biases can be understood from a coherence-based perspective in sequential decision making. It proceeds in three steps. First, it introduces prominent descriptive biases in information search. Second, it reviews existing conceptual accounts of information search and assesses their ability to capture dynamic coherence processes. Third, it discusses key limitations of current accounts of information search and outlines the requirements for extending coherence-based frameworks to account for biased information selection under sequential conditions.

3.1 Descriptive Biases in Information Search

Decision making often requires choosing between multiple alternatives under conditions of uncertainty. When options differ across several attributes and potential consequences, it is typically impossible to consider all relevant information simultaneously. As emphasized by Ratcliff et al. (2016), this sequential nature of decision making can arise because information becomes available only gradually over time or because cognitive processing capacity is limited, such that information must be extracted and evaluated piecemeal.

This structure characterizes many real-world choice contexts, including the policy decision example introduced in the General Introduction. When deciding which of two proposed policy reforms to support, initial information may tentatively favor one option, while additional considerations—such as economic efficiency, social consequences, or political feasibility—are encountered or processed step by step. As a result, decision makers must repeatedly integrate new information and decide which aspects of which option to consider next, often while a provisional preference between the alternatives is already beginning to guide the decision process (Hastie & Dawes, 2001).

Early experimental work demonstrated that variations in information order and presentation format can systematically alter preferences (Bettman, 1979; Hogarth & Einhorn, 1992; Plott & Levi, 1978) and that decisions differ depending on whether information is processed sequentially or simultaneously (Russo, 1977). Beyond these presentation effects, subsequent research has shown that information acquisition itself is systematically biased. Rather than sampling information evenly across alternatives, decision makers tend to guide search in ways that reflect strategy use, hypothesis testing, and emerging preferences.

To organize this heterogeneous literature, the following sections distinguish three classes of descriptive biases in information search. First, heuristic and cue-based search strategies describe how information acquisition can be structured by efficiency-oriented rules that specify search order, stopping, and decision criteria. Second, hypothesis-focused biases emphasize cognitive constraints that lead search to be directed toward testing focal hypotheses. Third, preference-driven and motivationally driven search patterns capture how emerging evaluations and goals bias information acquisition toward particular options.

3.1.1 Heuristic and Cue-based Search Strategies

Building on the multi-strategy perspective outlined in Chapter 1, heuristic approaches explain information search as part of decision strategies that trade off decision accuracy against cognitive effort. Rather than assuming a single general-purpose mechanism, this tradition posits that decision makers can flexibly select among qualitatively distinct strategies as a function of task demands and environmental structure (Gigerenzer et al., 1999; Payne et al., 1993; Svenson, 1979; Thorngate, 1980).

A central contribution of this framework was the systematic decomposition of decision strategies into elementary information processes, which allowed information acquisition to be treated as an integral component of strategy execution and enabled quantitative comparisons of cognitive effort and accuracy across strategies (Johnson & Payne, 1985). As a result, heuristic strategies came to be formalized as explicit process models that specify search order, stopping behavior, and decision rules, rendering them particularly amenable to process-tracing methods (Payne, 1976). These ideas provided the empirical basis for cost–benefit models of strategy selection (Beach & Mitchell, 1978) and were later integrated into a broader cognitive framework linking information search, strategy selection, and environmental structure (Payne et al., 1993).

Within this tradition, information search is commonly assumed to be guided by the expected usefulness of cues. Usefulness is defined with respect to an objective decision criterion and reflects the extent to which a cue is probabilistically related to correct inferences about that criterion under uncertainty (B. R. Newell et al., 2004). Accordingly, heuristic models assume that cues differ in their validity, that is, in how informative they are about which option better satisfies the decision criterion. The Take-the-Best heuristic formalizes this assumption as a non-compensatory, cue-wise search process in which cues are inspected in descending order of validity and information acquisition terminates once a cue discriminates between options (Gigerenzer & Goldstein, 1996). In contrast, compensatory strategies such as the Weighted Additive Rule assume option-wise information acquisition, in which multiple cues are sampled, weighted, and integrated into overall evaluations before alternatives are compared (Payne et al., 1993).

Subsequent research refined these models by demonstrating that cue usefulness cannot be reduced to validity alone. Instead, information search is also shaped by cue discriminability (B. R. Newell et al., 2004), expected gains in decision accuracy (Nelson, 2005; Nelson et al., 2010), and cue–criterion correlations (Rakow et al., 2005). The concept of cue success integrates validity and discriminability and captures the proportion of correct inferences that can be achieved by relying on a single cue (Martignon & Hoffrage, 1999).

Despite their explanatory power, heuristic models face systematic limitations. A robust body of empirical evidence shows that individuals typically acquire substantially more information than predicted by simple heuristics such as Take-the-Best (Bröder, 2003; Hausmann & Läge, 2008; B. R. Newell & Shanks, 2003), and that information search varies flexibly as a function of contextual demands, including task complexity (Böckenholt et al., 1991), information costs (Bröder & Schiffer, 2006), time pressure (Edland, 1994; Edland & Svenson, 1993), stress (Glöckner & Moritz, 2008), and age (Mata et al., 2007, 2011). Moreover, search termination is rarely governed by fixed stopping rules; instead, it depends dynamically on the information acquired during the search process itself (B. R. Newell & Bröder, 2008; Söllner et al., 2014; Söllner & Bröder, 2016).

Crucially for the present purposes, heuristic accounts conceptualize information search primarily as a function of cue properties and task structure. Consequently, they offer limited leverage for explaining systematic search patterns that arise independently of cue validity or discriminability—most notably, biases that depend on emerging preferences, prior beliefs, or the valence of available information. These limitations motivate the consideration of alternative accounts that conceptualize information search as dynamically shaped by cognitive constraints, hypothesis testing, and preference formation, which are addressed in the following sections.

3.1.2 Hypothesis-, Belief-, and Preference-Guided Biases in Information Search

Beyond efficiency-oriented heuristic strategies, information search can be systematically biased because acquisition is guided by focal hypotheses, beliefs, or emerging preferences rather than by cue properties alone. A substantial body of research documents that such biases arise from different sources and operate according to distinct principles, depending on whether search is organized around epistemic hypotheses, motivational commitments, or dynamically evolving preferences. Three prominent forms of biased information acquisition are positive hypothesis testing, confirmatory search (i.e. selective exposure), and attraction search.

A well-established cognitive bias in information search is positive hypothesis testing. Rather than sampling information that would be maximally diagnostic in a normative sense, individuals tend to preferentially examine cases in which a focal hypothesis is expected to hold, even when sampling disconfirming cases would yield greater informational value (Klayman & Ha, 1987, 1989; Navarro & Perfors, 2011). This tendency is closely related to pseudodiagnosticity, where evidence is evaluated primarily with respect to a focal hypothesis while neglecting its implications for alternative hypotheses (Fischhoff & Beyth-Marom, 1983). Positive hypothesis testing is generally interpreted as a cognitively grounded default of human reasoning rather than as a motivational bias. Dual-process accounts suggest that focusing on a single hypothesis at a time reflects a general constraint of human information processing, which can systematically shape both information search and evaluation (Evans et al., 2002; Mynatt et al., 1993). Classic demonstrations of this tendency stem from hypothesis-testing paradigms such as the Wason Selection Task, where participants predominantly select information consistent with the antecedent of a conditional rule instead of information that could falsify it (Johnson-Laird & Wason, 1970; Wason, 1968). From the perspective of Popper's falsificationist philosophy of science, such behavior is normatively suboptimal, as only evidence capable of refuting a hypothesis provides decisive information about its validity (Popper, 1934/2005).

A conceptually distinct form of biased information acquisition is confirmatory search, most prominently studied within the selective exposure paradigm. In the literature, the terms *confirmatory search* and *selective exposure* are both used to refer to the tendency of individuals to preferentially seek information that supports their existing beliefs, attitudes, or decisions while avoiding information that could challenge them (Fischer & Greitemeyer, 2010; Frey, 1986; Hart et al., 2009; Nickerson, 1998). In contrast to positive hypothesis testing, confirmatory search relies on valence-related expectations about the implications of available information for a focal belief or decision. This form of biased search can therefore only occur when individuals are able to anticipate whether information is likely to be supportive or threatening, rendering this form of biased search explicitly dependent on prior beliefs and their evaluative implications.

Theoretically, confirmatory search has been explained both by motivational and by cognitive–informational mechanisms. From a motivational perspective, confirmatory search serves to reduce the aversive state that can arise after committing to a decision by favoring belief-consistent over belief-inconsistent information, as postulated by cognitive dissonance theory (Chaiken, 1980; Chen et al., 1999; Festinger, 1957b). Meta-analytic evidence indicates that such defense motivation is a particularly strong driver of confirmatory search (Hart et al., 2009). From a cognitive–informational perspective, confirmatory search can arise because decision makers evaluate information asymmetrically, scrutinizing belief-inconsistent evidence more critically than belief-consistent evidence. As a consequence, preference-consistent information is often perceived as higher in quality and diagnosticity (Ditto & Lopez, 1992; Lord et al., 1979). This evaluative asymmetry can bias subsequent information acquisition, even when individuals are motivated to reach accurate decisions (Fischer et al., 2005; Fischer, Greitemeyer, et al., 2008), consistent with broader accounts of motivated reasoning (Kunda, 1990).

A third form of biased information search, central to the present dissertation, is attraction search (also termed *leader-focused search*; Carlson & Guha, 2011). In contrast to both positive hypothesis testing and confirmatory search, attraction search refers to the tendency to preferentially sample information about the option that currently appears more attractive, even when the evaluative valence of that information is unknown at the time of selection (Jekel et al., 2018; Scharf et al., 2019).

Evidence for attraction search has been reported across different decision domains, including probabilistic inference and preferential choice tasks (Jekel et al., 2018; Scharf et al., 2019). Across these settings, individuals reliably focus subsequent information acquisition on the currently leading option, irrespective of whether the additional information ultimately supports or undermines that option. Importantly, the effect is systematically stronger under conditions of restricted or costly information search, suggesting that option-level evaluations increasingly guide acquisition when cognitive or situational resources are constrained (Scharf et al., 2019).

Building on these findings, several factors have been shown to modulate the strength of attraction search. Directing attention to relative option attractiveness prior to search amplifies attraction-based information acquisition (Scharf et al., 2022). Similarly, manipulating the visual salience of information can bias search toward the currently more attractive option, indicating that attentional factors interact with option-level evaluations during information acquisition (Bröder et al., 2021). Attraction-based biases also extend to visual information processing: Eye-tracking studies show that information associated with the more attractive option is preferentially fixated, even in relatively unconstrained visual search environments (Bröder & Scharf, 2025). This pattern is consistent with the broader preferential viewing literature, which shows that visual attention is systematically biased toward subjectively preferred options and attributes during decision making (Orquin et al., 2021).

Conceptually, attraction search bears a superficial resemblance to pseudodiagnostic search, a phenomenon discussed in the literature on hypothesis testing that describes a selective focus on information relevant to a focal hypothesis (Doherty et al., 1979; Fischhoff & Beyth-Marom, 1983). However, rather than reflecting fundamentally different underlying processes, the two phenomena differ primarily in how this selectivity is embedded within the decision process. In pseudodiagnostic search, selectivity is typically framed in terms of testing a focal hypothesis, whereas in attraction search it arises from emerging preferences between options during information acquisition.

Taken together, the literature reviewed in this chapter demonstrates that information search in decision making is systematically biased. Across different lines of research, regularities have been identified showing that information acquisition is shaped by focal hypotheses, motivational considerations, and the relative attractiveness of decision alternatives, rather than being random or solely driven by normative informativeness.

While these descriptive accounts provide important insights into *what* kinds of biases occur in information search, they remain limited in specifying *how* such biases arise during the unfolding decision process. Addressing this question requires theoretical approaches that move beyond verbal characterizations of search tendencies and instead specify the processes underlying information acquisition more explicitly. The following section therefore turns to process-level accounts of information search, which aim to articulate the mechanisms by which information selection unfolds during decision making.

3.2 Process-Based Accounts of Information Search

Process-oriented accounts of information search seek to specify the cognitive mechanisms by which information acquisition unfolds over time and becomes systematically biased during decision making. Unlike descriptive approaches, which characterize regularities in search behavior, process accounts articulate explicit assumptions about the mechanisms that generate biased information selection. This process-level perspective is essential for adjudicating whether different forms of biased search reflect distinct mechanisms or emerge from shared principles.

Within the broader judgment and decision-making literature, relatively few frameworks have treated information search as an endogenous component of preference formation rather than as a preparatory stage or the outcome of strategy selection (Gigerenzer & Goldstein, 1996; Payne et al., 1993). This omission is notable given extensive evidence that emerging evaluations systematically guide what information decision makers seek next in sequential choice contexts. The present section therefore focuses on process accounts that explicitly link information acquisition to the evolving internal state of the decision maker.

Section 3.2.1 examines a coherence-based account of information search derived from the integrated coherence-based decision and search framework (iCodes; Jekel et al., 2018). Building on the architectural principles introduced in Chapter 1, this account specifies how preference-dependent search patterns can arise endogenously from coherence-maximizing dynamics. Section 3.2.2 then turns to the Cognitive Economy Model of selective exposure, which offers a complementary but theoretically distinct explanation by conceptualizing biased information acquisition as a response to cognitive constraints.

3.2.1 A Coherence-Based Model of Information Search

As outlined in Chapter 1, coherence-based models conceptualize decision making as the gradual stabilization of a maximally coherent representation through recurrent interactions between options and information (Glöckner & Betsch, 2008a; Holyoak & Simon, 1999; Thagard, 1989). The iCodes framework extends this perspective by treating information acquisition as part of the same coherence-seeking process that governs preference formation (Jekel et al., 2018). Rather than assuming that search is governed by fixed cue priorities or independent strategies, iCodes predicts that search decisions emerge from the current activation state of the coherence network.

As information is sequentially processed, even small initial differences in option attractiveness can become amplified through recurrent activation dynamics. Once one option gains an activation advantage, this advantage spreads through the network and influences which information is selected next. Information associated with the currently more attractive option receives stronger activation and therefore becomes more likely to be acquired.

Figure 2 illustrates this mechanism using the policy reform example introduced earlier. At the depicted point in the decision process, the decision maker has already encountered two pieces of information: economic efficiency strongly favors Reform A (++), whereas administrative complexity strongly disfavors Reform B (--). As a result, the preference node for Reform A currently exhibits higher activation than the preference node for Reform B.

Two additional pieces of information remain hidden: the administrative complexity of Reform A and the economic efficiency of Reform B. Importantly, these cues differ in validity, reflected by the thickness of the connections from the source node. Economic efficiency is the more valid cue and therefore receives stronger bottom-up activation than administrative complexity.

From the perspective of cue-based heuristics, this structure would predict that the next information search step is guided by cue properties such as validity or discriminability, favoring the inspection of the most informative cue—here, economic efficiency for Reform B. In the iCodes framework, however, search is not determined by cue validity alone. Instead, cue activation reflects the joint influence of bottom-up cue validity and top-down activation from the currently preferred option.

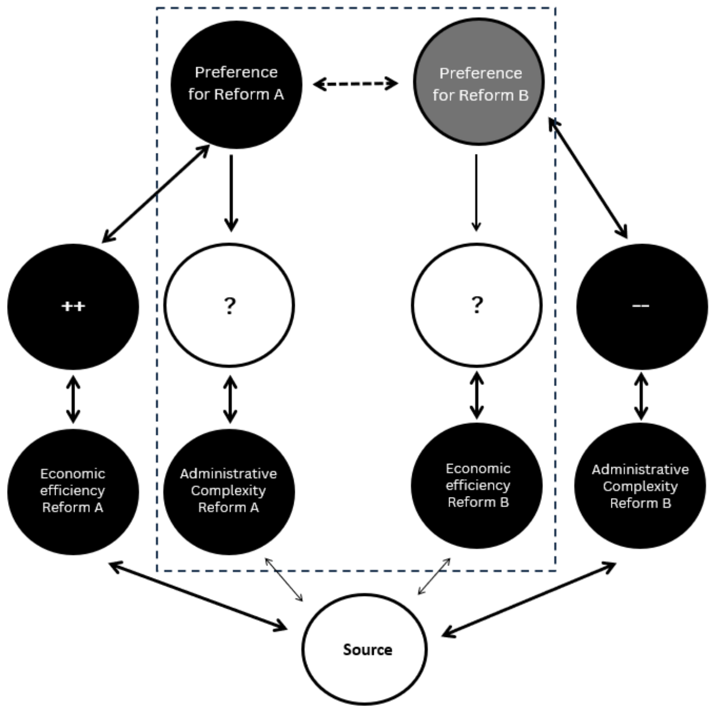
Because the preference for Reform A currently has higher activation within the network, activation spreads from the preference node for Reform A to cues associated with that option.

This top-down activation can counteract the bottom-up advantage of the more valid cue associated with Reform B. Consequently, the model predicts that the decision maker will reveal the administrative complexity of Reform A rather than the more valid economic efficiency cue for Reform B.

This dynamic dependency of information acquisition on the current relative attractiveness of options gives rise to the Attraction Search Effect (Jekel et al., 2018), introduced descriptively in Section 3.1. Crucially, attraction search does not rely on motivational goals such as confirmation or defense. Instead, it arises as a structural consequence of coherence maximization: once an option gains an activation advantage, cues associated with that option become more strongly activated within the network and therefore more likely to be selected.

Figure 2.

Simplified Illustration of Information Search Dynamics in the iCodes framework.



Note. Adapted from Bröder and Scharf (2025). The figure illustrates the coherence-based network architecture underlying the iCodes model using a hypothetical policy reform decision. Nodes are organized hierarchically from bottom to top. The bottom node represents the source, which provides bottom-up activation to cue nodes. The second row from the bottom represents cue nodes (e.g., economic efficiency, administrative complexity), with connection strength from the source reflecting cue validity (thicker arrows indicate higher validity). The third row from the bottom represents cue value nodes, which encode the evaluative implications of each cue for the respective option. The top row represents the competing options (Reform A and Reform B), connected by a bidirectional inhibitory link (dashed arrow). In the depicted state of the decision process, previously acquired information strongly favors Reform A (++), while information about Reform B is negative (--), resulting in higher activation of the preference node for Reform

A. Hidden cues are not yet integrated into the evaluative network; once revealed, they become activated and enter the bidirectional constraint-satisfaction process, allowing top-down and bottom-up influences to interact. Shading indicates relative activation levels of nodes.

Existing empirical tests of coherence-based search predictions have primarily been conducted in environments in which the evaluative valence of information is unknown at the time of selection. Under these conditions, decision makers can anticipate whether information pertains to the leading or trailing option and how diagnostic it may be, but not whether it will ultimately support or undermine that option. As a result, existing evidence primarily speaks to option-focused information acquisition under conditions of valence uncertainty.

By contrast, other accounts of biased information search have focused on situations in which the evaluative implications of information are known in advance. These approaches conceptualize biased information acquisition not as an emergent property of coherence dynamics but as a function of decision certainty and cognitive economy. The following section outlines this alternative perspective.

3.2.2 The Cognitive Economy Model of Selective Exposure

In contrast to coherence-based accounts, which derive biased information search from preference amplification under conditions of valence uncertainty, the Cognitive Economy Model of selective exposure addresses decision environments in which the evaluative implications of information are known in advance (Fischer, 2011). The model is not intended as a formal process model in the sense of network-based coherence frameworks. Rather, it provides a parsimonious conceptual account of confirmatory information search by integrating motivational and cognitive–informational considerations under a single explanatory principle.

The core assumption of the model is that confirmatory search is governed by the interaction between experienced decision certainty and considerations of cognitive economy. Decision certainty is conceptualized as confidence in the validity of one's decision, belief, or attitude (Tormala & Petty, 2002). According to the model, higher perceived decision certainty increases confirmatory information search, whereas decision uncertainty attenuates or reverses confirmatory tendencies (Fischer, 2011).

This relationship is explained by an economy-based logic of cognitive effort allocation. When individuals are confident in their decision, it is efficient to minimize cognitive effort by avoiding difficult-to-process decision-inconsistent information and relying on readily assimilable, decision-consistent material. In contrast, when decision certainty is low, avoiding inconsistent information becomes maladaptive. Under uncertainty, allocating additional cognitive resources to the processing of decision-inconsistent information serves to reduce the risk of erroneous decisions. Importantly, this account does not require a distinction between motivational and cognitive processes at the explanatory level; confirmatory search is treated as an adaptive response to subjective certainty and processing costs rather than as the outcome of defense motivation per se (Fischer, 2011).

Empirical support for this framework comes from studies showing that manipulations affecting decision certainty systematically modulate confirmatory information search. For example, gain–loss framing influences confirmatory search via its impact on perceived certainty, with gain frames increasing confirmatory search relative to loss frames (Fischer, Jonas, et al., 2008; Kahneman & Tversky, 1979). Similarly, threat manipulations that reduce decision certainty attenuate confirmatory search, with decision certainty mediating this effect (Fischer, Kastenmüller, et al., 2011). Complementary evidence indicates that constraints on cognitive resources increase confirmatory tendencies, consistent with the assumption that processing decision-inconsistent information is cognitively demanding (Fischer, Greitemeyer, et al., 2008; Muraven & Baumeister, 2000).

Taken together, the Cognitive Economy Model provides a principled account of confirmatory search in contexts where it is known in advance whether information will favor or disfavor an option. As such, it captures key regularities observed in selective exposure paradigms, while remaining conceptually distinct from coherence-based models, which have primarily been applied to situations in which it is not known in advance whether information will favor or disfavor an option.

3.2.3 Open Challenges in Process-Based Accounts of Information Search

Despite substantial progress in modeling biased information search, existing approaches remain limited in theoretically consequential ways. The following discussion focuses on those constraints that are directly relevant to the contributions of the present dissertation.

First, a limitation concerns the paradigms used to study information search in relation to evolving preferences. Although real-world decision making typically involves sequential evidence accumulation and dynamically emerging preferences, information search has rarely been examined within fully sequential paradigms that allow preference formation and information acquisition to unfold jointly. The SEP paradigm offers precisely this structure, yet applications of SEP-type designs to information search remain scarce. Existing paradigms have either examined leader-focused acquisition following an initial evaluation phase (e.g., Carlson & Guha, 2011) or studied confirmatory search in single-target settings in which all information pertains to one focal object or hypothesis (Chaxel et al., 2013; Mischkowski et al., 2021). While these designs permit clean tests of specific search biases, they do not allow attraction-based and confirmatory search tendencies to be examined simultaneously within a dynamically evolving preference process in which search both shapes and is shaped by emerging evaluations.

Second, although coherence-based models have been extended to information search—most prominently within iCodes (Jekel et al., 2018)—their integrative implications for distinct forms of biased search have not been systematically examined within a unified sequential paradigm. The iCodes framework formalizes attraction search as a structural consequence of coherence maximization: once an option gains an activation advantage, additional information about that option becomes more likely to be selected, regardless of whether it will ultimately confirm or disconfirm the current preference. Empirical tests of this prediction have largely focused on environments in which the evaluative valence of information is concealed at the time of choice.

By contrast, research on confirmatory search has predominantly examined contexts in which information valence is transparent or can be anticipated in advance, allowing decision makers to preferentially select preference-confirming and avoid preference-disconfirming information (Fischer, 2011). These paradigms isolate valence-based selection tendencies but are structurally distinct from those used to test attraction search. As a consequence, attraction search and confirmatory search have typically been investigated in separate experimental traditions, under different informational constraints, and without allowing their joint or competing contributions to be examined within the same evolving decision process.

This separation is theoretically consequential considering broader integrative proposals. Simon and Read (2025) advance coherence-based reasoning as a general process architecture capable of explaining diverse manifestations of biased reasoning, including biased information acquisition. Importantly, they suggest that coherence processes may operate not only during the integration of available information but also in shaping which attributes are selected or retrieved in the first place. If this integrative claim holds, attraction search and confirmatory search should be interpretable within a common coherence framework. Yet systematic empirical tests examining their distinct, additive, or potentially interacting contributions within a single sequential paradigm remain scarce.

Addressing this gap requires paradigms in which information integration and information search unfold jointly as preferences evolve over time, and in which the informational structure allows attraction-based and confirmatory-based tendencies to be disentangled. Chapter 6 responds directly to this challenge by building on and extending the iCodes framework to develop a sequential coherence model that jointly captures attraction search and confirmatory search within evolving decisions, enabling tests of their independent and additive contributions under systematically varied informational conditions.

Overview of the Dissertation Projects

The preceding chapters have outlined two interrelated challenges in current coherence-based approaches to judgment and decision making. Rather than reiterating these arguments, the present section consolidates them and situates the empirical chapters within this broader theoretical agenda.

First, as discussed in Section 2.3.3, coherence-based models articulate conceptual process assumptions about reciprocal updating between preference formation and information evaluation. Yet important aspects of these dynamics have not always been translated into model-implied statistical predictors within sequential paradigms. This gap does not reflect a lack of theory, but a misalignment between conceptual process commitments and empirical operationalization. Addressing this translation problem constitutes the first organizing principle of the dissertation.

Second, as outlined in Section 2.3.3 and 3.2.3, coherence-based reasoning has been proposed as a parsimonious and unifying framework for biased reasoning across domains (Simon & Read, 2025, see also Oeberst et al., 2025). As elaborated in the open challenges sections, this integrative claim raises two further questions: (a) how stable ideological belief systems intrude into dynamic coherence-based updating processes, and (b) whether diverse manifestations of biased information acquisition—such as attraction search and confirmatory search—can be understood as expressions of a shared coherence mechanism operating under different informational structures. Empirical tests directly targeting these integrative implications remain limited.

The present dissertation addresses these complementary challenges across three empirical projects. Chapter 4 advances theory construction by deriving and testing previously overlooked model-implied predictions of coherence-based accounts in sequential information integration, thereby aligning conceptual process models with statistical tests. Chapter 5 embeds stable ideological belief structures within a sequential coherence framework to examine how top-down priors interact with emerging preferences during political information evaluation. Chapter 6 extends sequential preference paradigms to information search by developing and testing a unified sequential coherence model that jointly captures attraction search and confirmatory search, allowing their independent and additive contributions to be examined within evolving decisions. Together, these projects progressively extend coherence-based models by (1) aligning their process assumptions with statistical operationalization, (2) incorporating stable belief structures as top-down constraints, and (3) extending the framework from information evaluation to information acquisition.

Chapter 4: Testing Overlooked Predictions of Coherence-Based Models

4.1 Chapter Introduction

Chapter 4 takes up the first challenge identified in the preceding sections: the alignment of conceptual coherence-based process assumptions with their statistical operationalization in sequential decision paradigms. Coherence-based models provide a detailed account of how preferences and evaluations co-evolve through bidirectional constraint satisfaction. Yet empirical tests within the SEP paradigm have typically operationalized information distortion as a function of prior preference states.

Connectionist models of predecisional information distortion explicitly assume continuous and reciprocal updating between preferences and information evaluation (DeKay, 2015). However, standard analytic practices approximate this parallel process using serial predictors that condition distortion on earlier preference levels. As a result, core model-implied dynamics—such as within-round amplification, attenuation, and reversals of distortion—may remain partially obscured.

This chapter addresses this translation problem directly by deriving statistical predictors that follow from the model's core assumption of parallel preference–evaluation updating. Rather than relying on paradigm-specific conventions, the analyses instantiate the model's process commitments at the trial level, allowing moment-to-moment changes in preference strength to predict corresponding changes in evaluative bias.

To address this issue, the chapter tests these model-derived predictors across multiple preregistered experiments and reanalysis of published datasets. By aligning statistical specification with process-level commitments, the analyses examine whether dynamic coherence effects become visible once preference change is modeled explicitly. Chapter 4.2 is based on the following article:

Häffner, C., Jekel, M., & Lisovoj, D. (2026). *From conceptual model to statistical test: Paradigm blindness in research on predecisional information distortion.* Manuscript under review.

Minor modifications were made to headings and formatting to align with the dissertation style. The content of the manuscript remains unchanged.

4.2. Manuscript I: From Conceptual to Statistical Model: Paradigm Blindness in Research on Information Distortion

Abstract

Psychological science aims to move beyond the accumulation of empirical regularities toward explanatory theories that specify the processes generating behavior. While recent theory-construction frameworks emphasize formal modeling, much research operates at an intermediate stage in which theories take the form of conceptual models without full formalization. At this stage, when theoretical assumptions are not translated into statistical predictors in a manner that preserves model-implied effects, the outcome of the statistical test ceases to be diagnostic of the conceptual model. In cases where paradigm-specific analytic conventions guide the analysis in ways that conflict with the conceptual model, model-implied effects remain untested—a mismatch we term *paradigm blindness*. We illustrate this issue in research on predecisional information distortion. Although coherence-based models posit bidirectional preference-evaluation coupling, standard analyses following the stepwise evolution-of-preference paradigm reduce distortion to a function of prior preference. Across three preregistered experiments (17,908 observations), we identify three previously untested model-implied effects—reversal, push, and pull—that follow from the parallel-processing assumption and when unmodeled, systematically bias estimates of information distortion, with additional converging evidence for the reversal effect in five reanalyzed published datasets (11,228 observations). Together, these findings show how deriving model-implied predictors from conceptual process models restores the diagnostic value of empirical tests for theory development.

Keywords: theory development, paradigm blindness, predecisional information distortion, coherence-based models, sequential decision making

From Conceptual Model to Statistical Test:
Paradigm Blindness in Research on Information Distortion

Concerns about the status of psychological theory have been a recurring theme in methodological and theoretical debates within psychology. These concerns have taken different forms, ranging from early discussions of weak integration to recent diagnoses of cumulative stagnation—the accumulation of many isolated findings without corresponding gains in theoretical coherence, constraint, or explanatory power (Eronen & Bringmann, 2021; Gigerenzer, 2010; Glöckner & Fiedler, 2024; Meehl, 1978). Several authors have also emphasized that even when theories are used to motivate empirical work, the link between theory and hypotheses is often weak because the theories themselves are insufficiently specified to strongly imply particular hypotheses (Fiedler & Trafimow, 2024; Oberauer & Lewandowsky, 2019). Some authors further argue that this lack of theoretical integration has also contributed to broader methodological problems, including poor replicability of findings (Klein, 2014; Muthukrishna & Henrich, 2019; Oberauer & Lewandowsky, 2019).

In response, recent debates on theory development have emphasized computational and other forms of formal modeling as a route to stronger theorizing (Farrell & Lewandowsky, 2010; Marewski & Olsson, 2009; Oberauer & Lewandowsky, 2019), and have given rise to frameworks and methodological tools aimed at the formalization of theory (e.g., Borsboom et al., 2021; Gigerenzer, 2017; Van Dongen et al., 2024; van Rooij & Blokpoel, 2020). As we argue here, however, an important bottleneck arises at an earlier stage of theory development—at the level of conceptual or prototheoretical models, understood as a verbally articulated explanatory model that specify a theory’s core assumptions and general principles prior to formalization, as suggested in *theory construction methodology* (Borsboom et al., 2021). In what follows, we use the term conceptual model to refer to such prototheoretical accounts, and conceptual process model when these accounts additionally specify the generative and temporal structure of the underlying phenomenon.

While prior work has stressed the importance of articulating prototheories and formalization, we focus on a distinct gap: even when a substantive conceptual process model is available, its implications may not consistently carry forward into hypothesis specification, statistical modeling, and testing. At this stage, moving to formalization can be premature if the implications of the conceptual model have not first been systematically articulated and tested. Theory development should therefore begin by using conceptual process models to constrain hypotheses and statistical predictors. When empirical tests are not derived in this way, a genuine mismatch between conceptual and statistical models can arise, such that the results of these analyses provide little leverage for developing, refining, or adjudicating the conceptual model. By contrast, a conceptual model whose implications have been tested provides a principled basis for subsequent formalization.

We argue that one important source of this mismatch is a complementary failure mode—*paradigm blindness*—whereby paradigm-specific conventions guide hypothesis specification and statistical testing loosely, and sometimes not at all, connected to the conceptual model's process assumptions. As a result, hypotheses and statistical models often track procedural features of the task rather than variables implied by the theory, allowing effects to replicate robustly while remaining theoretically underconstrained. This can create the impression of cumulative support for a conceptual model even when the empirical tests do not instantiate its core process assumptions. Whereas Borsboom et al. (2021) and other authors primarily diagnose a missing-theory problem, paradigm blindness concerns a breakdown in the translation of existing conceptual models into empirically constraining hypotheses and analyses.

We illustrate this issue in research on predecisional information distortion, a field that has developed the stepwise evolution-of-preference (SEP) paradigm (Russo, 2015) to investigate how individuals evaluate sequentially presented information about decision options and how emerging preferences bias information processing in favor of the emerging preference. Although coherence-based accounts motivating this literature posit parallel preference-evaluation dynamics (DeKay, 2015), hypotheses and statistical analyses within the SEP paradigm model information distortion as a function of preference states from the preceding round (Blanchard et al., 2014). As a result, statistical tests target a sequential process that is not implied by the conceptual model; statistical models may reproduce the effect, but in ways that are not clearly grounded in the model's process assumptions.

Illustrating Paradigm Blindness in Information Distortion Research

Information distortion refers to the systematic tendency to evaluate sequentially presented information in a manner that aligns with emerging preferences (Russo, 2015; Russo et al., 1996, 1998), driven by the tendency toward coherent interpretations of conflicting information (Heider, 1946; Holyoak & Simon, 1999; Read et al., 1997; Thagard, 1989). Over the past three decades, this phenomenon has been documented across diverse domains, including preferential decision making (Carlson et al., 2006; Chaxel et al., 2013; Russo et al., 1998), medical decision making (Kostopoulou et al., 2012, 2017; Nurek et al., 2014), risky choice (DeKay et al., 2009, 2011; Miller et al., 2013), and legal decision making (Carlson & Russo, 2001).

Early on in this line of research, the field established conventions in the stepwise evolution-of-preference paradigm regarding how to conduct studies on information distortion, what to measure, and which statistical procedures to use to investigate it (Russo, 2015; Russo et al., 1996, 1998). In SEP tasks, information about competing options is presented sequentially. After each piece of information, participants (a) evaluate the currently presented information and (b) report their current preference. Information distortion is quantified by comparing each participant's evaluation of a given piece of information to an unbiased baseline obtained from a control group, in which the same information is evaluated without forming preferences. The resulting difference score captures the degree to which information is evaluated more favorably for the currently preferred option (pro-leader distortion) or more unfavorably for the non-preferred option (anti-trailer distortion; Blanchard et al., 2014).

The conventional statistical procedure in SEP research relates information distortion at the current round of information presentation to the preference reported at the end of the previous round, thereby operationalizing distortion as a function of prior preference (e.g., DeKay et al., 2011; Russo, 2015). Across numerous studies, this analytic approach has yielded robust evidence for ID. Thus, research on information distortion has successfully achieved the first milestone of theory development (Borsboom et al., 2021): establishing a robust empirical phenomenon together with a widely shared measurement and analysis convention.

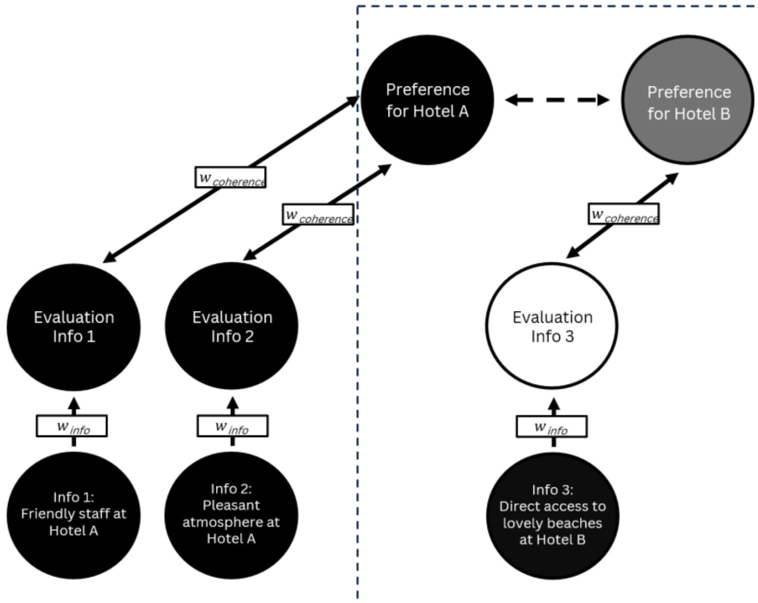
Evaluating whether statistical analyses meaningfully test a conceptual process model requires separating two issues. First, statistical models must adequately reflect the data structure imposed by the paradigm. In research on information distortion, this requirement was explicitly addressed by DeKay et al. (2009), who demonstrated that separation of within- and between-subjects variation in ID and accounting of dependencies in the observations requires multilevel modeling—a practice we also adopt in our analyses. Second, and more fundamentally for theory development, the specification of statistical predictors must correspond to the states and dependencies posited by the conceptual process model. As we elaborate next when introducing the conceptual model, it is precisely this second requirement that has not been systematically met in the SEP literature.

Building on these empirical findings, research on ID has advanced toward a mechanistic explanation through coherence-based process models of decision making (DeKay, 2015; Read et al., 1997; Simon et al., 2004). In these models, decision making is conceptualized as parallel constraint satisfaction within a network of mutually interacting representations. Preferences and information evaluations are dynamically coupled: emerging preferences bias the interpretation of new information, while newly interpreted information feeds back into preference updating.

DeKay (2015) provided an explicit network model for information distortion in sequential decision tasks. Figure 1 illustrates the core architecture of this conceptual process model. Information strength of each piece of information, as measured in a control condition, forms the input layer of the network. These inputs feed into a layer of subjective evaluations via weighted connections w_{info}. Subjective evaluations in turn connect to option nodes representing the competing choice alternatives through bi-directional coherence-based weighted connections $w_{coherence}$. The option nodes are negatively interconnected, as indicated by a dashed line, reflecting mutual exclusivity of the final choice. As each new piece of information is processed, activation spreads bidirectionally between option and evaluation nodes. Consequently, each information item is evaluated under the joint influence of its information strength and the current preference state, while at the same time contributing to preference updating, with the ratio between weights w_{info} and $w_{coherence}$ determining the relative influence of information and preferences on information evaluation.

Figure 1.

Conceptual Network Model of Coherence-Based Predecisional Information Distortion, adapted from DeKay (2015, Fig. 1, p. 406).



Note. A person has received two positive pieces of information about Hotel A in the previous two rounds and has formed an initial relative preference for Hotel A over Hotel B (black vs. gray option nodes), and then receives, in the third round of information presentation (dashed box), a positive piece of information about Hotel B. The conceptual model predicts that the evaluation of the new information (white node) depends on its strength (input layer) and on the already established preference for Hotel A over Hotel B (output layer), while simultaneously influencing preferences for the hotels through the coherence-based coupling of emerging preferences and subjective information evaluation.

In coherence-based network models, the evaluation of newly encountered information is not determined solely by a preference state carried over from the previous round. Instead, in the network state depicted in Figure 1, information evaluation and preference updating unfold in parallel through bidirectional interactions between evaluation nodes and option nodes. In the example, the model predicts that newly encountered information favoring the nonpreferred option (Hotel B) will be evaluated less positively than its information strength would warrant, reflecting anti-trailer distortion. Conversely, if the newly encountered information were to favor the currently preferred option, the same parallel updating mechanism would predict an evaluation that is more positive than warranted by its information strength (pro-leader distortion; Blanchard et al., 2014; DeKay, 2015; Simon et al., 2004).

Thus, coherence-based network models provide a conceptual process account of how dynamic preference–evaluation coupling gives rise to information distortion, marking a second milestone of the research program in the form of a prototheory (Borsboom et al., 2021) following the establishment of a robust empirical effect. A natural next step would have been to subject the model's implications—specifically, the parallel dynamics of information distortion and preference updating that are at the core of this model class—to empirical test, either through targeted new studies or by reanalyzing published data. Instead, subsequent studies adhered to paradigmatic conventions that operationalized distortion as a function of prior preference (e.g., Doh et al., 2025; Kostopoulou et al., 2017), thereby at least implicitly retaining a serial interpretation of the process.

Aligning Hypotheses and Statistical Tests with the Conceptual Process Model

Theoretical progress before formalization can already be achieved by completing a third milestone of theory development at the conceptual level—that is, by translating an existing conceptual process model into explicit, model-implied hypotheses and corresponding statistical tests. Systematically deriving statistical tests in this way sharpens empirical predictions, reveals previously unexamined implications, and subjects the conceptual process model to empirical tests that are diagnostic of its process assumptions.

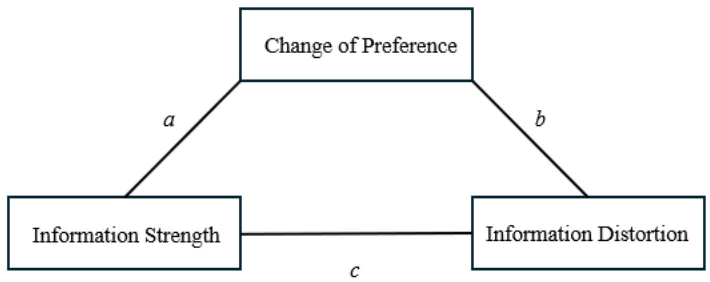
A direct implication of a parallel-processing account of information distortion is that within-round preference change, rather than prior preference alone, should be systematically related to both the magnitude and direction of ID. This implication is empirically informative in situations where serial and parallel models yield divergent predictions for information evaluation. Under a serial model, preferences are assumed to remain stable during information processing, such that ID is fully determined by the preference state at the beginning of the round. In contrast, a parallel model allows preferences to evolve and update while the information is being evaluated and therefore predicts that ID covaries with how preferences shift within the round. Preference change is thus not a secondary outcome of the decision process but a model-diagnostic predictor of ID.

From the conceptual model, we derive three related effects that capture distinct manifestations of within-round preference dynamics. First, the *push effect* refers to increases in the magnitude of pro-leader and anti-trailer distortion when preference for the leading option strengthens within a round due to the presentation of preference consistent information. Second, the *pull effect* refers to corresponding decreases in information distortion when preference weakens due to the presentation of preference inconsistent information. Third, the *reversal effect* occurs when preference changes sign, such that the previously trailing option becomes the leader due to preference inconsistent information, resulting in a predicted reversal in the direction of information distortion. Together, these effects specify how both the magnitude and the direction of ID depend on within-round preference dynamics. By contrast, a serial model, as implied by the conventional analysis relating preferences from the previous round to ID in the current round of information presentation, predicts that changes in preferences in the current round do not correlate with the strength of ID and preference reversals do not lead to directional changes in ID.

A common challenge in testing push, pull, and reversal effects arises from the temporal structure in the SEP paradigm: Information is evaluated before preferences are reported at the end of the round. Consequently, within-round preference change cannot be treated as a simple causal antecedent of ID without violating measurement order. To address this challenge, we employ a mediation-based identification strategy (Figure 2) that maps directly onto the assumed parallel-processing dynamics. In this model, information strength—as determined independently in the control condition—serves as an exogenous driver. The a-path captures how the strength and valence of the current information update preferences within the round. The b-path reflects how these preference updates covary with the magnitude and direction of ID, corresponding to the predicted push, pull, and reversal effects. The c-path represents the total effect between information strength and ID, which is expected to be partially accounted for by within-round preference change.

Figure 2.

Mediation-based identification of within-round preference–evaluation dynamics.



Note. The figure illustrates the mediation-based identification strategy used to test within-round preference–evaluation dynamics. Information strength refers to the objective strength and valence of the currently presented information (as assessed in the control condition). Preference change represents the difference in preference from the previous to the current round. The *a*-path reflects how information strength affects preference change within a round, the *b*-path reflects how preference change is related to information distortion (corresponding to push, pull, and reversal effects), and the *c*-path represents the total relation between information strength and information distortion. Preference change is treated as an indicator of within-round dynamics rather than as an exogenous cause.

Crucially, preference change is not interpreted as an exogenous cause of ID in this model but as a process indicator of the evolving preference-evaluation coupling within a round. From this perspective, the mediation model is not a substantive theory of information distortion but a formal test strategy that allows us to recover parallel preference–evaluation dynamics—including both graded changes in distortion magnitude and qualitative reversals in distortion direction—despite the serial ordering of measurements in the SEP paradigm.

Overview of the Present Research

The present research tests three theoretically diagnostic predictions derived from a parallel-processing account of ID: (a) Push and pull effect, reflecting how within-round preference strengthening or weakening modulates the magnitude of information distortion; (b) the reversal effect, reflecting a change in the direction of distortion when preferences switch; and (c) a mediation-based test of parallel processing. We conducted three preregistered experiments in which participants made binary decisions in three health- and consumer-related domains (dentists, hotels, Bluetooth speakers) while information was presented sequentially. Following DeKay et al. (2014), all studies used an item-wise presentation format rather than the attribute-wise format typical of earlier SEP studies (Russo et al., 1996, 1998, 2008). Presenting information about each option separately allows pro-leader and anti-trailer distortion to be examined independently and makes within-round preference dynamics observable at the level required by the parallel-processing account.

Study 1 provides an initial test of the push and pull effects while replicating established pro-leader and anti-trailer distortion. Study 2 extends this test by covering the full range of information strength, including weak and neutral information, and by testing, as an a-priori hypothesis, whether preference reversals are accompanied by corresponding reversals in the direction of distortion. Study 3 addresses a key alternative explanation—selection-based differences between individuals who do versus do not change their preference—by using a within-participants design that evaluates distortion relative to participants' own unbiased baseline evaluations.

We also test whether the three effects of parallel processing generalize beyond our studies tailored to detect them. To this end, we complement the experimental studies with reanalyses of published information-distortion datasets from other labs originally collected for different purposes.

Transparency and Openness

The studies were reported in accordance with APA Journal Article Reporting Standards (JARS). All studies were preregistered prior to data collection, and preregistration documents are accessible via the Open Science Framework (OSF). Materials, anonymized data, analysis scripts, power analyses, and figures for Studies 1–3, as well as the reanalyzed datasets, are publicly available through the same OSF repository. An anonymous view-only link is provided for review purposes: (project overview: https://osf.io/28z7n/overview?view_only=2772a72c358243ae899dac2bb39124e2). The repository is organized into functional components (Final Analysis, Data Preprocessing, Materials, Software, and Figures), each subdivided into study-specific folders ("Study 1," "Study 2," "Study 3," and "Reanalysis"), enabling full reproducibility of the reported results. For transparency and ease of navigation, direct links to study-specific preregistrations and materials are provided in Appendix A.

Ethical Considerations

The present studies involved anonymous, minimal-risk behavioral procedures. Participants provided informed consent prior to participation. Data were stored without reference to participants' identities. All procedures complied with APA ethical standards for research with human participants.

Study 1

Study 1 tests push and pull effects predicted by a parallel-processing account of information distortion, while replicating established pro-leader and anti-trailer distortion.

Method

Participants

Two hundred ten people participated in the experiment (M_{age} = 28.1 years, SD_{age} = 18.7, mainly students from the University of Cologne), 154 of which were female. Recruitment of participants was organized using the software hroot (Bock et al., 2014) using participant pool of the Decision Lab Cologne. The study took roughly about 25 minutes and participants received a fixed payment of 6 Euros (approx. 6.1 USD) for participation.

Sample size determination was guided by two considerations. First, the primary analyses include a mediation model, for which previous work suggests that a minimum sample size of approximately 100 participants is required to detect at least moderate indirect effects with adequate power (Koopman et al., 2015). Second, to ensure sufficient power to replicate the established interaction between preference strength and pro-leader versus anti-trailer information distortion, we conducted a bootstrap-based power analysis using pilot data. At N = 100 participants, all bootstrap samples yielded a statistically significant regression coefficient for the interaction effect. Given uncertainty about effect sizes for the more complex mediation and interaction effects, and because sufficient resources were available, we conservatively doubled the target sample size.

Design

The hypotheses were tested in a fully crossed 3 (decision category) x 2 (leader vs. trailer) within-subject design. The dependent variable was ID, measured according to the well-established SEP-paradigm. Participants completed six decision tasks, with two tasks in each decision context. Each task involved a binary choice between two options and the presentation of eight information items, four describing each option. Information was presented sequentially, with one information item per option at each step, allowing distortion of information about the leading and trailing option to be examined separately.

Within each decision context, the two tasks differed in the net valence of the available information. In one task, both options had a neutral net valence, consisting of two positive and two negative information items per option. In the other task, both options had a mildly positive net valence, consisting of three positive and one negative information item per option. This manipulation was implemented to ensure that, in at least some trials, participants were required to choose between two desirable alternatives.

Materials and Procedure

Decision Contexts and Information Pool

Decision tasks were based on a validated pool of information items describing unfamiliar options in three decision contexts (hotels, dentists, Bluetooth speakers). All information items were adapted from real online customer reviews and referred to concrete, decision-relevant attributes of a single option (e.g., a review indicating that Hotel A has a "lovely beach with clear-blue water").

Following the SEP paradigm, the stimulus pool was developed and validated in a control group in an independent online pre-study ($N = 210$; 157 female; $M_{age} = 27.77$, $SD_{age} = 9.01$). In this pre-study, participants rated a total of 144 information items (48 per decision context) independently, without forming preferences or making choices. Ratings were provided on a continuous scale ranging from −100 (strong negative influence) to +100 (strong positive influence), indicating how strongly each piece of information would influence a decision for or against the respective option. This procedure yielded undistorted, preference-independent estimates of information valence and evaluative strength, which served as the baseline for calculating ID.

For Study 1, information was selected to systematically vary in valence (positive, negative, neutral) and strength. Items were paired across options within each decision task by matching information pieces with highly similar mean ratings and confidence intervals and, where necessary, adjusting pairings to ensure concent consistency, resulting in alternatives that were approximately equally attractive overall while preserving sufficient variability for preference formation and change. Full details on stimulus development, validation, and matching procedures are available on OSF.

Labels for Decision Options

Because the formation of preferences requires identifiable options, a pre-test was conducted to screen for neutral fictitious names used in the experiment to label options. In a pilot study, thirty-three participants (21 female; $M_{\text{age}} = 28.87$, SD = 8.69) rated 20 randomly generated, pronounceable nonwords following German phonotactic rules from three content domains (hotels, dentists, Bluetooth speakers) on a −100 to +100 scale, indicating "how strongly this name would influence [them] personally in a potential decision for or against the option." Only names with mean ratings not significantly different from zero and narrow confidence intervals were retained. For the current study, four names (two per context) were selected to ensure neutral labels.

Procedure

The study was conducted online using Qualtrics. After a brief practice task, participants completed six decision tasks, two in each of the three decision contexts. Both the order of decision contexts and the sequence of information within each context were randomized. In each task, eight information (four per option) was presented sequentially. After each information, they rated its influence on their decision on a continuous scale from –100 (strong negative influence) to +100 (strong positive influence). Participants then indicated their current preference between the two options on a continuous scale from –100 (strong preference for Option A) to +100 (strong preference for Option B), with the left/right orientation of option labels randomized. After all information for a given decision had been presented, participants reported their final choice and their confidence in that choice on a continuous scale ranging from 50 (absolutely toss-up) to 100 (absolutely confident).

Results

All statistical analyses were conducted using R (R Core Team, 2019). Visualizations were created with the *ggplot2* package (Wickham, 2016), and mixed-effects models were estimated using the *lme4* package (Bates et al., 2015) in conjunction with *lmerTest* (Kuznetsova et al., 2017). Analyses were conducted using multilevel regression models to account for repeated observations within participants. All models included random intercepts for participants. Unless noted otherwise, effects are reported as fixed effects with corresponding standard errors and test statistics.

Proleader and Antitrailer ID

We first tested whether ID differed systematically for information about the leading versus the trailing option. ID was defined as the deviation between a participant’s evaluation of an information item and the item's baseline evaluation obtained in the independent pre-study. Information was classified, following the SEP paradigm, as leader or trailer information based on the participant's preference state at the end of the preceding round.

We estimated a multilevel regression model with ID as the dependent variable and information type entered as an effect-coded predictor (leader = +1, trailer = −1), including random intercepts for participants. As predicted, ID was on average 6.35 points higher for information about the leading option than for information about the trailing option ($b = 3.17$, $t(7541) = 8.51$, $p < .001$). Complementarily, evaluations of leader information were more positive than baseline control evaluations ($b = 2.59$, $t(188.92) = 3.84$, $p < .001$), whereas evaluations of trailer information were substantially more negative ($b = -3.99$, $t(193.53) = -4.64$, $p < .001$), as tested with two intercept-only random intercept multilevel models.

To examine whether preference strength in the preceding round modulated the contrast between pro-leader and anti-trailer distortion, we extended the baseline model by including absolute preference strength from the previous round (mean-centered) and its interaction with information type. The interaction was not statistically significant ($b = 0.02$, $t(7555) = 1.60$, $p = .11$), indicating that preference strength alone did not reliably amplify the pro-leader versus anti-trailer distortion contrast.

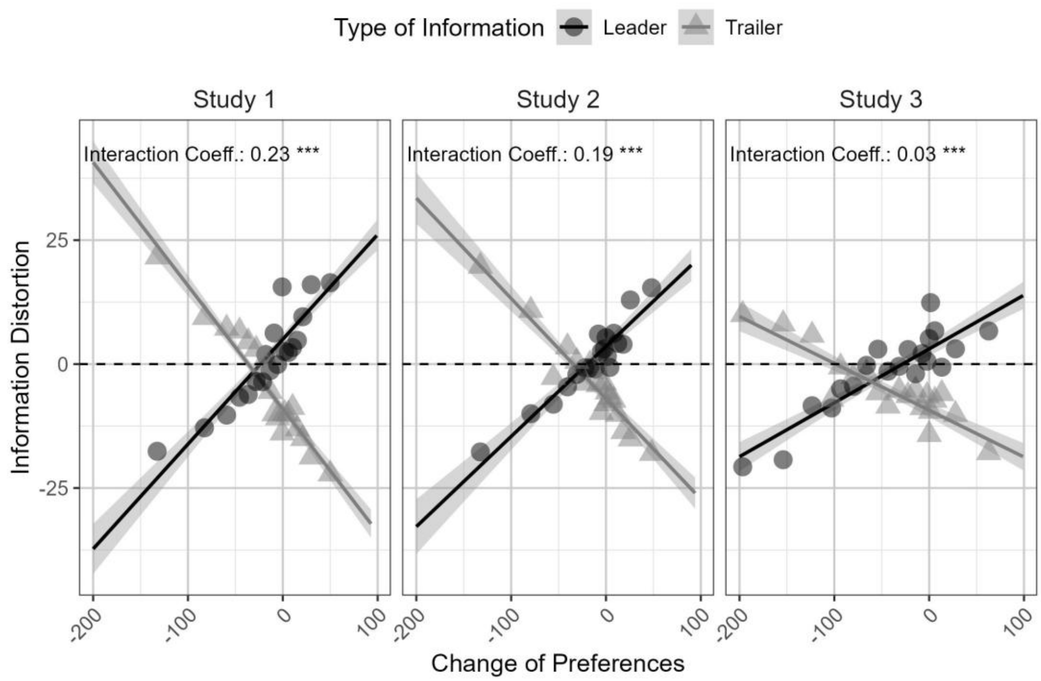
Push & Pull Effects

To test whether within-round preference change is associated with information distortion, we computed the change in preference strength between the previous round and the current round for each observation. Preference change was mean-centered and entered into a multilevel regression model predicting ID, with information type effect-coded (leader = +1, trailer = −1) and random intercepts for participants.

As preregistered, preference change significantly moderated the contrast between pro-leader and anti-trailer distortion ($b = 0.23$, $t(7544) = 26.41$, $p < .001$). Figure 3 (left panel) illustrates that when preferences strengthened during the processing of new information, the magnitude of pro-leader and anti-trailer distortion increased (push effect), whereas when preferences weakened, the magnitude of distortion decreased (pull effect).

Figure 3.

Preference Change and Information Distortion Across Studies.



Note. Panels display Studies 1–3. Lines depict linear model fits for *Leader* versus *Trailer* information based on trial-level data; shaded bands indicate 95% confidence intervals. Points show binned means of distortion across 20 equally sized intervals of preference change. The dashed horizontal line marks zero distortion. Reported values denote the interaction coefficient between preference change and information type for each study (*** $p < .001$).

In an exploratory analysis, we tested whether this interaction persists when accounting for the influence of prior preference strength on leader versus trailer distortion. To this end, we additionally included the preference strength from the preceding round and its interaction with information type. Both interactions were statistically reliable: Prior preference strength moderated the leader–trailer contrast ($b = 0.13$, $t(7548) = 10.41$, $p < .001$), and the interaction between preference change and information type remained robust ($b = 0.26$, $t(7552) = 28.69$, $p < .001$). Thus, within-round preference change explains systematic variation in information distortion that is not fully captured by models conditioning only on prior preference strength.

Mediation-Based Test of Parallel Processing

To address the temporal ordering problem inherent in the SEP paradigm and to provide a causal test of within-round dynamics, we estimated a mediation model in which information strength of the current information served as an exogenous predictor, within-round preference change indexed preference updating as a mediator, and ID constituted the outcome (see Figure 2). All analyses were conducted at the level of individual information evaluations and included random intercepts for participants. Specifically, the *a*-path tested whether information strength predicted changes in preference from the preceding to the current round, whereas the *b*-path tested whether preference change predicted ID while controlling information strength. The indirect effect thus captures the extent to which information strength affects ID via its impact on preference updating, as predicted by a parallel-processing account.

The mediation analysis revealed a robust indirect effect of information strength on ID via preference change (ACME = 0.12, 95% CI [0.11, 0.14], $p < .001$). More positive information reliably shifted preferences within the round, which in turn amplified ID. A significant residual direct effect of information strength on ID remained (ADE = −0.07, 95% CI [−0.09, −0.06], $p <$

.001), indicating partial mediation. Together, these results show that within-round preference change accounts for a substantial portion of the effect of information strength on ID.[1] This mediation pattern supports the assumption that preference updating and information evaluation unfold in parallel within a round, rather than in a strictly serial sequence.

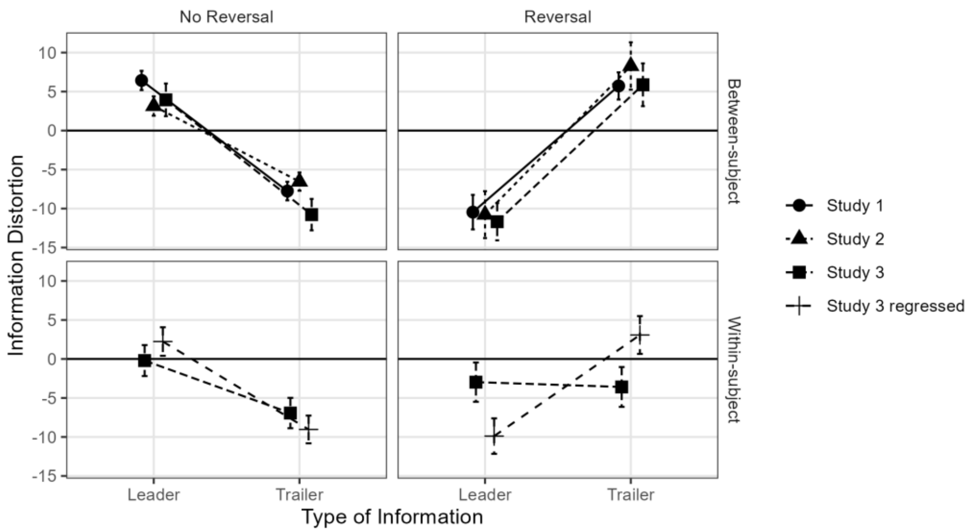
Exploratory Analysis: Reversal Effect

Although Study 1 was preregistered to test push and pull effects, the conceptual model also predicts a reversal of information distortion when preferences switch between options within a round. We therefore conducted an exploratory test of this prediction. Preference reversals were defined as cases in which the preferred option changed from the preceding to the current round and were effect-coded (reversal = 1, no reversal = −1). As predicted by a parallel-processing account, information distortion reversed when preferences reversed (see Figure 4, top panel). Specifically, information about the option that had been preferred in the preceding round was distorted negatively, whereas information about the previously non-preferred option was distorted positively ($b = -7.67$, $t(7553) = -18.29$, $p < .001$).

[1] Detailed path-level analyses corresponding to preregistered hypotheses are reported in the OSF repository.

Figure 4.

ID by Leader vs. Trailer Information and Preference Reversals across three Studies.



Note. Points show mean distortion scores with 95% confidence intervals, plotted separately for the between-subject designs (Studies 1–2) and the within-subject design (Study 3), and for trials without versus with a preference reversal (columns). Distortion is defined as the deviation of an evaluation from its baseline. In Studies 1–2, the baseline is provided by an independent control condition without prior preference commitment. In Study 3, distortion reflects change from each participant’s pretest evaluation of the same statement. For comparability, Study 3 is additionally displayed in a between-subject operationalization using the same independent control group as in Studies 1–2. The data labeled *Study 3 (RTM-corrected)* display distortion scores after controlling for regression to the mean. Specifically, following Yu and Chen (2015), pretest–posttest change scores were regressed on mean-centered pretest evaluations in a linear mixed-effects model with random intercepts for participants. The fixed effect of pretest on change (i.e., the expected regression-to-the-mean component) was subtracted from the raw change scores, yielding RTM-adjusted distortion estimates.

Critically, this effect contextualizes a null result reported above: When testing whether preference strength in the preceding round modulated the strength of pro-leader versus anti-trailer distortion without accounting for preference reversals, the interaction was not statistically significant (p = .11). However, when preference reversals were included as a moderator in an extended exploratory model, preference strength reliably interacted with information type conditional on whether preferences reversed (three-way interaction: $b = -0.11$, $t(7556) = -7.98$, $p < .001$).

Together, these findings indicate that averaging across qualitatively distinct preference dynamics—specifically, reversal versus non-reversal cases—can obscure theoretically meaningful relations between preference strength and information distortion. This pattern is consistent with the notion that incomplete predictor specification may mask effects predicted by a parallel-processing account.

Discussion of Study 1

Study 1 provides initial support for the three theoretically diagnostic predictions derived from a parallel-processing model of information distortion. We find evidence for push and pull effects, indicating that within-round preference change modulates the magnitude of distortion, as well as for a mediation-based test showing that the effect of information strength on distortion is accounted for by within-round preference change. An exploratory analysis further revealed a reversal effect: When preferences switched within a round, the direction of information distortion reversed accordingly. Together, these findings indicate that information evaluation covaries with dynamically emerging preferences within a round, consistent with parallel-processing accounts and inconsistent with a serial model as implied by standard statistical procedure in the SEP paradigm.

Exploratory analyses additionally suggest that neglecting preference reversals may obscure theoretically relevant relations between preference strength and distortion. Although preference strength alone did not reliably modulate the pro-leader versus anti-trailer contrast, it did so when preference reversals were considered. While exploratory, this pattern highlights the importance of adequately capturing within-round preference dynamics.

Study 1 is subject to an important limitation. Decision tasks exclusively involved moderately to highly diagnostic information according to control ratings, restricting the range of information strength available for testing the mediation framework. Study 2 addresses this limitation by also including weak and neutral information, enabling a more comprehensive test of the proposed model and assessing the extent to which the results generalize across information strength ranges.

Study 2: Extending the Test to Weak and Neutral Information

Study 2 addresses this limitation by systematically extending the range of information strength to include weak and neutral information. This extension enables a more comprehensive test of the mediation-based account and assesses whether the observed within-round preference dynamics generalize beyond highly diagnostic decision environments. In addition, we varied the number of information items presented to test whether effects generalize beyond that factor. We also preregistered the reversal effect for Study 2, allowing for a confirmatory test of this prediction and replication of the effect.

Participants and Design

Two-hundred sixty-one participants (162 female; M_{age} = 29.02 years, SD_{age} = 10.34, 5 participants did not indicate their gender) took part in Study 2. Participants were again recruited via the Cologne Decision Lab using the hroot recruitment system (Bock et al., 2014), and none had participated in Study 1. The study was conducted online using Qualtrics, took about 20 minutes and participants received a show-up fee of 4€ (~ 4.33 USD).

Sample size was determined a priori to ensure sufficient power to replicate the central effects observed in Study 1 under a modified study design. Although the overall design followed Study 1, only one decision in each context was used to reduce participant workload. To this end, we conducted a bootstrap power analysis based on a subset of Study 1 data that matched the present design (i.e., one decision task per context). The analysis focused on the key effects of interest: (a) Push and pull effects of within-round preference change on information distortion, and (b) the indirect effect of information strength on information distortion via preference change. At a target sample size of 250 participants, more than 95% of bootstrap samples yielded statistically significant effects.

Materials

Study 2 closely followed the design of Study 1 but extended the material to cover a broader range of information strengths, including neutral and weak information, to address a key limitation of the initial study. To generate less diagnostic information items, we conducted an additional pre-study in which 30 weak to neutral pieces of information were developed and pre-rated. This independent pre-study involved $N = 214$ participants (136 female; $M_{age} = 29.25$, $SD_{age} = 8.92$). For each decision category, participants rated 30 pieces of information, consisting of 10 neutral items with slightly positive valence, 10 neutral items with slightly negative valence, and 10 ambivalent items combining positive and negative aspects of varying strengths (2 to 4 items for weak, medium, and strong predicted strength).

In the main study, decision tasks covered the same three contexts as in Study 1 (hotels, Bluetooth speakers, dentists). Information items were selected from the combined stimulus pool developed across both pre-studies. Within each context, six item pairs were created by matching items of the same valence with comparable diagnosticity – i.e., minimal differences in their mean ratings averaged across both pre-studies, yielding one pair for each combination of strength (strong, medium, neutral) and valence (positive, negative).

Procedure

The procedure for Study 2 largely followed Study 1. Participants completed one decision task in each of three decision contexts (hotels, Bluetooth speakers, and dentists). All information was presented sequentially in an item-wise format. To extend the range of information strength, the number of information items presented per option varied across decision contexts. Participants received either four, five, or six pieces of information per option, resulting in sequences of 8, 10, or 12 information items presented in individually randomized order. Sequence length was randomly assigned across the three contexts such that each participant was exposed to each sequence length once. As in Study 1, after each piece of information, participants rated its influence on their decision on a continuous scale ranging from –100 (strong negative influence) to +100 (strong positive influence). Immediately thereafter, participants indicated their current preference between the two options on a separate screen using a continuous scale ranging from –100 (strong preference for Option A) to +100 (strong preference for Option B), with scale orientation randomized. After completing each decision task, participants reported their final choice and confidence.

Results

Proleader and Antitrailer ID

We estimated a multilevel regression model with ID as the dependent variable and information type entered as an effect-coded predictor (leader = +1, trailer = −1), including random intercepts for participants. Replicating Study 1, ID differed reliably between information about the leading and trailing option. On average, ID was 5.03 points higher for leader than for trailer information ($b = 2.52$, $t(6124) = 6.16$, $p < .001$). Relative to baseline evaluations, ID for leader information unexpectedly did not differ reliably from zero ($b = 0.97$, $t(248) = 1.46$, $p = .15$), whereas ID for trailer information was significantly negative ($b = -4.03$, $t(245) = -6.42$, $p < .001$), again indicating robust anti-trailer distortion.

To examine whether preference strength in the preceding round modulated the contrast between pro-leader and anti-trailer distortion, we extended the baseline model by including absolute preference strength from the previous round (mean-centered) and its interaction with information type. In contrast to Study 1, this interaction was statistically significant ($b = 0.05$, $t(6150) = 3.57$, $p < .001$), indicating that stronger preferences in the preceding round were associated with a larger divergence between leader and trailer distortion.

Push & Pull Effects

To test whether within-round preference change is associated with information distortion, we computed the change in preference strength between the previous round and the current round for each observation. Preference change was mean-centered and entered into a multilevel regression model predicting ID, with information type effect-coded (leader = +1, trailer = −1) and random intercepts for participants. Replicating the results of Study 1, preference change significantly moderated the contrast between pro-leader and anti-trailer distortion ($b = 0.19$, $t(6171) = 18.77$, $p < .001$). When preferences strengthened during the processing of new information, the magnitude of pro-leader and anti-trailer distortion increased (push effect). When preferences weakened, the magnitude of distortion decreased (pull effect; see Figure 3, middle panel).

In an exploratory analysis, we examined whether this interaction remains significant when accounting for the influence of prior preference strength on leader versus trailer distortion by additionally including prior preference strength and its interaction with information type. Both interactions were statistically reliable: Prior preference strength moderated the leader–trailer contrast ($b = 0.13$, $t(6154) = 9.47$, $p < .001$), and the interaction between preference change and information type remained robust ($b = 0.22$, $t(6170) = 20.83$, $p < .001$). Thus, as in Study 1, within-round preference change explains systematic variation in information distortion that is not fully captured by models conditioning only on prior preference strength.

Reversal Effect

As in Study 1, preference reversals were defined as cases in which the preferred option changed from the preceding to the current round and were effect-coded (reversal = 1, no reversal = −1). Consistent with predictions from a parallel-processing account, information distortion reversed when preferences reversed. Specifically, information about the option that had been preferred in the preceding round was distorted negatively, whereas information about the previously non-preferred option was distorted positively ($b = -7.09$, $t(6181) = -13.05$, $p < .001$). For completeness, we conducted an exploratory extension mirroring Study 1 by testing whether preference strength interacted with information type conditional on preference reversals. The corresponding three-way interaction was again significant ($b = -0.09$, $t(6162) = -4.55$, $p < .001$).

Mediation-Based Test of Parallel Processing

As in Study 1, we addressed the temporal ordering problem inherent in SEP tasks by estimating a mediation model in which the information strength of the current information served as an exogenous predictor, within-round preference change indexed preference updating as the mediator, and ID constituted the outcome (see Figure 2). All analyses were conducted at the level of individual information evaluations and included random intercepts for participants.

The mediation analysis again revealed a robust indirect effect of information strength on ID via within-round preference change (ACME = 0.12, 95% *CI* [0.11, 0.13], $p < .001$). More positive information reliably shifted preferences within the round, which in turn amplified distortion. A residual direct effect of information strength on ID remained (ADE = −0.11, 95% *CI* [−0.14, −0.09], $p < .001$), indicating partial mediation. Importantly, the total effect of information strength on ID was small and not statistically reliable in Study 2, reflecting opposing direct and indirect pathways. As established in contemporary mediation theory (Hayes, 2009; MacKinnon et al., 2007), the presence of a significant indirect effect does not require a significant total effect and is fully consistent with situations in which multiple processes exert countervailing influences. From a process perspective, the indirect pathway is theoretically central: It captures the predicted mechanism by which information strength affects ID through its impact on within-round preference updating. Together, these results replicate the mediation pattern observed in Study 1 and provide converging evidence that preference updating and information evaluation unfold in parallel within a round, rather than in a strictly serial sequence.

Discussion of Study 2

Study 2 replicated and extended the central findings of Study 1 under conditions that covered the full range of information strength, including weak and neutral information. As in Study 1, within-round preference change robustly modulated the magnitude of information distortion (push and pull effects), preference reversals altered its direction, and mediation analyses demonstrated that information strength affects distortion indirectly via within-round preference updating. Together, these results reinforce the conclusion that information distortion reflects dynamically evolving preferences rather than being fully determined by prior preference states.

Interim Discussion

Across two studies, we find converging evidence for three diagnostic predictions derived from a parallel-processing account of information distortion: (a) distortion scales with within-round preference strengthening or weakening (push and pull effects), (b) distortion reverses direction when preferences reverse, and (c) information strength affects distortion indirectly via within-round preference updating. These effects generalize across decision domains and procedural variations, indicating that information evaluation tracks dynamically emerging preferences within a round.

Alongside these findings, following the SEP paradigm, Studies 1 and 2 use a control group to calculate information distortion; thus, from a measurement perspective, it is unclear to what extent differences between ratings in the experimental group and ratings in the control group reflect information distortion rather than genuine interindividual differences in information evaluation. To address this issue, we deviated from the typical SEP design and used a within-subjects design to track differences in information evaluations within participants in Study 3.

Study 3: A Within-Participant Test of Preference Dynamics

Although Studies 1 and 2 provide converging evidence for dynamic preference-dependent information distortion, an alternative explanation remains. Observed differences in information evaluation between reversal and non-reversal cases could reflect selection effects rather than information distortion: some participants may a priori evaluate preference-inconsistent information more extremely due to genuine differences in how they value information (e.g., some participants may value a lovely beach at the trailing hotel more strongly than others), which both increases the likelihood of a preference reversal and leads to more extreme information ratings. In this case, differences in information evaluations relative to a control group would precede—and potentially drive—preference change rather than result from distortion during the decision process.

Study 3 addresses this concern by adopting a within-participant design, thereby deviating intentionally from the typical control-group design in the SEP paradigm. By comparing each participant's information evaluations under preference-consistent versus preference-inconsistent states to their own preference-independent baseline evaluations, this design allows us to isolate changes in distortion attributable to preference dynamics while controlling for stable individual differences in evaluative tendencies. This design strengthens the internal validity of the inferences about whether information distortion arises from within-round preference updating, as predicted by parallel-processing models.

Participants and Design

Three-hundred fifty-one participants (174 female; M_{age} = 54.49 years, SD_{age} = 14.41) were recruited using the provider Toluna (www.toluna-group.com). The study was conducted online and programmed in Qualtrics and lasted approximately 30 minutes; compensation was administered through the panel provider. Sample size was determined a priori using a bootstrap power analysis based on a subset of Study 1 data matched to the present design (one decision task in each of the two contexts). With N = 250, at least 95% of bootstrap samples yielded statistically reliable estimates for the central preference-dynamics effects (leader–trailer distortion and their modulation by preference change and reversals). Because the within-participant design required complete data, and the task was cognitively demanding, and therefore potentially noisier, we targeted a final sample of approximately N = 350 participants.

Study 3 used a fully crossed within-participant 2 (decision context: hotels vs. dentists) × 2 (information type: leader vs. trailer) design. Participants first provided preference-independent baseline evaluations of all 24 items, completed a filler task, and then performed the SEP-based decision tasks (as in Studies 1–2). ID was computed within participants as the deviation between each information evaluation in the decision phase and the same participant's baseline evaluation of the same information.

Materials

Stimulus Materials

Study 3 used the exact same 12 validated matched item pairs from Study 2 for the two remaining contexts (hotels, dentists). Each pair consisted of two items of the same valence with comparable diagnosticity (i.e., minimal differences in their mean ratings averaged across the two pre-studies), covering all combinations of evaluative strength (strong, medium, neutral) and valence (positive, negative). All 12 information items per context were used in the baseline phase, allowing preference-independent evaluations of each information. In the subsequent decision phase, four of the six matched pairs (eight information in total) were randomly selected per context and assigned to the two decision options, ensuring that decision-phase evaluations could be directly compared to each participant's own baseline ratings for the identical information items.

Procedure

The study consisted of three consecutive phases: a baseline phase, a filler phase, and a decision phase. In the baseline phase, participants rated all available information without making any choices. The items from the two decision contexts (hotels, dentists) were presented blockwise. Within each block, the 12 information items were presented in individually randomized order, with each information item referring to a different, unrelated option from the respective context. Participants rated how strongly each information would influence a decision for or against the respective option on a continuous scale ranging from −100 (strong negative influence) to +100 (strong positive influence). This phase provided preference-independent baseline evaluations for each information. Next, to minimize carryover effects between baseline evaluation and the decision phase, participants completed a brief working-memory filler task lasting approximately 8–10 minutes. The task comprised standardized forward and backward digit span trials (16 each) and one cycle each of 1-back and 2-back tasks, serving as a temporal and cognitive buffer between phases.

In the following decision phase, participants completed two decision tasks—one per decision context—implemented using the SEP paradigm as in Studies 1 and 2. For each context, eight information items (four per option) were randomly selected from the set of 12 pre-evaluated information and presented sequentially in an individually randomized order. After each information, participants rated its influence on their decision using the same scale as in the baseline phase and indicated their current preference between the two options. After all information had been presented, participants reported their final choice as well as their confidence in that choice.

Results

Proleader and Antitrailer ID

As in Studies 1 and 2, we estimated a multilevel regression model with ID as the dependent variable and information type entered as an effect-coded predictor (leader = +1, trailer = −1), including random intercepts for participants. In contrast to the previous studies, ID was computed as a within participant deviation from each participant's own preference-independent baseline evaluation.[2] Replicating the basic ID effect, ID differed reliably between information about the leading and trailing option, with ID on average 4.38 points higher for leader than for trailer information ($b = 2.22$, $t(4024) = 3.97$, $p < .001$). Relative to baseline and replicating results from Study 2, ID for leader information did not differ significantly from zero ($b = -1.30$, $t(328) = -1.25$, $p = .21$), whereas ID for trailer information was significantly negative ($b = -5.69$, $t(287) = -6.93$, $p < .001$), again indicating robust anti-trailer distortion.

To examine whether preference strength in the preceding round modulated the contrast between pro-leader and anti-trailer distortion, we extended the baseline model by including absolute preference strength from the previous round (mean-centered) and its interaction with information type. In contrast to Study 2 but in line with Study 1, this interaction was not statistically significant ($b = 0.01$, $t(4009) = 0.66$, $p = .51$), indicating that prior preference strength did not amplify the divergence between leader and trailer distortion in the within-participant design.

Push & Pull Effects

To test whether within-round preference change is associated with ID, we again computed the change in preference strength between the previous round and the current round for each observation. Preference change was mean-centered and entered into a multilevel regression model predicting ID with information type effect-coded (leader = +1, trailer = −1) and random intercepts for participants. Replicating the central pattern observed in Studies 1 and 2, preference change significantly moderated the contrast between pro-leader and anti-trailer

[2] As a robustness check, we additionally replicated the main analyses using the mean-based distortion measure employed in Studies 1 and 2. The pattern of results remained unchanged (see full analysis script on OSF).

distortion ($b = 0.03$, $t(4004) = 3.47$, $p < .001$). When preferences strengthened while processing new information, the magnitude of pro-leader and anti-trailer distortion increased (push effect). When preferences weakened, the magnitude of distortion decreased (pull effect).

In an exploratory extension, we additionally accounted for the influence of preference strength from the previous round by including it and its interaction with information type as predictors. In this extended model, preference strength reliably moderated the leader–trailer contrast ($b = 0.05$, $t(3984) = 2.53$, $p = .011$), while the interaction between preference change and information type remained robust ($b = 0.04$, $t(3979) = 4.24$, $p < .001$). This pattern indicates that effects of preference strength from the previous round on information distortion become more clearly detectable when within-round preference change is explicitly modeled, consistent with a parallel-processing account in which stable preference states and dynamic preference updating jointly shape information evaluation.

Reversal Effect

Similar to Studies 1 and 2, preference reversals were defined as cases in which the preferred option changed from the preceding to the current round and were effect-coded (reversal = 1, no reversal = −1). Consistent with predictions from a parallel-processing account, information distortion reversed when preferences reversed. Specifically, information about the option that had been preferred in the preceding round was distorted more negatively, whereas information about the previously non-preferred option was distorted more positively. This interaction was statistically significant in the within-subject analysis ($b = -1.60$, $t(3982) = -2.79$, $p = .005$) and is reflected in the lower within-subject panel of Figure 4.

For completeness, we conducted an exploratory extension mirroring the analyses in Studies 1 and 2 by additionally accounting for prior preference strength. The three-way interaction between information type, preference reversal, and prior preference strength was statistically significant ($b = -0.06$, $t(4014) = -3.61$, $p < .001$), indicating that the magnitude of the reversal effect varied with the strength of the preference carried into the round. As in previous studies, this analysis serves as a consistency check rather than as a reinterpretation of the preregistered effect.

Because Study 3 involved repeated evaluation of identical information (baseline and decision phase), we conducted an exploratory analysis to examine potential regression-to-the-mean (RTM) effects following the procedure recommended by Yu and Chen (2015). Specifically, change scores (decision-phase evaluation minus baseline evaluation) were modeled as a function of mean-centered baseline evaluations in a linear mixed-effects model with random intercepts for participants. Baseline evaluations were mean-centered across all observations to operationalize extremity relative to the overall mean. As expected under RTM, baseline evaluations strongly predicted change, $b = -0.20$, $t(4043) = -24.38$, $p < .001$, indicating that initially extreme ratings exhibited systematic shifts toward the mean independent of preference dynamics.

For descriptive purposes, we additionally constructed RTM-adjusted distortion scores by subtracting the fixed-effect component of change predicted by baseline evaluations from the raw change scores. The resulting variant ("Study 3 regressed") is displayed in Figure 4. The reversal pattern remains visible and is numerically more pronounced after adjustment.

Mediation-Based Test of Parallel Processing

As in Studies 1 and 2, we addressed the temporal ordering problem inherent in SEP tasks by estimating a mediation model in which the objective evaluative strength of the current information served as an exogenous predictor, within-round preference change indexed preference updating as the mediator, and ID constituted the outcome (see Figure 2). All analyses were conducted at the level of individual information evaluations and included random intercepts for participants.

The mediation analysis again revealed a reliable indirect effect of information strength on ID via within-round preference change (ACME = 0.06, 95% CI [0.05, 0.07], $p < .001$). More positive information shifted preferences within the round, which in turn amplified information distortion. A residual direct effect of information strength on ID was also observed (ADE = −0.26, 95% CI [−0.28, −0.25], $p < .001$). Consistent with Study 2, the direct and indirect paths pointed in opposite directions, resulting in a negative total effect of information strength on ID. As discussed above, this pattern reflects the operation of countervailing processes during information evaluation and is fully consistent with a parallel-processing account. Together, these results extend the mediation pattern observed in Studies 1 and 2 to a within-participant design and provide further support for the assumption that preference updating and information evaluation unfold in parallel within a round, rather than in a serial sequence.

Discussion of Study 3

Study 3 extended the previous findings by providing a within-participant test of preference-dependent information distortion. By comparing each participant’s information evaluations during decision making to their own preference-independent baseline evaluations, this design controls for stable individual differences in evaluative tendencies and directly addresses concerns about selection bias. Despite this stricter test, the central qualitative patterns observed in Studies 1 and 2 were replicated.

Consistent with predictions derived from a parallel-processing account, Study 3 again revealed push and pull effects, such that within-round preference changes were associated with corresponding shifts in information distortion. In addition, a reliable reversal effect was observed: When preferences switched, the direction of distortion reversed accordingly, with information about the newly preferred option evaluated more positively and information about the previously preferred option evaluated more negatively. These findings demonstrate that preference dynamics shape information evaluation even when comparisons are made within individuals.

At the same time, observed effects were attenuated relative to Studies 1 and 2. This pattern is theoretically informative rather than problematic. Study 3 employed a within-participant design in which the same information was evaluated twice—once during a preference-independent baseline phase and once during the decision phase. Such repeated measurement is known to introduce regression-to-the-mean (RTM) effects (Campbell et al., 2003; Yu & Chen, 2015), which operate independently of preference dynamics and reduce the observable magnitude of preference-consistent shifts in evaluation. As shown in the Results and illustrated in Figure 4 (bottom panel), RTM was present in the data.

From a process perspective, differences across designs are therefore expected. Whereas between-participant designs allow coherence dynamics to unfold without repeated measurement of the same information, within-participant designs introduce baseline-dependent correction effects that limit the observable magnitude of preference-dependent distortion. The persistence of distortion effects in Study 3 despite this attenuation indicates that the underlying coherence processes are robust and likely underestimated in within-participant designs.

Parallel Processing in Information Distortion: Evidence from Reanalyzed Datasets

The three experimental studies reported above provide converging evidence for parallel processing dynamics in information distortion derived from a coherence-based conceptual processing model, including reversal effects, push and pull dynamics, and mediation via within-round preference updating. These studies were specifically designed to isolate and test these mechanisms. If parallel processing reflects a general property of information integration rather than a feature of our specific paradigms, at least some of its core predictions should also be detectable in published datasets from studies not originally designed to test our parallel-processing account.

Importantly, most published datasets assess preferences dichotomously, precluding tests of graded push and pull dynamics. However, dichotomous preference measures allow identification of within-round preference reversals. Accordingly, the present reanalyses focus on the reversal prediction derived from coherence-based process models—specifically, whether information distortion changes direction when the leading option switches.

To this end, we reanalyzed five published studies conducted by three independent research groups: Miller et al. (2013; Experiments 1 and 2), Blanchard et al. (2014; Experiments 1 and 2), and Kostopoulou (2012; Experiment 1). These studies differ from the present experiments in various methodological respects, including content domain, task structure, measurement timing, and the absence of explicit manipulations targeting within-round preference dynamics. Key deviations are briefly summarized below.

Miller et al. (2013)

In Miller et al. (2013), Study 1, $N = 431$ participants were randomly assigned to one of five decision contexts (monetary gambles ($n = 86$), song downloads ($n = 94$), frequent flyer miles ($n = 83$), political decisions ($n = 83$), or medical decisions ($n = 85$)), or a no-choice control condition ($n = 84$). Each participant completed four decision tasks, with five attributes per option. Unlike our design, attributes were presented in a blockwise format, meaning information for both Option A and Option B on a given attribute was shown simultaneously before proceeding to the next. The attributes included two highly diagnostic, two weakly diagnostic, and one neutral information. The presentation order was partially randomized, with alternating positions for high- and low-diagnostic attributes. Participants evaluated each information on a single bipolar nine-point scale (1 = strongly favors Option A; 9 = strongly favors Option B), followed by a binary preference indication. Leader versus trailer classification was determined based on the preference indicated in the preceding round. Because each information was evaluated on a single bipolar scale anchored by the two options, positive evaluations for one option mean negative evaluations for the other, precluding a separation of pro-leader amplification and anti-trailer devaluation (see Blanchard et al., 2014). In Study 2, only the task with monetary gambles was used, with four fixed attribute orders randomly assigned to *n* = 171 participants in the experimental condition, with again *n* = 43 participants serving as a control. The task structure, evaluation scale, and preference recording were identical to Study 1.

Blanchard et al. (2014)

In Blanchard et al. (2014**)** Study 1, participants were assigned to either a choice condition (*n* = 90) with two decision contexts (backpacks vs. dry cleaners) or a no-choice control condition (*n* = 90). Participants completed one decision task per context. The number of attributes per option and the blockwise presentation format were the same as in Miller et al.'s studies, but all attributes had comparable levels of diagnosticity across both options. In contrast to Miller et al.'s single bipolar scale, each attribute was evaluated on two separate nine-point scales assessing the extent to which the information favored Option A and Option B, respectively. Following the evaluations, participants ranked the competing options according to their preference. In Study 2, Blanchard examined only one decision context (restaurants) with six decision options, including a no-choice control condition. The number of attributes and evaluation procedures mirrored those used in Study 1. As in Study 1, participants evaluated each option on two nine-point scales and provided a ranking of the six options according to their preference.

Kostopoulou et al. (2012)

In Kostopoulou (2012), the study focused on a single decision domain—medical decisions—where $n = 102$ physicians were presented with three patient scenarios (fatigue, chest pain, and dyspnea) and $n = 36$ participants served as unbiased control. For each scenario, they had to choose between two potential diagnoses based on a set of diagnostic information. Each scenario began with a diagnostic information to induce an initial preference, followed by a fixed order of information: (1) a diagnostic information favoring Diagnosis A, (2) a neutral information, and (3) in the final scenario, a diagnostic information favoring Diagnosis B. While the order of the information was fixed, the information within each set were presented in random order. Participants evaluated each scenario using a single 21-point scale (-10 = strongly favors Diagnosis A; 0 = favors neither; 10 = strongly favors Diagnosis B). Following the evaluation, they indicated their final preference on a 21-point scale, rating the likelihood of the cause for the patient's symptoms.

Results Reanalysis

To ensure direct comparability with the preregistered Studies 1–3, ID was reanalyzed using the same modeling strategy employed in the present experiments. Although ID was operationalized as in the original studies, the statistical approaches used to estimate distortion effects varied across publications (e.g., level of aggregation, contrast coding, or experimentally imposed versus preference-based leader definitions). The goal of the present reanalyses was therefore not to reproduce the originally reported test statistics, but to re-estimate distortion effects within a unified analytic framework aligned with the process-level predictions tested in Studies 1–3.

Across all reanalyses, relative distortion served as the dependent variable and was modeled as a function of information type (leader vs. trailer) and preference reversals. Both predictors were effect-coded following the conventions used in the preregistered experiments (leader = +1, trailer = −1; reversal = +1, no reversal = −1). Random intercepts for participants were included in all models. To facilitate comparison across studies using different evaluation scales, distortion scores were rescaled to ranging from −1 to 1, preserving the direction and relative magnitude of distortion while eliminating scale differences.

Replication of Pro-Leader and Anti-Trailer Distortion Effects

We first examined whether the reanalyzed datasets replicate information distortion observed in the preregistered experiments—namely, more favorable evaluation of information about the currently leading option relative to information about trailing alternatives. To illustrate the robustness of this effect across different measurement formats, we highlight results from one study using a single bipolar evaluation scale (Miller et al., 2013, Study 1) and one study using alternative-specific evaluations on separate scales (Blanchard et al., 2014, Study 1), which closely mirrors the measurement approach employed in the present experiments. In Blanchard et al. (2014), Study 1, information about the leading option was evaluated more favorably than information about trailing options, $b = 0.03$, $t(595) = 4.75$, $p < .001$. A comparable pro-leader distortion effect emerged in Miller et al. (2013), Study 1, $b = 0.06$, $t(3614) = 19.99$, $p < .001$. The same qualitative pattern was observed across the remaining reanalyzed datasets—including Blanchard et al. (2014), Study 2; Miller et al. (2013), Study 2; and Kostopoulou et al. (2012), Study 1 (all *ps* < .001)—despite substantial differences in task structure, number of options, and evaluation format. Thus, across heterogeneous decision contexts, study designs, and measurement approaches, the reanalyses consistently replicated a distortion bias−also reported in the respective articles−favoring information about the currently preferred option.

Preference Reversals and the Reversal of Distortion

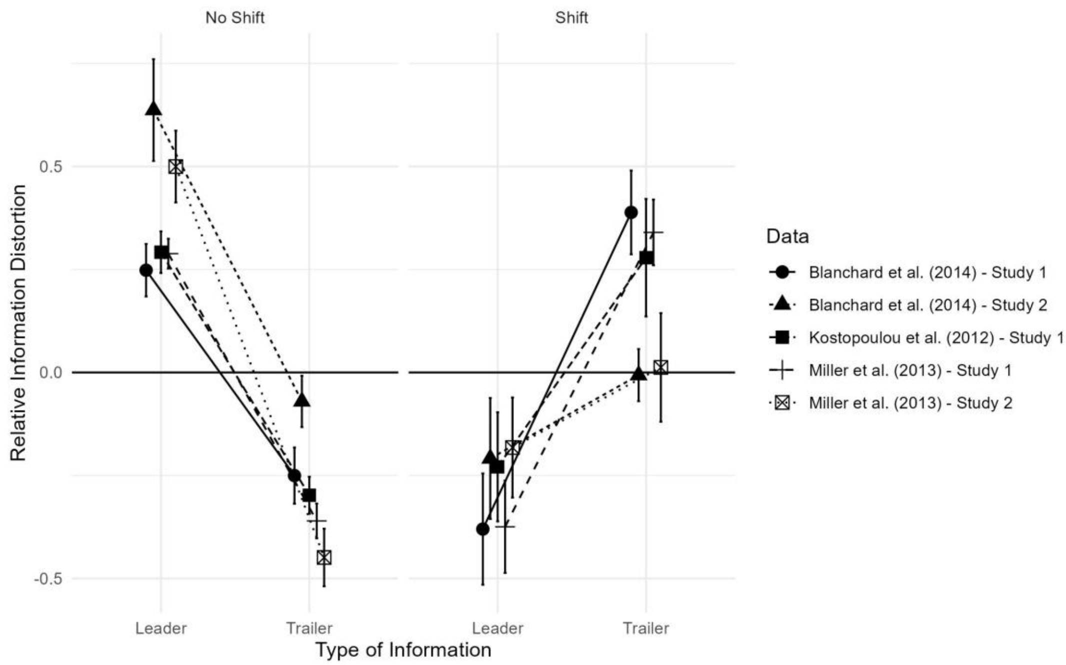
The central prediction of coherence-based accounts is that information distortion is dynamically linked to emerging preferences and should reverse when preferences reverse. To test this prediction, we estimated mixed-effects models including the interaction between information type and preference reversal.

Consistent with this prediction, a robust interaction between information type and preference reversal emerged across all reanalyzed datasets. In Blanchard et al. (2014), Study 1, the interaction was significant, $b = -0.071$, $t(593) = -11.64$, $p < .001$, indicating that the direction of distortion reliably reversed following a preference shift. The same pattern was observed in Miller et al. (2013), Study 1, $b = -0.087$, $t(3827) = -21.78$, $p < .001$. All remaining studies showed the same reversal effect (all *ps* < .001).

Figure 5 visualizes this pattern by plotting mean relative distortion separately for trials with and without preference reversals. Across all datasets, distortion favored the currently leading option and systematically reversed when preferences changed, in line with coherence-based process accounts.

Figure 5.

ID for Leader vs. Trailer information in Five Reanalyzed Studies, Moderated by Preference Reversals.



Note. The figure shows study-level mean relative distortion scores from the reanalyzed datasets, standardized within each study to ensure comparability across rating scales and decision contexts. Points correspond to estimated marginal means for leader and trailer information, and error bars denote 95% confidence intervals. Standardization preserved the direction and relative magnitude of distortion while eliminating scale differences across studies.

General Discussion

Summary of Findings

Across three preregistered experiments using complementary between- and within-subjects designs, all three predictions following from a parallel-processing account of predecisional information distortion were supported. Information distortion increased when preference-consistent information strengthened the currently leading option (push effect), decreased when preference-inconsistent information weakened existing preferences (pull effect), and reversed direction when the leading option changed within a round (reversal effect). Consistent with this interpretation, mediation analyses across all three studies showed that information strength predicted information distortion mediated by preference updating within the round.

Complementary reanalyses of five published datasets showed that distortion was systematically aligned with the currently preferred option and reversed following preference shifts. This pattern emerged consistently across decision domains, measurement formats, participant samples, and research groups. The convergence of these findings indicates that the reversal effect constitutes a robust, model-implied dynamic that generalizes beyond the present experiments and remains observable even in datasets not originally designed to test this prediction.

Operational Bias in Estimates of Information Distortion

Building on the qualitative pattern summarized above, we discuss the quantitative consequences of misalignment between statistical operationalization and the conceptual process model. When information distortion is operationalized in a manner that does not correspond to the conceptual model, resulting estimates do not accurately reflect the effect the model is intended to capture. Consequently, effect sizes of information distortion reported in the literature may be biased. We examine the extent of this potential bias using all eight studies reported in this article, either novel or reanalyzed. We report two quantities as effect size measures of information distortion (Table 1).

Table 1.

Biased and Unbiased Estimation of Information Distortion Metrics Across Studies.

Study	Domain(s)	*N*	Prevalence of Hypothesis-Consistent Information Distortion (% of Trials)		Information Distortion (Mean % of Maximum Possible Deviation)	
			Biased	Unbiased	Biased	Unbiased
Blanchard et al. (2014), Study 1	Consumer	180	57.5	65.6	2.62	6.24
Blanchard et al. (2014), Study 2	Consumer	121	52.8	53.6	3.91	1.98
Kostopoulou et al. (2012), Study 1	Medical	102	61.2	63.9	2.64	3.09
Miller et al. (2013), Study 1	Mixed	431	64.6	72.0	6.49	9.16
Miller et al. (2013), Study 2	Monetary	171	64.1	65.7	9.71	10.48
Study 1 in this article	Consumer/ Medical	210	54.1	60.9	5.87	11.31
Study 2 in this article	Consumer/ Medical	261	54.0	59.3	2.59	6.73
Study 3 in this article	Consumer/ Medical	351	51.3	61.1	4.54	11.51

Note. For Study 3, values are based on regression-adjusted estimates controlling for regression-to-the-mean effects in the within-participants design.

The prevalence of hypothesis-consistent information distortion denotes the proportion of trials in which participants evaluated preference-consistent information more positively and preference-inconsistent information more negatively. The biased estimate, following the SEP paradigm, defines information as preference-consistent or inconsistent based on the preference indicated in the prior round. The unbiased estimate, consistent with the conceptual process model, diverges from the biased estimate when preferences reverse within the current round. In such cases, evaluations are classified relative to the updated preference state rather than the outdated preference from the previous round, resulting in different trial classifications and, consequently, different prevalence estimates.

The second indicator reflects the operationalization of information distortion typically reported in studies using the SEP paradigm: the mean deviation of information evaluations from a control-group baseline, expressed as a proportion of the maximum possible deviation on the rating scale (Russo, 2015). When preference reversals are ignored, as in the SEP paradigm, trials in which distortion favors the newly preferred option are coded relative to an outdated preference state and therefore enter the analysis with the opposite sign relative to the conceptual model (i.e., yielding a biased estimate). Averaging across such misclassified trials attenuates the mean estimate.

Across all eight datasets, both indicators converge on the same pattern: analyses that ignore preference reversals consistently underestimate information distortion. Within each dataset, we computed the difference between unbiased and biased estimates and then averaged these differences across datasets, weighting by sample size.

For the prevalence of hypothesis-consistent distortion, reversal-ignorant analyses yield a weighted mean of 57.0%, indicating that evaluations favored the preferred option in 57% of trials. Incorporating reversals increases this estimate by 6.45 percentage points, resulting in an adjusted weighted mean of 63.5%, indicating that preference-consistent distortion occurs in 63.5% of trials when within-round reversals are taken into account.

For distortion magnitude (expressed as percentage of the maximum possible deviation), reversal-ignorant analyses yield a weighted mean of 4.29% of the maximum possible scale deviation. Incorporating reversals increases this estimate by 3.89 percentage points, resulting in an adjusted weighted mean of 8.19%, indicating a substantially larger average deviation in evaluations toward the preferred option when reversals are modeled. Thus, reversal-ignorant estimates (4.29%) fall slightly below the lower bound of values typically reported in the SEP literature (≈5%–40%; Russo, 2015), whereas reversal-adjusted estimates (8.19%) lie within that descriptive range. The consistency of this upward shift across decision domains, measurement scales, and independent research groups suggests that the attenuation reflects a structural consequence of reversal-ignorant analytic specifications rather than idiosyncratic properties of particular datasets.

Taken together, this pattern indicates that how information distortion is operationalized shapes how the phenomenon appears. Conventional SEP analyses define distortion relative to prior preference states and summarize it as a single aggregate effect. This operationalization conflates qualitatively distinct dynamic regimes—preference strengthening, weakening, and reversal—thereby compressing heterogeneous processes into a single summary index. Modeling within-round preference change makes this structure explicit: preference reversals function as a previously unmodeled moderator, and distortion covaries not simply with initial preference states but with how preferences evolve during information processing. Aggregating across these regimes preserves the overall direction of the effect but obscures its internal structure and attenuates its magnitude.

Paradigm Blindness in Analytic Practice

The preceding section discussed that conventional SEP-based operationalizations systematically attenuate estimates of information distortion by aggregating across qualitatively distinct within-round dynamics. This pattern points beyond effect estimation to the analytic practices that generate these estimates. It raises the question of how such practices can persist even when a conceptual process model is available that implies a different structure.

Empirical findings are typically presented as observations about the phenomenon under study. Yet the statistical models used to generate those findings are not theoretically inert. Statistical operationalizations encode structural assumptions about the phenomenon under study: they determine which states are represented, how those states are temporally ordered, and which dependencies are permitted. Statistical models are therefore not neutral containers of effects; they instantiate commitments that shape which aspects of a process can be empirically detected. From a measurement-theoretic perspective, statistical models articulate substantive assumptions about the structure of the phenomenon under study rather than merely summarizing data (Borsboom et al., 2004).

When analytic conventions are inherited from paradigmatic practice rather than derived from the conceptual process model, these structural commitments may remain unarticulated. In the present case, modeling information distortion as a function of prior preference states implicitly assumes a temporally ordered, state-based process. Coherence-based accounts, by contrast, posit dynamic and parallel updating of preferences and evaluations within a decision round. Notably, this implied serial structure was neither endorsed nor discussed in the information-distortion literature; it entered through the analytic convention itself.

This pattern exemplifies what we refer to as paradigm blindness. Paradigm blindness does not arise from the absence of theory, nor from a lack of theoretical ambition or weak link between theory and hypothesis (Oberauer & Lewandowsky, 2019). Rather, it reflects the stabilization of analytic practices at the level of the paradigm, such that their structural implications escape scrutiny. Consequently, conceptual process models and statistical implementations can drift apart, allowing empirical findings to accumulate while central process assumptions remain only weakly tested.

Paradigm blindness thus constitutes a structural failure mode of theory testing. The phenomenon under study may be robust and replicable, thereby fulfilling all laudable criteria of reproducible research, yet the analytic framework through which it is examined may fail to preserve distinctions implied by the conceptual model used to explain it. Identifying and correcting such misalignments is therefore a prerequisite for empirical tests to become diagnostic of process-level theories. Theoretical progress can accordingly follow from refining how existing theories are tested. More broadly, analytic practice is not external to theorizing but part of it: analytic choices are theoretically consequential because they determine which aspects of a process model are supported, which remain untested, and where refinement is required.

Scope and Limits

The present work is situated at a stage of theory development that precedes formalization (Borsboom et al., 2021). As argued by Farrell and Lewandowsky (2010), limits of human cognition constrain verbal theorizing and can obscure the implications of complex process assumptions, motivating formalization because a model's predictions may only become clear once it is implemented (e.g., through computational simulation).

In the present case, direct formalization would likely have avoided the analytic mismatch discussed in this article. Coherence-based network models have already been implemented in formal and computational terms (Glöckner & Betsch, 2008; McClelland et al., 2014; Read et al., 1997; Thagard & Verbeurgt, 1998), and adapting and testing such a computational model directly would have made the implications of parallel processing explicit. The present analyses therefore do not challenge the value of formalization.

Instead, the present contribution addresses a distinct question: whether central, model-defining implications of a conceptual framework are supported by the statistical tests used to evaluate it, independent of specific functional forms or parameterizations in a formal model. Formal models require a range of auxiliary assumptions, and empirical misfit can arise from implementation-specific choices that are secondary to a theory's core commitments. In such cases, it remains unclear whether a model fails because a central theoretical principle is incorrect (e.g., the assumption of parallel preference–evaluation coupling underlying coherence effects) or because of a particular instantiation (e.g., the choice of update function, weighting scheme, noise specification, or parameter constraints).

More generally, many research areas involve conceptual process models that guide empirical work before a corresponding family of formal implementations exists. Under such conditions, early formalization introduces substantial auxiliary structure, increasing the risk that theoretical evaluation becomes dominated by implementation choices rather than by the theory's core commitments. The present work targets this intermediate stage of theory development by translating a conceptual process model into explicit, model-implied hypotheses and corresponding empirical tests.

Recent proposals for integrative theory building (Betsch & Bröder, 2025) likewise emphasize establishing support for basic principles (Van Lange et al., 2022) before introducing additional formal complexity. The analyses therefore focus on testing implications that are characteristic of a class of models—for coherence-based models, amplification, attenuation, and reversal under parallel processing—that must hold across admissible implementations. Establishing or rejecting these constraints provides a principled basis for subsequent formalization.

Conclusion

The present research shows that deriving statistical predictors directly from a conceptual process model reveals theoretically implied dynamics of information distortion that paradigmatic partially atheoretical analytic practices misrepresent and underestimate. Across preregistered experiments and reanalysis of published datasets, incorporating within-round preference dynamics yields a more complete and theoretically faithful account of coherence-driven distortion in sequential decision making. More broadly, these findings identify a critical but underappreciated step in theory development: ensuring that conceptual process assumptions are translated into statistical predictors in a way that preserves their model-implied dynamics. The present case study shows that completing this step—prior to any type of formalization—can expose empirically diagnostic patterns that remain invisible under paradigm-guided analyses, even in well-established research programs. Paradigm blindness, in this sense, does not reflect a shortage of data or theory, but a failure to align statistical models with the assumptions they are meant to test. Addressing this mismatch is essential for cumulative, process-sensitive theorizing in judgment and decision making and beyond.

Appendix 4A

Open Science Framework (OSF) Study-Specific Links

The project overview is accessible via the Open Science Framework (OSF):

https://osf.io/28z7n/?view_only=2772a72c358243ae899dac2bb39124e2

Preregistrations

Study 1: https://osf.io/ubqet/files/8hz9x

Study 2: https://osf.io/atndk/files/kry68

Study 3 (registration record): https://osf.io/7ps68/overview

Study 3 (archived preregistration PDF): https://osf.io/fua9t/files/8j32a

For Study 3, both the registration record and the archived preregistration document are provided to allow verification of the OSF timestamp and direct access to the frozen document.

Pre-Studies Diagnosticity of Information

Information pool validation (Pre-study 1):

https://osf.io/xpnqs/?view_only=1e3e91d19f5f4b89a0ec7f89ed151343

Information pool validation (Pre-study 2):

https://osf.io/f9rn5/?view_only=279620ab5ac0456394e7f01c8d755498

Pre-Study Neutral Option Labels

Neutral option label pre-study:

https://osf.io/fm295/?view_only=1faba23cf110478a836f1d949f46ea20

4.3 Summary and Implications

Chapter 4 demonstrates that coherence-driven information distortion in sequential decision making cannot be fully characterized by conditioning evaluations on prior preference alone. Building on coherence-based network models of sequential integration (DeKay, 2015), the chapter derives and tests dynamic predictions that follow directly from the assumption of parallel preference–evaluation updating. These include push effects, whereby distortion increases as preferences strengthen within a round; pull effects, whereby distortion attenuates as preferences weaken; and reversal effects, whereby changes in preference direction are accompanied by corresponding reversals in evaluative bias.

Across preregistered experiments and reanalysis of published datasets, the results show that these model-implied effects remain systematically underestimated—or statistically masked—under standard analytic practices in the SEP paradigm (Russo, 2015). Treating preferences as stable during information evaluation effectively imposes a serial approximation on a fundamentally parallel process. This operationalization obscures model-implied dynamics that only become visible when changes in preference strength are modeled explicitly. Consequently, distortion effects may partially cancel out at the aggregate level, leading to conservative and theoretically incomplete estimates of coherence-driven bias.

The present findings identify a specific translation bottleneck in the study of sequential decision making. Although coherence-based process assumptions are widely acknowledged, standard statistical operationalizations in the SEP paradigm fail to instantiate their core parallel updating dynamics. This misalignment does not reflect a lack of theory; rather, it demonstrates how even theoretically mature paradigms can obscure process-level effects when statistical models track procedural conventions more closely than underlying dynamics.

By making these dynamics statistically explicit, Chapter 4 clarifies what it means to test coherence-based models at the process level. Rather than requiring immediate computational formalization, theoretical refinement can proceed through disciplined derivation of model-implied predictors from conceptual process assumptions. This clarification provides the methodological groundwork for the subsequent chapter, which examines how additional constraints—such as stable ideological beliefs—interact with dynamic preference formation within the same coherence-based framework.

Chapter 5: Extending Coherence-Based Models: Top-Down Constraints

5.1 Chapter Introduction

Chapter 5 addresses the second major challenge identified in the Open Challenges sections: integrating stable belief systems into coherence-based models of sequential decision making. Whereas Chapter 4 focused on translating dynamic process assumptions into statistically diagnostic predictors, the present chapter examines how higher-order ideological commitments function as top-down constraints within an otherwise unified coherence-driven updating process.

Recent theoretical work proposes that partisan motivated reasoning and related political biases can be understood as manifestations of coherence-based constraint satisfaction, in which beliefs, preferences, and evaluations mutually influence one another (Oeberst et al., 2025; Simon & Read, 2025). Simon and Read (2025) emphasize that coherence formation is inherently temporal: networks evolve toward attractor states whose trajectory depends on the configuration and strength of prior commitments. Extending this logic, Oeberst et al. (2025) argue that belief-consistent information processing may constitute a special case of coherence-based reasoning and call for empirical tests that distinguish between predominantly unidirectional coherence shifts—where prior beliefs constrain evaluation—from bidirectional shifts in which belief-based constraints and evaluative processes reciprocally interact. They further recommend jointly examining belief-based constraints and evaluative distortion within process-tracing paradigms.

Despite these theoretical proposals, empirical research on political judgment has predominantly relied on outcome-level or pre–post designs that treat partisan bias as a relatively stable predecisional constraint. Such approaches leave open whether ideological priors impose a fixed directional bias on evaluation or instead modulate a dynamic updating process in which preferences and evaluations co-evolve over time. To our knowledge, no study has yet examined this interaction within a fully sequential, trial-level paradigm.

Chapter 5 addresses this gap by embedding ideological and partisan commitments within the sequential coherence framework developed in Chapter 4. Building on connectionist models of sequential information integration (DeKay, 2015), ideological priors are conceptualized as higher-order constraints within the same network that governs evaluative distortion. Using a sequential process-tracing paradigm, the chapter tests whether partisan bias is best characterized as a predominantly unidirectional imposition of belief-based constraints or as a dynamic, bidirectional interaction between top-down ideological commitments and bottom-up preference construction during ongoing information integration. Chapter 5.2 is based on the following article:

Häffner, C., Sperlich, L., & Lammers, J. (2026). *Top-down intrusions on coherence-driven distortion: Rethinking partisan motivated reasoning.* Manuscript submitted for publication.

Minor modifications were made to headings and formatting to align with the dissertation style. The content of the manuscript remains unchanged.

5.2. Manuscript II: Top-Down Intrusions on Coherence-Driven Distortion: Rethinking Partisan Motivated Reasoning

Abstract

Citizens often evaluate statements more positively if they think they were made by members of their own party than if they were made by opponents. This pattern is commonly interpreted as reflecting partisan motivated reasoning, in which people uncritically accept in-group information while rejecting opposing views. However, this literature typically treats such biases as stable tendencies, leaving the underlying evaluative process underspecified. The present research instead draws on work on predecisional information distortion, which shows that evaluative bias can arise from coherence-seeking during sequential information integration, even without strong prior beliefs. We propose a coherence-based network model that conceptualizes partisan bias as the outcome of interacting top-down ideological priors and bottom-up preference construction. Two experiments (N_{total} = 1,734) used a sequential information integration paradigm to test five sets of process-based predictions derived from that model. Participants sequentially evaluated statements by political candidates while preferences evolved stepwise over time. Strength and alignment of partisan cues were manipulated to vary the influence of ideological priors. Results show that evaluations are shaped jointly by ideological priors and emerging preferences: partisan cues amplified evaluative distortion in a top-down manner, while emerging preferences fed back into belief-based categorization. Conflicts between partisan labels and ideological content reliably redirected choices away from the option matching participants' political orientation. Even under explicit partisan labeling, trial-level changes in preference continued to predict evaluative distortion. Together, these findings suggest that partisan bias unfolds dynamically and reflects coherence-based information integration under ideological constraints.

Public Significance Statements

Our research shows that political bias in favor of statements by own party politicians and against opponents emerges dynamically as people consider information step by step: prior beliefs about politicians influence judgments, but emerging preferences also shape how new information is interpreted and which political side it seems to support. By linking research on political bias with research on preferences development during decisions, we demonstrate that partisan bias arises from the interaction between ideological beliefs and evolving preferences rather than from fixed partisan biases alone. This perspective helps explain why political judgments can shift and intensify as people encounter new information over time.

Keywords: motivated reasoning, coherence-based reasoning, information distortion, political judgment, partisan bias, sequential decision making

Top-Down Intrusions on Coherence-Driven Distortion: Rethinking Partisan Motivated Reasoning

People have evaluative biases; that is, they tend to view new information as confirming what they already believe (Festinger, 1957b; Lord & Taylor, 2009; Oeberst & Imhoff, 2023; Sanbonmatsu et al., 1998; Sassenberg et al., 2014; Zuckermann et al., 1995). Such biases are of particular concern when they arise in domains with direct societal implications. In political contexts, for example, citizens' evaluations of information are frequently attributed to partisan motivated reasoning, implying that judgments are shaped primarily by prior ideological commitments rather than by informational content (Campbell et al., 1960; Kunda, 1990; Lodge & Taber, 2013; Taber & Lodge, 2006).

Within this literature, partisan bias is typically conceptualized as a belief-driven phenomenon exerting a predominantly top-down influence on evaluation. In many studies, the strength of this bias is expected to vary as a function of relatively stable individual characteristics such as partisan identification (Ditto et al., 2019; Goren et al., 2009; Kim et al., 2010; Lord et al., 1979; Lord & Taylor, 2009; Taber & Lodge, 2006). In this sense, biased acceptance or rejection of information is often examined primarily as an outcome of prior beliefs rather than as a dynamic process unfolding during information integration. This is not to say that work in political psychology has failed to provide sophisticated insights in the psychological processes that shape this biased acceptance or rejection. For example, research has identified differences in cognitive effort or time spent on confirming or disconfirming messages in an ideological motivated manner. As another example, it has identified voters' tendency to select news sources in a biased manner that ensures that prior positions are likely to be supported (Iyengar et al., 2008; Iyengar & Hahn, 2009; Taber & Lodge, 2006). Nonetheless, that literature treats party identification (e.g. Democrat vs. Republican) and ideological self-identification (e.g. between liberals and conservatives) as the basis on which biases build. In other words, biases start *once* voters self-identify as liberal or conservative.

In contrast, insights from judgment and decision-making research demonstrate that systematic evaluative bias can also arise even in the absence of such prior beliefs. Work on predecisional information distortion shows that when information is encountered sequentially, emerging preferences bias the evaluation of subsequent evidence: information supporting the currently favored option is evaluated more positively, whereas information favoring alternatives is discounted (DeKay, 2015; Russo et al., 1996). Crucially, this bias unfolds dynamically during preference construction rather than reflecting the influence of pre-existing beliefs, suggesting that biased evaluation can emerge incrementally as preferences stabilize rather than being imposed all at once. For example, when choosing between two consumer alternatives (e.g., products or services), decision makers may initially approach the task without a clear preference. However, as product attributes are evaluated sequentially, even small emerging differences can lead one option to become temporarily preferred. Subsequent attribute information is then interpreted more favorably when associated with the currently leading option than when associated with the competing alternative, a pattern widely documented in research on predecisional information distortion (Russo et al., 1996, 2000).

This is not to say that political science has not tested how preferences react in response to differences in what candidate is currently leading or trailing. However, in the political scientific tradition the focus has been on influence by information about the candidates’ leading versus trailing position in the polls (Bartels, 1988; Simon, 1954). A phenomenon referred to as the *bandwagon effect* refers to the tendency that people sometimes follow the majority, for example because a leading candidate is likely to be a superior option (Ansolabehere & Iyengar, 1994; Barnfield, 2020; Bartels, 1988). Another phenomenon, which consists of shifts away from the majority and toward one of the trailing candidates, is typically referred to as the *underdog effect* (Ceci & Kain, 1982; Fleitas, 1971; Laponce, 1966). Critically, though, these shifts are shifts in response to other voters' opinions, meaning that they reflect a tendency to unquestionably follow the majority opinion as a heuristic for quality or reflect a tendency to move away from the majority to support trailing candidates out of a strategic, moral, empathic, or other motives (Lammers, Bukowski, et al., 2022; Vandello et al., 2007). This thus reflects a tendency to be influenced in response to a candidate's relatively stable leading or trailing position in the opinion of others (i.e., in the polls).

Thus, although both research traditions address biased evaluation, they differ fundamentally in their explanatory focus. Research on political motivated reasoning has often examined biased evaluation as a function of relatively stable partisan identification and as biases in response to the relatively stable aggregate of others' opinions (e.g. in the polls). In contrast, research on predecisional information distortion has focused on how evaluative asymmetries emerge dynamically during the sequential construction of preferences, within the same person and during the seconds or few minutes at most needed to do so. Because these literatures have largely developed in isolation, it remains unclear whether belief-driven and preference-driven biases reflect distinct cognitive mechanisms or different operating conditions of a shared underlying process. Consequently, it remains unclear to what degree bias in political judgment reflects relatively stable belief-driven tendencies or dynamic evaluative processes that unfold during sequential information integration.

This separation mirrors a broader theoretical challenge in the bias literature. Decades of research have documented systematic deviations from normative models of rational inference under the umbrella of biased reasoning (Arkes, 1991; Klayman, 1995; Nickerson, 1998), often discussed within the error paradigm (Funder, 1995). At the same time, this literature has been criticized for theoretical fragmentation and limited process specification (Simon & Read, 2025). As Glöckner and Betsch (2011) argue, verbal theories frequently accommodate findings post hoc but yield few precise and falsifiable predictions, leaving it unclear whether different biases reflect distinct mechanisms or context-dependent manifestations of more general principles. These concerns have motivated calls for more strongly specified process theories that generate precise predictions about how judgments and decisions evolve over time (Oberauer & Lewandowsky, 2019; Russo, 2015).

Responding to concerns about theoretical fragmentation and limited process specification in the bias literature, recent work has emphasized unifying, process-level accounts of biased reasoning. Coherence-based models constitute fully specified process models of judgment and decision making, that formalize how beliefs, preferences, and evaluations are jointly updated through parallel constraint satisfaction. These models can make predictions not only about choices, but also about process-sensitive parameters such as confidence, reaction times, and information search (Glöckner et al., 2010b; Glöckner & Betsch, 2008b; Holyoak & Simon, 1999; Simon, Snow, et al., 2004). Building on this process foundation, Simon and Read (2025) proposed coherence-based reasoning as an integrative framework for the bias literature, conceptualizing diverse reasoning biases as context-dependent manifestations of a shared coherence-seeking process. Related integrative approaches emphasize belief-consistent information processing and have outlined conditions under which belief-consistent distortions may be particularly likely, including strongly held or situationally dominant prior beliefs (Oeberst et al., 2025; Oeberst & Imhoff, 2023). The extent to which such conditions favor unidirectional distortions over bidirectional coherence dynamics in sequential information integration, in the context of reading political information, however, remains an open empirical question.

Introducing a Sequential Information Integration Paradigm in Political Judgment

The present research advances this theoretical agenda by testing coherence-based models in the context of people's judgment of political information, using a sequential information integration paradigm. Doing so is important because it can contribute to the literature in two ways. First, it strengthens theory integration by examining whether motivated reasoning and predecisional information distortion—traditionally studied in separate literatures—can be accounted for within a single coherence-based framework. Building on established coherence-based models that formalize predecisional information distortion as a systematic bias in information evaluation during sequential preference construction (DeKay, 2015) we use these models as a baseline for extending coherence-based accounts to contexts characterized by strong ideological priors. Second, the present work advances the motivated-reasoning literature by empirically testing whether partisan bias reflects predominantly belief-consistent distortions associated with stable partisan identification or instead dynamic coherence processes in which ideological priors interact with evolving preferences during sequential information integration.

Accordingly, the central contribution of this article is to show how belief-driven and preference-driven biases can be integrated within a common process model. Using political motivated reasoning as a critical extension case, we examine whether partisan bias can be formally understood as a political instantiation of coherence-based information integration under ideological priors, deriving process-based predictions about how beliefs, preferences, and evidence jointly evolve over time to shape political judgment. In this sense, the present work speaks directly to longstanding concerns about political ignorance (Bartels, 1996) by clarifying not only that political judgments are biased, but how and when such biases emerge during ongoing information integration.

In the remainder of this paper, we first provide a more detailed account of how political reasoning has traditionally been investigated. We then introduce a coherence-based model that highlights the dynamic development of preferences and beliefs over time. Building on established models of predecisional information distortion, this approach allows us to derive testable predictions about the emergence of partisan bias over time. These predictions are then examined through two experiments, which test whether partisan motivated reasoning constitutes a distinct cognitive process or reflects a graded manifestation of general coherence-driven mechanisms

From Outcome to Process-Level Accounts in Political Reasoning

Political motivated reasoning has been extensively documented as a robust and societally consequential phenomenon. Across paradigms, citizens evaluate political information more favorably when it aligns with their partisan or ideological commitments and more critically when it originates from opposing groups (Lodge & Taber, 2013; Taber & Lodge, 2006). Classic evidence suggests that partisan source cues can dominate evaluations even when policy content is held constant (Goren et al., 2009; Kam, 2005; Lord et al., 1979; Rahn, 1993), supporting the view that political judgment is strongly shaped by prior beliefs.

Interpretations of this pattern diverge. Some accounts emphasize that party labels can serve as heuristics that enable relatively efficient decisions in complex environments (Lupia, 1994; Popkin, 1991). People cannot be expected to form complex opinions on all political issues at hand. Simply following the heuristic that politicians from their own party are right and opponents wrong, allows them to arrive at a conclusion that they would have likely arrived at through detailed processing – but saves them the mental energy of having to actually do so.

On the other hand, others stress the normative risks: if evaluations primarily track group allegiance, citizens may fail to update beliefs in response to relevant evidence and may endorse positions that are poorly aligned with their own interests (Bartels, 2002; Iyengar et al., 2019; Van Bavel & Pereira, 2018). These concerns are amplified by findings that ideological orientations are often deeply entrenched—linked to dispositional, genetic, and structural correlates (Alford et al., 2005; Gerber et al., 2010; Hatemi et al., 2010, 2011; Kanai et al., 2011; Mondak & Halperin, 2008) and that partisan bias may even increase with political sophistication (Nyhan et al., 2013; Nyhan & Reifler, 2010; Taber & Lodge, 2006). Taken together, this literature is often read as implying a largely static, belief-driven distortion: people selectively evaluate evidence to defend what they already believe or desire to be true (Ditto & Lopez, 1992; Kunda, 1990; Taber & Lodge, 2006).

This conclusion currently rests, however, predominantly on outcome-level designs that capture political judgment as a single evaluative response. While such designs are well suited to documenting partisan bias, they offer limited insight into the processes through which judgments unfold—despite the fact that political information is typically encountered sequentially and preferences evolve over time.

In judgment and decision-making research, comparable limitations are now starting to be addressed only after process-oriented theories and methods have rendered within-deliberation variables theoretically interpretable. Shifts in attention, information search, or evaluation dynamics were long treated as noise until frameworks such as Query Theory and coherence-based network models identified them as signatures of preference construction (Glaholt & Reingold, 2011; Jekel et al., 2018; Johnson et al., 2007; Schulte-Mecklenbeck et al., 2017; Simon, Snow, et al., 2004; Zilker & Pachur, 2023). As Meehl (1967) noted, effects that are not tied to explicit theoretical predictions are easily dismissed rather than explained.

Reflecting this limitation, recent work has called for a shift away from adjudicating the (ir)rationality of partisan favoritism toward identifying its predictors and clarifying the cognitive mechanisms underlying motivated political reasoning (Ditto et al., 2025). Without an explicit process model, it remains unclear whether partisan bias reflects a fixed predecisional constraint imposed by ideological commitments or a dynamic outcome that emerges during information integration itself.

Sequential decision-making paradigms provide a principled way to address this gap. The stepwise evolution-of-preference paradigm (Russo, 2015; Russo et al., 1996, 1998)—which is the standard experimental approach for studying predecisional information distortion—tracks how preferences and evaluations co-evolve as information is encountered step by step. Applied to political judgment, this paradigm allows one to test whether partisan bias is imposed prior to information processing or instead unfolds dynamically as preferences develop.

Introducing a Coherence-Based Model of Partisan Motivated Reasoning

The present work seeks to bridge two traditions by testing whether political–ideological bias can be formalized as a political instantiation of the same coherence-driven process that underlies predecisional information distortion, or whether partisan motivated reasoning represents a qualitatively distinct bias. From a coherence-based perspective, decision making reflects an iterative constraint-satisfaction process that strives for internal consistency among beliefs, preferences, and evidence (DeKay, 2015; Glöckner & Betsch, 2008b; Holyoak & Simon, 1999; Simon, Snow, et al., 2004). Recent theorizing has formalized this view as a general framework of biased reasoning, in which both motivated and seemingly unbiased distortions emerge from a common drive toward cognitive coherence (Simon & Read, 2025).

Within Simon and Read's (2025) framework, mental representations are modeled as interconnected elements whose activation levels reflect the strength and salience of each belief, preference, or evaluation. Competing conclusions and their connected elements form "attractors"—areas in the network that gather strong activation and naturally pull the system toward them. Highly accessible or affectively charged beliefs—such as partisan identities—function as such attractors, shaping the interpretation of new information while simultaneously being reinforced by it. These dynamics help explain why partisans interpret identical information in ideologically congruent ways (Cohen, 2003; Meffert et al., 2006) and selectively accept worldview-consistent evidence while discounting opposing information (Hornsey et al., 2020; Kahan et al., 2012). Importantly, ideological priors do not unilaterally determine interpretation: new information can feed back into the network, modulating belief activation through parallel updating. This bidirectionality distinguishes coherence-based reasoning from purely belief-consistent accounts by capturing the dynamic interplay between stability and change in political judgment.

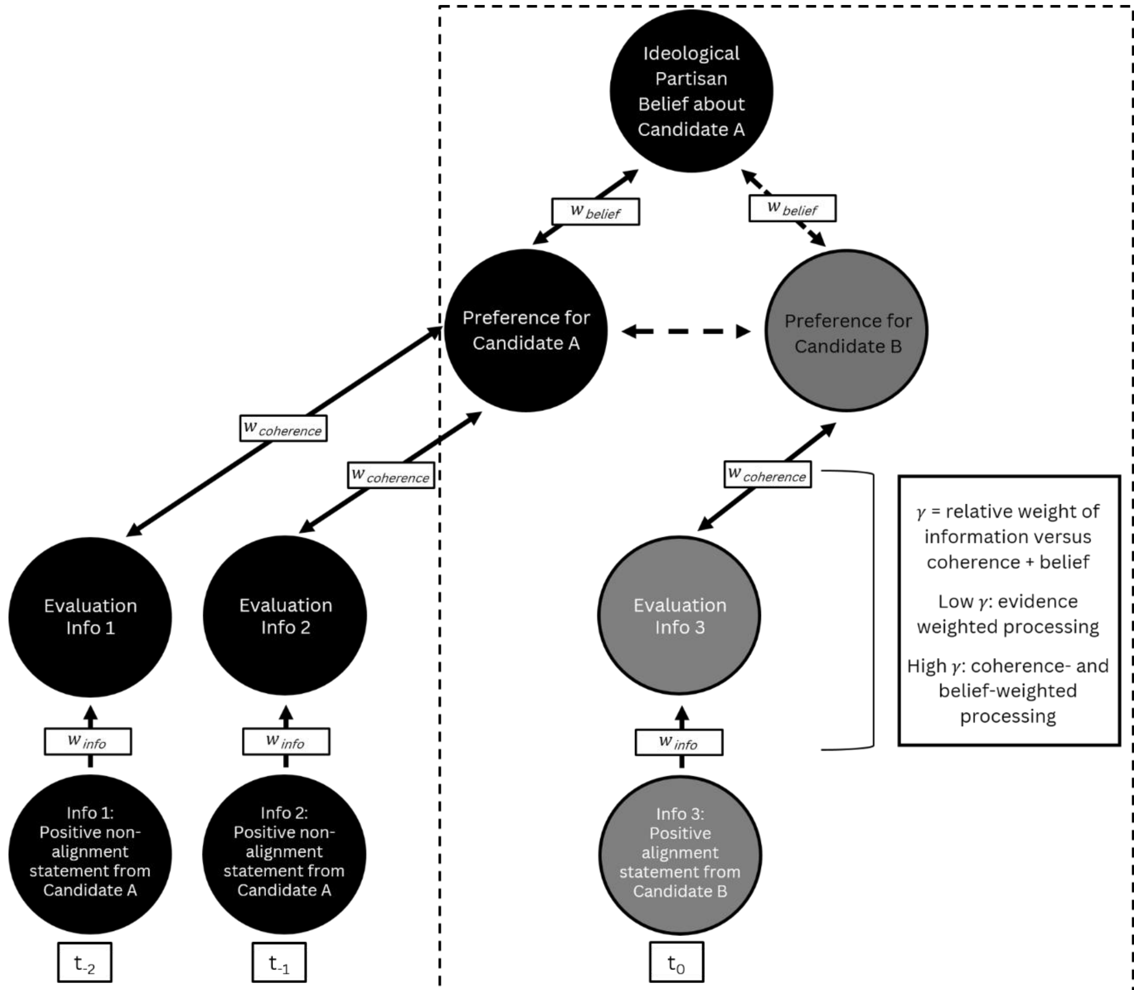
Building on these principles, the present model extends the connectionist framework of predecisional information distortion proposed by DeKay (2015). In DeKay's original framework, evaluations of information and evaluations of choice options mutually influence one another through bidirectional connections, such that incoming information shapes emerging preferences, which in turn bias the evaluation of subsequent information. The system gradually converges on a coherent evaluative state in which evidence, preferences, and judgments are aligned.

The key extension introduced here is the incorporation of ideological–partisan beliefs as higher-order priors within the same coherence network. Rather than treating partisan motivated reasoning as a separate mechanism, ideological beliefs are modeled as structured constraints that enter the same bidirectional updating process that governs coherence-driven information integration. Partisan identity thus operates as a prior whose influence depends on its activation strength and its alignment with incoming evidence, allowing partisan bias to be conceptualized as a parametric variation within a unified coherence-based process.

Figure 1 illustrates this proposed coherence network. At its core are two candidate nodes representing the competing political options. These nodes are reciprocally connected by inhibitory links, reflecting the mutually exclusive nature of preference formation in choice: increasing activation of one candidate suppresses activation of the alternative.

Figure 1.

A Coherence-Based Network Model of Partisan-Ideological Political Information Integration



Note. Nodes represent candidate preferences, evaluations of information, incoming statements, and ideological–partisan beliefs. Node shading indicates activation state: dark (black) nodes denote activated representations, whereas light (gray) nodes denote inhibited representations. Solid lines indicate excitatory influences, dashed lines inhibitory influences. The dotted rectangle highlights the network state at time t_0, when a new statement about Candidate B is evaluated after prior information has established a preference for Candidate A. The parameter γ denotes the relative influence of informational input versus coherence- and belief-based constraints.

At the lower level, input nodes represent the objective content of sequentially presented political statements. In the present implementation, we focus on positive statements, reflecting the empirical regularity that political candidates predominantly communicate affirming claims about their intentions and policies. Each input node feeds into an evaluation node via a bottom-up connection weighted by w_{info} capturing the diagnostic contribution of the statement's content to its subjective evaluation. Evaluation nodes are bidirectionally linked to candidate preference nodes through coherence connections weighted by $w_{coherence}$.

These links implement the core coherence mechanism: evaluations influence candidate preferences, and emerging preferences feedback to bias subsequent evaluations. As a consequence, information favoring the currently leading candidate is progressively evaluated more positively, whereas information favoring the trailing candidate is attenuated—yielding coherence-driven information distortion.

At the top of the network, ideological–partisan belief nodes represent prior political identity (e.g., "Candidate A aligns with my party"). Following Simon and Read (2023), these belief nodes function as priors that impose top-down constraints on the network. Belief nodes connect to candidate representations via belief links weighted by w_{belief}. Positive activation excites the ideologically consistent candidate and inhibits the opposing candidate; negative activation reverses this pattern through signed constraints. Because candidate nodes mutually inhibit each other, belief activation about one candidate indirectly suppresses the alternative even in the absence of explicit beliefs about that alternative.

The relative influence of bottom-up evidence versus coherence- and belief-based constraints is governed by a scaling parameter γ, which captures the balance between information-weighted processing (w_{info}) and constraint-based processing ($w_{coherence}+ w_{belief}$).

Low γ values correspond to evidence-dominant integration, whereas high γ values reflect increasing dominance of coherence maintenance and ideological priors.

Crucially, ideological–partisan belief activation is not assumed to be constant. Instead, the model treats w_{belief} as an experimentally manipulable parameter. By varying the availability, clarity, and diagnosticity of partisan cues, the experimental context systematically modulates the strength with which ideological priors constrain coherence-based updating. This formalization directly follows Simon and Read's (2025) proposal that motivated reasoning arises from variations in prior strength within a coherence network rather than from a separate reasoning process. At the model level, systematic variations in the relative strength and alignment of informational (w_{info}), coherence-based ($w_{coherence}$), and belief-based (w_{belief}) constraints give rise to distinct coherence dynamics, which form the basis for five concrete predictions tested in the present studies.

Model-Derived Predictions

To clarify the structure of the five model-derived predictions, we illustrate the network shown in Figure 1 using a concrete example. Consider a participant who evaluates two unfamiliar political candidates from two unknown parties, such that no explicit partisan labels are available. The task is to evaluate a series of sequentially presented statements and, based on this information, to assess the candidates' perceived political position in order to form a voting intention. At earlier time points (t_{-2} and t_{-1}), participants encountered two positive but ideologically non-diagnostic statements about Candidate A (e.g., competence, experience). Through the bottom-up information links (w_{info}) these statements increase activation in the corresponding evaluation nodes, which—via coherence links ($w_{coherence}$) gradually increase activation of the preference node for Candidate A. Because candidate preference nodes are bidirectionally linked to ideological belief nodes, this emerging preference is sufficient to induce an inferred ideological belief (e.g., “the candidate I currently prefer is probably closer to my political orientation”), even though no explicit ideological cues have been provided. Thus, bottom-up preference formation can trigger top-down belief activation.

At time t_0, participants encounter a new positive statement about Candidate B. Crucially, the *content* of this statement is ideologically congruent with the participant’s own orientation (i.e., its policy implication matches the participant’s ideological stance), yet it is associated with the currently trailing candidate. In network terms, the statement provides evidence in favor of Candidate B through w_{info} (bottom-up evidence input from the statement node to its evaluation), while the current coherence state favors Candidate A via $w_{coherence}$ and—once activated—also exerts top-down constraint via w_{belief}. The parameter γ captures the relative dominance of evidence-weighted processing (w_{info}) versus constraint-based processing ($w_{coherence} + w_{belief}$) such that higher γ implies stronger coherence-/belief-weighted evaluation under this conflict configuration.

H1: Top-Down Intrusions on Information Distortion

In the network state illustrated in Figure 1 (t_0), participants encounter a new positive statement about Candidate B while Candidate A already holds a preference advantage. Because the candidate nodes mutually inhibit each other, the higher activation of the currently preferred candidate reduces activation of the competing candidate. Via the coherence links ($w_{coherence}$) connecting candidate and evaluation nodes, this activation imbalance biases the evaluation of the B-associated statement downward relative to the actual diagnosticity of the statement (i.e., w_{info}).

This yields a basic leader–trailer asymmetry in information evaluation (H1a). Statements about the leading candidate are evaluated more positively than statements about the trailing candidate. Importantly, this can occur even in the absence of explicit ideological beliefs or cues.

In the present example, this asymmetry is further amplified because the emerging preference for Candidate A induces an inferred ideological belief (e.g., "the candidate I currently prefer is probably closer to my orientation"). When such partisan beliefs are activated—either inferentially (as in this example) or via explicit cues (e.g., telling participants that Candidate A belongs to their preferred party)—they strengthen belief-based constraints in the network (i.e., higher effective w_{belief}) and thereby increase the leader–trailer asymmetry (H1b). Moreover, this amplification should be strongest among participants with stronger partisan commitment, because they place greater weight on belief-consistent constraints and thus show an even stronger asymmetry (H1c).

H2: Preference-Induced Belief-Based Categorization

Beyond evaluative distortion, the model predicts systematic bias in the perceived ideological alignment of information. In the network state illustrated in Figure 1 (t_0), the emerging preference for Candidate A is sufficient to activate an inferred ideological belief linking the preferred candidate to the participant's own political orientation. Consistent with the bidirectional architecture of coherence-based models, this belief activation both follows from the emerging preference and feeds back into subsequent interpretation of subsequent information.

The consequence of this is that a statement about Candidate B—despite being ideologically congruent in its content—is perceived as less aligned with the participant's political orientation because it is associated with the trailing candidate. Conversely, information associated with the currently preferred candidate is perceived as more ideologically aligned, even when ideological cues are weak or absent. Thus, perceived ideological alignment follows candidate preference rather than informational content alone (H2a).

More generally, the model predicts that ideological categorization of information is increasingly shaped by belief-based constraints within the coherence network. As partisan beliefs become more strongly activated—either inferentially through emerging preferences or through explicit partisan cues—candidate association exerts greater influence on perceived ideological alignment. Accordingly, stronger belief activation (higher effective w_{belief}) yields more pronounced misalignment of trailing-candidate information and stronger perceived alignment of leader-associated information, with the effect expected to scale with individual partisan commitment (H2b).

H3: Attenuated Distortion Under Partial Constraint Conflict

In the network state at t_0, the ideological content of the newly encountered statement supports the participant's political orientation, which—via the belief node—induces a positive belief-based constraint (w_{belief}) in favor of the statement's content. At the same time, the statement is associated with the currently trailing candidate, whose preference node is suppressed by coherence-driven competition with the leading option. As a result, belief-based constraints and coherence-based constraints exert opposing influences on the evaluation of the same piece of information.

In the present example, this partial constraint conflict attenuates evaluative distortion: although coherence-driven feedback from the candidate layer biases evaluation of trailer-associated information downward, concurrent activation of a belief-consistent ideological interpretation counteracts this bias. The evaluation of the B-associated statement is therefore expected to be less negatively distorted than it would be if ideological content and candidate association both favored the same option.

More generally, the model predicts that the magnitude of information distortion depends on the alignment of belief-based and coherence-based constraints. When ideological content and candidate association converge, belief activation reinforces coherence-driven preference dynamics, yielding stronger distortion. When they diverge, belief-based constraints partially offset coherence-based suppression, resulting in reduced—but not eliminated—distortion. This interaction between source-based belief activation and preference-based coherence dynamics constitutes a central test of the model's claim that ideological bias operates through graded constraint satisfaction rather than through uniform belief-driven discounting.

H4: Identity–Evidence Conflict and Motivated Blindness

As discussed above, the network is characterized by a conflict between belief activation and statement-based information. At t_0, the participant in our example has developed a preference for Candidate A, while the newly encountered statement provides ideologically congruent evidence in favor of Candidate B. Supporting Candidate B at this point would introduce inconsistency with already active preference and belief nodes. This is where coherence-based constraints come into play, stabilizing the participant's existing preference for Candidate A. Consequently, the participant's choice is guided more strongly by the internally coherent preference state than by the informational implications of the newly presented evidence, which should result in selecting Candidate A.

More generally, the model predicts that when belief- and coherence-based constraints dominate information integration, individuals may select an option that is less well aligned with ideological content. In other words, internally coherent preferences can override the informational implications of new evidence, giving rise to what we term motivated blindness.

H5: Dynamic Updating Under Differential Partisan Constraint

In the example configuration at t_0, the ideologically congruent statement about Candidate B increases activation of the trailing candidate via bottom-up information input, thereby reducing the preference imbalance between the two candidates. As a consequence, the negative distortion (i.e., tendency to undervalue information about the trailing candidate) of B-associated information is attenuated relative to situations in which new information would further strengthen the existing preference for Candidate A.

More generally, the model predicts that evaluative distortion dynamically tracks trial-by-trial changes in preference strength: information that shifts preferences toward the trailing option reduces preference-consistent information distortion, whereas information that strengthens the leading option increases preference-consistent information distortion (H5a). Crucially, this dynamic sensitivity distinguishes the model from purely prior-driven accounts of motivated reasoning, which would predict stable distortion determined by ideological priors alone.

At the same time, the influence of trial-level updating is moderated by partisan constraint strength. When ideological beliefs are weak or merely inferred, preference updating exerts a strong influence on subsequent evaluations. When ideological beliefs are more strongly activated through salient partisan cues, belief-based constraints stabilize the current preference state, dampening the impact of trial-level updating on distortion (H5b).

Overview of the Present Research

Across two experiments, we test these five central predictions derived from a coherence-based model of political reasoning. Both studies employ a sequential decision-making paradigm in which preferences evolve stepwise as new information is integrated, allowing us to examine information distortion as a dynamic process rather than a static evaluative bias. Across studies, we systematically vary the strength and alignment of ideological–partisan belief activation, thereby implementing graded changes in belief-based constraints within a single coherence-driven network. This approach allows us to test how top-down ideological priors interact with bottom-up evidence during preference formation.

Study 1 provides an initial test of the model's core mechanisms under conditions of increasing partisan activation, while ideological content is not consistently linked to a specific candidate. By comparing conditions with minimal, inferred, and explicit partisan cues, Study 1 examines how strengthening belief-based constraints amplifies coherence-driven information distortion as preferences emerge over time. Specifically, Study 1 test the model's core predictions concerning evaluative distortion, belief-based categorization, and their moderation by constraint alignment (H1–H3). Study 2 extends this approach by holding ideological content constant at the candidate level while varying whether partisan cues converge or conflict with informational evidence. This design imposes systematically different constraint structures on the coherence network and thus enables a direct test of motivated blindness, as well as a distinction between static partisan influence and dynamic coherence-based updating at the trial level. Study 2 provides a comprehensive test of all five model-derived predictions, including motivated blindness and dynamic trial-level updating (H1–H5).

Transparency and Open Science

The studies are reported in accordance with APA Journal Article Reporting Standards (JARS). All primary hypotheses and analysis plans were preregistered prior to data collection. Preregistrations, anonymized data, analysis code, materials, and power analysis scripts are publicly available on the Open Science Framework (OSF). An anonymous view-only link is provided for review purposes (https://osf.io/hdf9a/overview?view_only=e76b08bd21af43e39252048c898b04f4). The repository is organized into functional components (Data and Analysis, Materials, Software, and Figures), each subdivided into study-specific folders ("Study 1," "Study 2," "Pre-Study"), enabling full reproducibility of the reported results. For transparency and ease of navigation, direct links to study-specific preregistrations and materials are provided in Appendix A. Participants for all studies were recruited via Prolific. To avoid non-naiveté, participants were barred from taking part in more than one study. The experiments were programmed in jsPsych and hosted via JATOS.

Ethical Considerations

The present studies involve anonymous, minimal-risk behavioral procedures. Participants provided informed consent prior to participation and were fully debriefed at the end of the study. No identifying information was collected, and all data are stored and analyzed in anonymized form. Ethical approval for the research protocol was obtained through an external ethical review procedure conducted by the German Association for Experimental Economic Research (see https://gfew.de/ethik/MJXiv4uP). All procedures comply with the ethical standards of the American Psychological Association for research involving human participants.

Study 1

Study 1 serves as an initial test of the proposed coherence-based model under conditions of increasing party identity cue accessibility. The study examines how top-down ideological activation interacts with ongoing coherence-driven updating when political preferences evolve through sequential information integration. Specifically, it tests whether increasing accessibility of party identity cues amplifies predecisional information distortion as preferences evolve sequentially, and whether emerging preferences themselves induce belief-based categorization in the absence of explicit party labels. To test these predictions, Study 1 systematically manipulates the accessibility of partisan beliefs while participants evaluated two political candidates described by a sequence of policy statements. This design enables a process-level examination of how top-down ideological activation, bottom-up preference formation, and statement congruence jointly contribute to coherence-based distortion during political decision making (H1-H3).

Method

Participants

A representative sample (according to age, gender, and political affiliation) total of N = 1215 U.S. adult Prolific users took part in the study. In line with preregistered exclusion criteria, participants who indicated that their data should not be used (n = 2) or failed an attention check (n = 53) were excluded, resulting in a final sample of N = 1160 participants (age 45.7y; 577 female, 557 male; remaining participants reported another or no gender category; 71.0% White, 14.1% African American, 5.6% Asian, 9.2% other/mixed) who were evenly distributed across the three between-subject conditions (same party: n = 386; different parties: n = 387; explicit party knowledge: n = 390). Two attention checks were embedded in the task; participants who failed an attention check were excluded from the analyses. Participants received £1.25 as compensation for their time (approximately 8 minutes), independent of their exclusion or inclusion in the final sample.

Sample size was determined a priori to ensure adequate power to detect small interaction effects in trial-level mixed-effects models, particularly interactions between partisan-belief condition and information type (leader vs. trailer). Because no precise prior estimates of these interaction effects were available, we adopted a conservative sampling strategy targeting approximately .80 power for small effects. We therefore aimed to recruit about 400 participants per between-subject condition (total N = 1,200), allowing for stable estimations, even after anticipated exclusions.

Design

The experiment employed a mixed design with one between-subjects factor and two within-subjects factors. Specifically, the design comprised a 3 (party-identity-belief condition: no partisan belief vs. inferred partisan belief vs. explicit partisan belief; between-subjects) × 2 (ideological congruence of the statement: congruent vs. incongruent; within-subjects) × 2 (information type: leader vs. trailer; within-subjects) structure. Predecisional information distortion was assessed using the stepwise evolution of preference paradigm (Russo et al., 1996, 1998; for an overview see Meloy & Russo, 2004). Participants were randomly assigned to one of the three party-identity-belief conditions. The order of political statements, their assignment to candidates, and all visual labels were fully randomized for each participant.

Materials

All political statements were sourced from *The Living Room Candidate* archive (http://www.livingroomcandidate.org/). A total of 40 statements from 21st-century U.S. presidential campaigns were pretested in an independent validation study to obtain unbiased estimates of their persuasive strength and ideological alignment. The pool included 20 statements originally made by Democratic and 20 by Republican candidates. To ensure comparability and avoid identity cues, statements were selected for generic content and lightly edited so that each could stand alone.

In the pre-study, a separate U.S. sample (N = 325; M_{age} = 34.83, SD = 10.55; 50.2 % women, 70 % White; politically balanced) rated 20 randomly presented statements on one of two dimensions: (a) voting influence—the extent to which the statement would affect one's decision to vote for (+100) or against (-100) the speaker's party, and (b) political alignment—the perceived ideological orientation of the statement (–100 = strongly Democratic to +100 = strongly Republican). Each participant rated 20 statements on a single dimension only.

These data were used to match statements across parties based on mean voting-influence ratings, resulting in six Democratic–Republican pairs of comparable argumentative strength spanning a wide diagnosticity range (means ≈ 2–42). Three pairs were politically aligned with their source party, and three were misaligned, providing systematic variation in ideological congruence. For the main study, one statement from each pair was randomly selected for each candidate, yielding six statements per candidate (12 total per participant). The final stimulus set represented diverse candidates and election years, ensuring both comparability and generalizability of the presented information.

Procedure

After providing informed consent, participants were informed that they would complete a political decision-making task involving two candidates and a sequence of political statements presented one at a time. Participants were randomly assigned to one of three between-subject party-identity-belief conditions. Depending on condition, the candidates shown were presented to belong to 1) the same party (i.e., Political Party 1), 2) different parties, without specifying which candidate belonged to which party (i.e., Political Party 1 / Political Party 2) or 3) belonged to different parties with explicit party labels (Democrat vs. Republican). Candidate Labels (i.e., Candidate XY / Candidate YZ) and visual identifiers (i.e., dark- vs. light-blue icons) were counterbalanced.

Participants first completed a brief practice trial to familiarize themselves with the task and response format. The main task then proceeded trial by trial. In each trial, participants read one political statement referring to one of the two candidates and evaluated that statement. After each evaluation, participants indicated their current preference between the two candidates. Based on these preference reports, the statement presented in the subsequent trial was classified as referring to either the currently preferred (leader) or non-preferred (trailer) candidate.

Measures

In each trial, participants evaluated the presented political statement by indicating the extent to which it influenced their overall attitude toward the candidate on a scale from −100 (strongly negative) to +100 (strongly positive). After each statement evaluation, participants reported their current preference between the two candidates on a separate scale ranging from −100 (strong preference for Candidate A) to +100 (strong preference for Candidate B). These preference ratings were used to dynamically classify each statement as pertaining to the currently preferred (leader) or non-preferred (trailer) candidate.

After completing all trials, participants rated the perceived ideological alignment of each candidate on a scale from −100 (strongly Democratic) to +100 (strongly Republican), based on the statements they had seen. Participants then indicated their final choice between the candidates and reported their confidence in this decision on a continuous scale from 50 (absolutely toss-up) to 100 (absolutely confident).

Evaluative distortion. Evaluative distortion was computed as the deviation between a participant's evaluation of each statement and its baseline evaluation obtained from the independent pre-study sample. Positive values indicate more favorable evaluations than warranted by the statement's objective diagnosticity, whereas negative values indicate downward distortion.

Alignment distortion. Alignment distortion captured bias in perceived ideological categorization. It was computed as the difference between a participant's perceived ideological position of each candidate (−100 = strongly Democratic to +100 = strongly Republican) and the objective ideological benchmark derived from the control group's mean alignment ratings of that candidate's statements. Positive values indicate that a candidate was perceived as more Republican than expected based on statement content, and negative values indicate the opposite.

Ideological congruence. Ideological congruence was defined based on the mean ideological alignment ratings of each statement obtained in the independent validation sample and calculated relative to participants' self-reported partisan orientation. Crucially, congruence was defined contextually as a function of the participant's current preference state. Ideological content was coded as congruent when the ideological content of a statement aligned with the participant's political orientation and was associated with the currently preferred candidate, or when ideologically opposing content was associated with the currently non-preferred candidate. Conversely, ideological content was coded as incongruent when ideological alignment conflicted with current candidate preference (i.e., ideologically agreeable content associated with the non-preferred candidate, or ideologically disagreeable content associated with the preferred candidate).

Results

All analyses were conducted using R (R Core Team, 2019). Data preparation and restructuring were conducted using the tidyverse packages dplyr and tidyr (Wickham et al., 2019). Visualizations were generated with the ggplot2 package (Wickham, 2016b), and all mixed model analyses were carried out using the lme4 (Bates et al., 2015a) and lmerTest (Kuznetsova et al., 2017) packages. Effect sizes for individual fixed effects were quantified as semi-partial R^2 using the r2beta function (Jaeger et al., 2017), following recommendations for multilevel models (Rights & Sterba, 2019). For ease of interpretation, we additionally report the square root of these R^2 values as correlation-based effect sizes and interpret their magnitude in line with recent guidelines for individual-differences research (Gignac & Szodorai, 2016). Results are organized according to the three model-derived predictions tested in Study 1: coherence-driven evaluative distortion (H1), preference-driven belief activation and ideological categorization (H2), and the moderation of distortion under partial constraint conflict (H3).

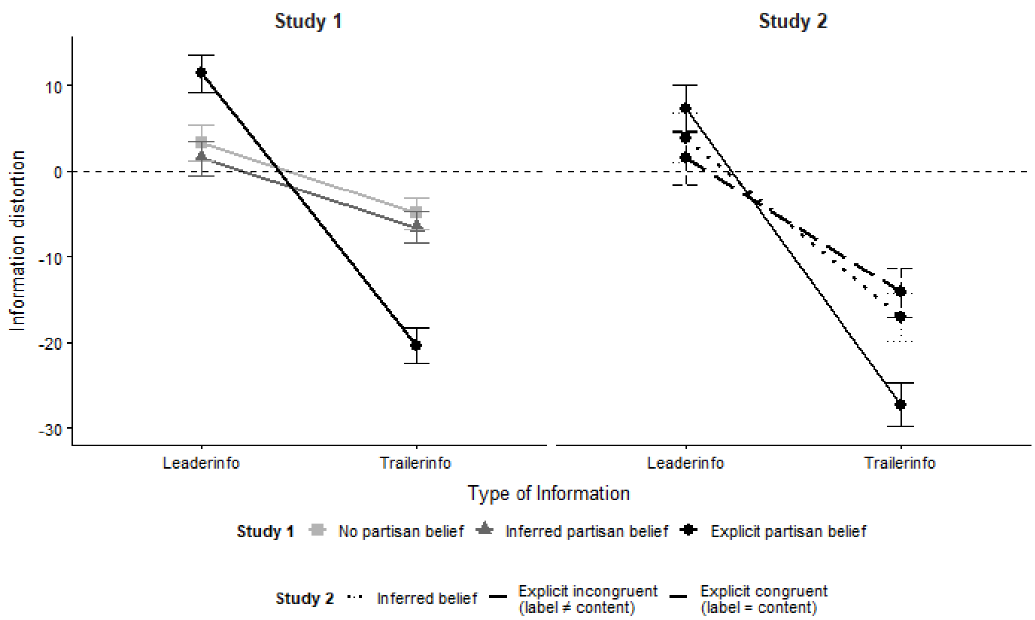
H1: Top-Down Intrusions on Information Distortion

We first examined whether the coherence-driven information distortion predicted by the model emerged in the present political decision-making paradigm. A linear mixed-effects model with a random intercept for participants and effect-coded information type (leader = +0.5, trailer = −0.5) revealed a robust leader–trailer asymmetry: statements associated with the currently preferred (leading) candidate were evaluated more positively than statements associated with the non-preferred (trailing) candidate, $b = 17.19$, $SE = 0.77$, $t(10832) = 22.29$, $p < .001$. This effect accounted for $R^2 = .04$, 95% CI [.03, .05] ($r \approx .21$), indicating a small-to-moderate distortion effect that emerged reliably across conditions. Because statements were randomly assigned to candidates, this asymmetry cannot be attributed to systematic differences in statement content but instead reflects preference-consistent distortion in information evaluation (H1a).

We next examined whether this leader–trailer asymmetry was amplified by the activation of party-identity beliefs. Party-identity belief was modeled as a linear trend across the three experimental conditions (no partisan belief = −1, inferred partisan belief = 0, explicit partisan belief = +1), with information type again effect-coded. As predicted, we observed a significant interaction between information type and partisan-belief condition, $b = 11.87$, $SE = 0.93$, $t(10805) = 12.74$, $p < .001$. This interaction accounted for a semi-partial $R^2 = .01$, 95% CI [.01, .02] ($r \approx .12$). As shown in Figure 2 (left panel), the evaluative gap between leader- and trailer-associated information increased systematically as party-identity cues shifted from absent to inferred to explicit, consistent with the model's assumption that belief-based constraints strengthen coherence-driven biases in information evaluation. Specifically, participants overvalued leader-associated information and undervalued trailer-associated information. This evaluative gap was largest in the explicit partisan-belief condition and smaller in the inferred and no-partisan conditions, consistent with stronger belief-based constraints amplifying coherence-driven biases (H1b).

Figure 2.

Information Distortion by Leader vs. Trailer Statements and Party-Identity Belief Manipulation across two studies.



Note. The figure displays mean information distortion for statements associated with the currently leading versus trailing candidate across partisan-belief conditions. Leader statements refer to statements associated with the candidate preferred at the end of the preceding trial, whereas trailer statements refer to statements associated with the currently nonpreferred candidate. In Study 1, partisan belief activation varied from absent to inferred to explicit: In the *no partisan belief* condition, both candidates belonged to the same party; in the *inferred* condition, candidates belonged to different parties, but party affiliation had to be inferred from ideological content; in the *explicit* condition, party labels were provided. In Study 2, ideological content was held constant within candidates while partisan labels were either *congruent* with the ideological implications of the statements, *incongruent*, or *inferred* (no explicit labels). Error bars represent ±1 SE. The dashed horizontal line indicates zero distortion.

Finally, we tested whether the amplification of distortion by partisan cues depended on the strength of participant orientation. The original party preference scale ranged from 1 = "Strongly prefer Democrats" to 7 = "Strongly prefer Republicans." Partisan identification strength was quantified as the absolute distance from the scale's neutral midpoint (4 = "Neutral") and recoded to range from 0 (no partisan preference) to 3 (strongest partisan preference). A three-way interaction between information type, partisan-belief condition, and partisan strength was significant, $b = 7.02$, $SE = 0.86$, $t(10760) = 8.19$, $p < .001$, semi-partial $R^2 = .005$, 95% CI [.003, .008] ($r \approx .08$). This pattern indicates that partisan cues amplified evaluative distortion more strongly among participants with higher partisan identification, whereas participants with weaker partisan commitments showed comparatively attenuated sensitivity to partisan cue activation (H1c). Although small in magnitude, this moderation effect is theoretically consistent with the model's claim that belief-based constraints exert graded influence on information integration as a function of prior belief strength.

H2: Preference-Induced Belief-Based Categorization

We next examined whether emerging preferences systematically biased how participants categorized the ideological alignment of information, as predicted by the model's assumption that bottom-up preference formation can activate top-down belief-based categorization. In other words, the model predicts that once a candidate is preferred, information associated with that candidate is more likely to be perceived as ideologically closer to one's own political orientation, even when ideological cues are weak or mixed.

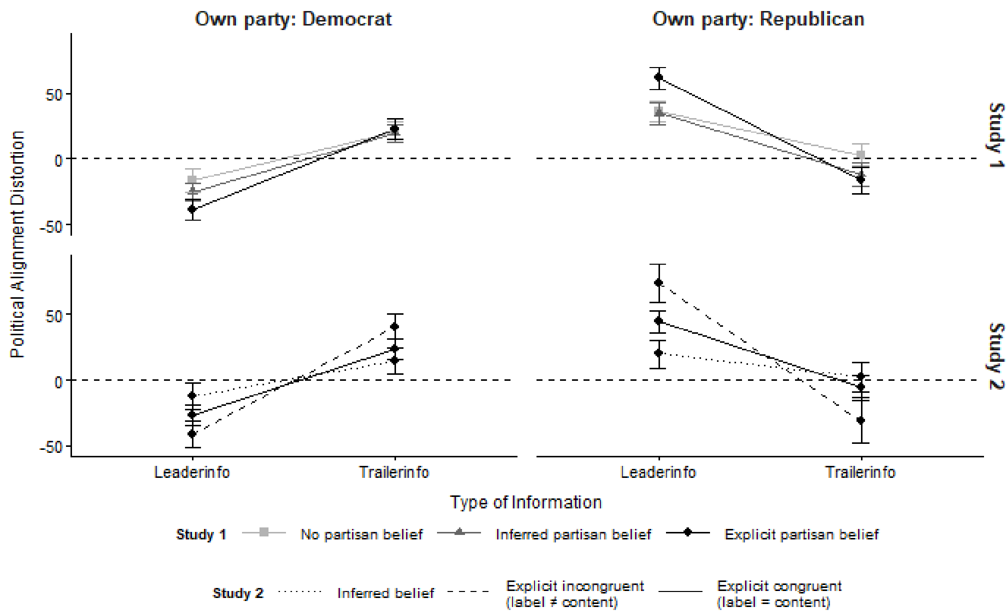
To test this prediction, we analyzed alignment distortion as a function of information type and participants' partisan orientation. Information type was effect-coded (leader = +0.5, trailer = −0.5). Partisan orientation was derived from a 7-point party preference scale (1 = *strongly prefer*

Democrats, 7 = *strongly prefer Republicans*) and effect-coded for analysis (+0.5 for Republican-leaning participants, −0.5 for Democratic-leaning participants). Participants who reported a neutral party preference (scale midpoint) were excluded from this analysis. This analysis revealed a strong interaction between information type and partisan orientation, $b = 103.54$, $SE = 4.93$, $t = 21.00$, $p < .001$. Leader-associated candidates were perceived as more closely aligned with participants' own political orientation, whereas trailer-associated candidates were perceived as more ideologically opposed. This interaction accounted for substantial unique variance (semi-partial $R^2 = .20$, 95% CI [.17, .23]; $r \approx .43$), indicating a pronounced belief-based categorization effect whereby ideological meaning followed candidate preference rather than informational content alone (H2a).

We next examined whether this preference-induced categorization was further shaped by the activation of party-identity belief, as predicted by the model's assumption that belief-based constraints increase in strength when partisan knowledge becomes more salient. To this end, we extended the model by adding the experimental party-identity belief contrast as a continuous moderator (−1 = no partisan belief, 0 = inferred partisan belief, +1 = explicit partisan knowledge), treating the manipulation as an ordered increase in belief salience. The extended model included all two- and three-way interactions among information type, partisan orientation, and party-identity belief. Consistent with this, the predicted three-way interaction was significant, $b = 35.71$, $SE = 6.04$, $t = 5.91$ $p < .001$, and accounted for a small but meaningful amount of additional variance (semi-partial $R^2 = .02$, 95% CI [.01, .03]; $r \approx .13$). Ideological categorization was most pronounced when party-identity belief was made explicit, indicating that stronger belief activation increases the extent to which ideological judgments are guided by candidate preference rather than by statement content (H2b). See Figure 3 (top panels).

Figure 3.

Political Alignment Distortion by Leader vs. Trailer Statements and Party-Identity Belief Manipulation across two studies.



Note. The figure displays mean political alignment distortion for statements associated with the currently leading versus trailing candidate, separated by participants' own party identification (Democrat vs. Republican) and study. Leader statements refer to statements associated with the candidate preferred at the end of the preceding trial, whereas trailer statements refer to statements associated with the currently nonpreferred candidate. Political alignment distortion reflects deviations in perceived candidate orientation from the control baseline; positive values indicate shifts toward perceiving a candidate as more Republican-aligned and negative values toward perceiving a candidate as more Democratic-aligned. In Study 1 (top panels), partisan belief activation varied from absent to inferred to explicit. In Study 2 (bottom panels), ideological content was held constant while partisan labels were inferred, explicitly incongruent with the ideological implications of the statements, or explicitly congruent with them. Error bars represent ±1 SE.

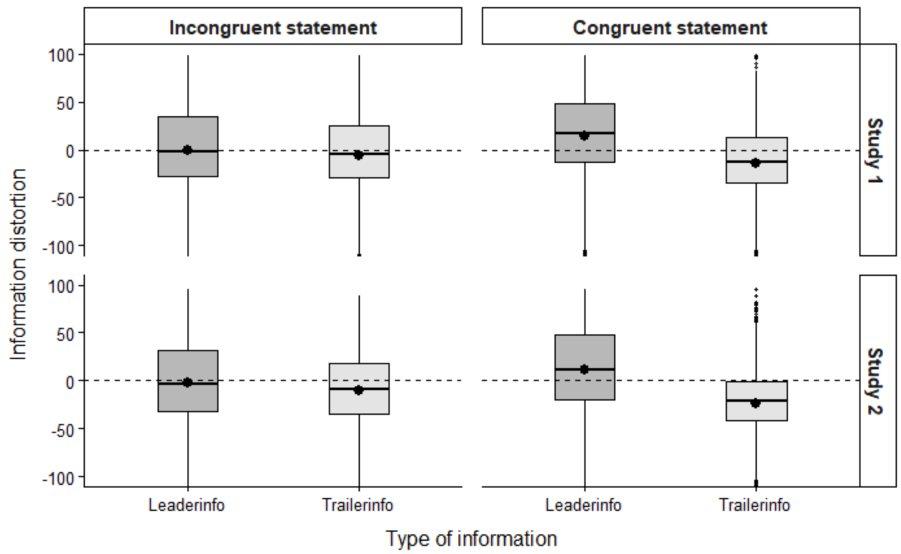
H3: Attenuated Distortion Under Partial Constraint Conflict

We next examined whether the magnitude of information distortion depended on the alignment between ideological content and current candidate preference. In practical terms, distortion should be amplified when ideological content and candidate preference point in the same direction (i.e., Ideological content was coded as congruent when a statement's ideological content matched the participant's political orientation and was linked to the currently preferred candidate) and reduced when ideological content and candidate preference diverge.

To test this prediction, we examined whether ideological congruence moderated the leader–trailer distortion effect. A linear mixed-effects model with a random intercept for participants and effect-coded information type (leader = +0.5, trailer = −0.5) and congruence (congruent = +0.5, incongruent = −0.5) revealed a significant interaction between information type and congruence, $b = 22.66$, $SE = 2.84$, $t(1128) = 7.98$, $p < .001$. Ideologically congruent information was evaluated more positively when associated with the leading candidate and more negatively when associated with the trailing candidate, relative to ideologically incongruent information. This interaction accounted for unique variance in distortion (semi-partial $R^2 = .02$, 95% CI [.01, .02]; $r \approx .13$), indicating that convergence between ideological content and candidate preference amplified evaluative distortion, whereas divergence between these signals attenuated it (H3). Figure 4 (upper panels) provides an illustration of this pattern for Study 1.

Figure 4.

Information Distortion by Leader vs. Trailer Statements and Ideological Congruence across two studies.



Note. Boxplots display the distribution of information distortion for statements associated with the currently leading versus trailing candidate, separately for ideologically incongruent and congruent content and by study. Leader statements refer to statements associated with the candidate preferred at the end of the preceding trial, whereas trailer statements refer to statements associated with the currently nonpreferred candidate. Ideological congruence was defined relative to participants' own political orientation and current preference state: statements were coded as congruent when ideologically aligned content was associated with the currently preferred candidate (or opposing content with the nonpreferred candidate) and as incongruent when ideological content conflicted with current preference. Information distortion reflects the deviation between a participant's evaluation of a statement and the corresponding baseline evaluation obtained from the independent control sample. The dashed horizontal line at 0 indicates no distortion. Boxes represent medians and interquartile ranges, whiskers extend to 1.5 interquartile ranges, and points denote means.

To assess the robustness of this effect, we conducted a complementary exploratory analysis using an alternative operationalization of ideological congruence based solely on the objective partisan source of each statement rather than on the subjective alignment ratings provided by the pretest sample. Using this definition, congruence was again calculated contextually as a function of information type. Replicating the primary analysis, a linear mixed-effects model with a random intercept for participants again showed a significant interaction between information type and congruence, $b = 7.97$, $SE = 1.72$, $t(8776) = 4.64$, $p < .001$. Although smaller in magnitude (semi-partial $R^2 = .002$, 95% CI [.001, .004]; $r \approx .03$), this effect confirms that attenuation of distortion under partial constraint conflict generalizes across different operationalizations of ideological content.

Together, these findings support the model's prediction that ideological bias in information evaluation reflects graded constraint satisfaction rather than uniform belief-driven discounting. Evaluative distortion varies systematically with whether ideological signals reinforce or counteract current preference states.

Discussion

Study 1 provides converging support for the central predictions of the coherence-based account. Information distortion reliably favored the currently leading candidate over the trailing candidate, replicating the information-distortion effect in a political decision context (H1a). Importantly, this distortion increased systematically with experimentally induced activation of party-identity beliefs (H1b) and was further moderated by individual partisan identification strength (H1c), indicating that ideological beliefs act as graded top-down constraints on evaluative processing. These findings indicate that preferences emerge sequentially and can independently bias evaluations. Moreover, this bias is progressively amplified as belief-based constraints increase in strength and salience.

Beyond evaluative distortion, we found clear evidence for preference-induced belief-based categorization. Information associated with the currently preferred candidate was perceived as more closely aligned with participants' own political orientation, whereas information associated with the trailing candidate was perceived as more opposed (H2a). This categorization effect was further strengthened with increasing activation of party-identity beliefs (H2b), demonstrating that the interpretation of ideological content is dynamically shaped by emerging preferences rather than determined solely by the objective informational content.

Finally, information distortion was systematically moderated by the alignment between ideological content and candidate preference (H3). Distortion was strongest when the ideological content of statements aligned with participants' current preferences and attenuated when it conflicted with them, providing direct evidence that ideological bias reflects graded constraint satisfaction rather than uniform belief-driven discounting.

Together, these findings indicate that political bias in information processing arises from reciprocal interactions between coherence-driven preference updating and ideological belief activation. Rather than acting as static priors that uniformly filter information, ideological beliefs can be understood as top-down priors that dynamically interact with bottom-up evolving preference states to shape both evaluative distortion and ideological categorization over the course of sequential decision making.

Study 2

Study 2 extends Study 1 by testing all five model-derived predictions while holding ideological content constant and systematically varying the relation between partisan labels and informational evidence. Whereas Study 1 presented candidates with ideologically mixed statements, Study 2 assigns each candidate ideologically consistent information and independently varies whether partisan labels align with or contradict this content. As a result, belief-based and evidence-based constraints either converge or conflict. This design allows us to examine how partisan labels shape information evaluation and choice when they reinforce versus oppose the informational evidence. In addition, Study 2 assesses whether trial-level preference updating continues to predict information distortion under different degrees of partisan constraint, thereby testing whether coherence-based updating remains operative when partisan cues are salient. Apart from this manipulation, Study 2 closely follows the design and procedure of Study 1.

Method

Participants

A representative sample (age, gender, and political affiliation) of N = 623 adult U.S. Prolific participants took part in the study, in return for £1.50 (for approximately 10 minutes). In line with preregistered exclusion criteria, participants who indicated that their data should not be used (n = 1) or failed the attention check (n = 32) were excluded, resulting in a final sample of N = 590 participants (M age = 46.4 years, SD = 15.60; 291 female, 284 male; 74.7% White, 9.0% African American, 5.9% Asian; 5.9% Hispanic or Latino; 4.5% other/mixed) distributed across the three between-subject conditions (explicit party congruent: n = 204; explicit party incongruent: n = 190; inferred party: n = 197).

Sample size was determined using a simulation-based parametric bootstrap power analysis with the mixed-effects models from Study 1 as the data-generating framework. Power was evaluated for the two focal effects tested again in Study 2: the baseline information-distortion effect (leader vs. trailer) and the interaction between information type and trial-level change in preference (ΔPreference). Based on these simulations, we conservatively targeted n = 200 participants per condition (N = 600 total). Simulation code and data are available on OSF.

Design

The experiment again employed a mixed design with one between-subjects factor and two within-subjects factors. Specifically, the design comprised a 3 (party-identity-belief condition: inferred vs. explicit congruent vs. explicit incongruent; between-subjects) × 2 (ideological congruence of the statement; within-subjects) × 2 (information type: leader vs. trailer; within-subjects) structure. Participants were randomly assigned to one of the three party-identity-belief conditions. Study 2 used the same sequential decision paradigm, trial structure, and response measures as Study 1 but introduced a manipulation of the relation between partisan labels and ideological content. In the inferred party condition, participants were informed that the two candidates belonged to different political parties, without specifying which candidate belonged to which party, as in Study 1. In the explicit conditions, candidates were labeled as either Democratic or Republican. In the explicit congruent condition, party labels were consistent with the ideological content of the candidates' statements, whereas in the explicit incongruent condition, party labels contradicted the ideological content. The order of political statements, their assignment to candidates, and all visual labels were fully randomized for each participant.

Materials and Procedure

To enable this manipulation, Study 2 employed a newly assembled set of political statements drawn from the same validated stimulus pool used in Study 1. As in Study 1, statements were paired such that each pair consisted of one Democratic-source and one Republican-source statement matched on diagnosticity (i.e., mean voting-influence ratings), ensuring comparable persuasive strength across parties. Across the six pairs, diagnosticity spanned a broad range (means ≈ 5–43), and statements were drawn from multiple election years and political figures (see open materials).

In contrast to Study 1, statements were assigned to candidates such that each candidate received only ideologically consistent information. Ideological content was determined jointly by the objective source of each statement and its mean ideological-alignment rating from the pre-study sample; only statements for which both indicators converged (i.e., clearly Democratic or clearly Republican) were eligible for assignment. This assignment ensured that partisan labels could either align with or contradict the ideological content of the information.

Measures

Evaluation, alignment, and preference measures, as well as the computation of evaluative distortion, alignment distortion, and ideological congruence, were identical to Study 1. Study 2 allowed for two additional calculations which are described next.

Orientation-Consistent Choice. Orientation-consistent choice captured whether participants' final decision aligned with their own partisan orientation. Final choice was coded as orientation-consistent (1) if the selected candidate's ideological position matched the participant's self-reported partisan orientation and as orientation-inconsistent (0) otherwise. Participants with a neutral partisan orientation were excluded from these analyses (n = 103).

Trial-Level Preference Change (ΔPreference). Trial-level preference change captured dynamic updating of preferences across trials. It was computed as the signed difference between preference strength on the current and the immediately preceding trial (ΔPreference = $\text{preference}_t - \text{preference}_{t-1}$). Positive values indicate increasing preference for the currently leading candidate, whereas negative values indicate shifts in preference away from the leader and toward the trailing candidate. ΔPreference was mean-centered prior to analysis.

Results

Results are again organized according to the five model-derived predictions outlined in the Introduction (H1–H5); all primary analyses testing these predictions were again preregistered prior to data collection, and any deviations are explicitly noted.

H1: Top-Down Intrusions on Information Distortion

To test H1 in Study 2, we examined whether the coherence-driven leader–trailer asymmetry in information evaluation persists when ideological content is held constant within candidates, and whether this asymmetry is systematically amplified by party-identity belief and partisan orientation strength.

A linear mixed-effects model with a random intercept for participants and effect-coded information type (leader = +0.5, trailer = −0.5) revealed a robust leader–trailer asymmetry: information associated with the currently leading candidate was evaluated more positively than information associated with the trailing candidate, $b = 25.32$, $SE = 1.06$, $t(5427) = 23.86$, $p < .001$. This effect accounted for a substantial proportion of variance (semi-partial $R^2 = .08$, 95% CI [.07, .10], $r \approx .30$), indicating that preference-consistent distortion emerged reliably under ideologically consistent information (H1a).

To test H1b, we next examined whether this asymmetry was amplified by the activation of party-identity belief. Party-identity belief was modeled as a preregistered linear contrast capturing increasing effective strength of belief-based constraints on evaluation, ranging from no explicit partisan prior (inferred = −1), to a conflicting partisan prior (explicit–incongruent = 0), to a congruent partisan prior (explicit–congruent = +1). As predicted, the interaction between information type and party-identity belief was significant, $b = 6.95$, $SE = 1.28$, $t(5436) = 5.43$, $p < .001$, indicating a systematic widening of the pro-leader and anti-trailer evaluative gap as belief-based constraints became stronger (semi-partial $R^2 = .005$, 95% CI [.002, .009], $r \approx .07$; see Figure 2, right panel).

Finally, addressing H1c, we tested whether the belief-based amplification of evaluative distortion depended on individual partisan orientation strength. Partisan orientation strength was coded as the absolute distance from the neutral midpoint of a 7-point party preference scale (recoded to range from 0 to 3). The three-way interaction between information type, party-identity belief, and partisan orientation strength was significant, $b = 5.56$, $SE = 1.19$, $t(5418) = 4.68$, $p < .001$, accounting for a small but reliable amount of additional variance (semi-partial $R^2 = .003$, 95% CI [.001, .007], $r \approx .05$). This pattern indicates that the top-down amplification of evaluative distortion by party-identity belief was strongest among participants with more strongly held partisan commitments, consistent with the model's assumption that belief-based constraints exert graded influence on information integration.

H2: Preference-Induced Belief-Based Categorization

To test H2 in Study 2, we examined whether emerging candidate preferences were accompanied by systematic bias in the perceived ideological alignment of information, and whether this preference-induced categorization was amplified by party-identity belief activation.

We first tested whether ideological alignment judgments followed current candidate preference rather than informational content alone. Alignment distortion was computed as in Study 1, capturing deviations between participants' perceived ideological position of each candidate and the objective alignment benchmark derived from the control sample (−100 = Democratic, +100 = Republican). Information type (leader = +0.5, trailer = −0.5) and participants' partisan orientation (Republican = +0.5, Democratic = −0.5) were effect-coded. Consistent with our prediction, this analysis revealed a strong interaction between information type and partisan orientation, $b = 111.92$, $SE = 6.48$, $t(912) = 17.28$, $p < .001$. Leader-associated candidates were perceived as more ideologically aligned with participants' own political orientation, whereas trailer-associated candidates were perceived as more ideologically opposed (H2a). This effect accounted for substantial unique variance (semi-partial $R^2 = .25$, 95% CI [.20, .29], $r \approx .49$), indicating a pronounced preference-driven bias in ideological categorization.

We next examined whether this preference-induced categorization was further shaped by party-identity belief activation. Party-identity belief was modeled using the same preregistered linear contrast indexing increasing effective strength of belief-based constraints as in H1. The predicted three-way interaction between information type, partisan orientation, and party-identity belief was significant, $b = 72.28$, $SE = 7.61$, $t(908) = 9.50$, $p < .001$ and explained a substantial amount of additional variance (semi-partial $R^2 = .09$, 95% CI [.06, .13], $r \approx .27$). This effect indicates that belief-based categorization became increasingly pronounced as party-identity beliefs imposed stronger constraints on interpretation (see Figure 3, bottom panels), with the strongest alignment distortion observed when partisan labels were explicitly congruent with ideological content (H2b).

H3: Attenuated Distortion Under Partial Constraint Conflict

To test H3 in Study 2, we examined whether ideological congruence moderated the leader–trailer distortion effect, as predicted by the model. Ideological congruence was defined relative to participants' own partisan orientation and their current preference state. Statements aligned with a participant's partisan orientation were coded as congruent (+0.5) when associated with the currently preferred (leading) candidate and as incongruent (−0.5) when associated with the non-preferred (trailing) candidate; the reverse applied to ideologically misaligned statements. Information type was effect-coded (leader = +0.5, trailer = −0.5).

A linear mixed-effects model with a random intercept for participants revealed a significant interaction between information type and ideological congruence, $b = 27.45$, $SE = 2.27$, $t(5704) = 12.08$, $p < .001$. As predicted, congruent information was evaluated more positively when associated with the leading candidate and more negatively when associated with the trailing candidate, relative to incongruent information (H3). This interaction accounted for unique variance in evaluative distortion (semi-partial $R^2 = .024$, 95% CI [.017, .033], $r \approx .13$). As illustrated in Figure 4 (bottom panel), the magnitude of distortion varied systematically depending on whether belief-based and preference-based constraints reinforced or opposed one another, even under conditions of clear ideological content.

Together, these results support the model's prediction that ideological bias in information evaluation reflects graded constraint satisfaction rather than uniform belief-driven discounting. Even under conditions of clear ideological content, the extent of distortion varied systematically with whether belief-based and preference-based constraints reinforced or opposed one another.

H4: Identity-Evidence Conflict and Motivated Blindness

H4 was tested for the first time in Study 2 by examining whether choice behavior systematically varies as party-identity beliefs shift from conflicting with to reinforcing ideological content. The model predicts an ordered pattern: when ideological content and party-identity beliefs converge, choices should predominantly align with participants' political orientation; when party affiliation must be inferred from content alone, choices should primarily reflect the informational evidence; and when party-identity beliefs conflict with ideological content, belief- and coherence-based constraints should exert greater influence on choice, resulting in motivated blindness. Accordingly, H4 tests whether the relative influence of ideological content versus party-identity beliefs on choice changes monotonically as belief-based constraints become stronger and more diagnostic.

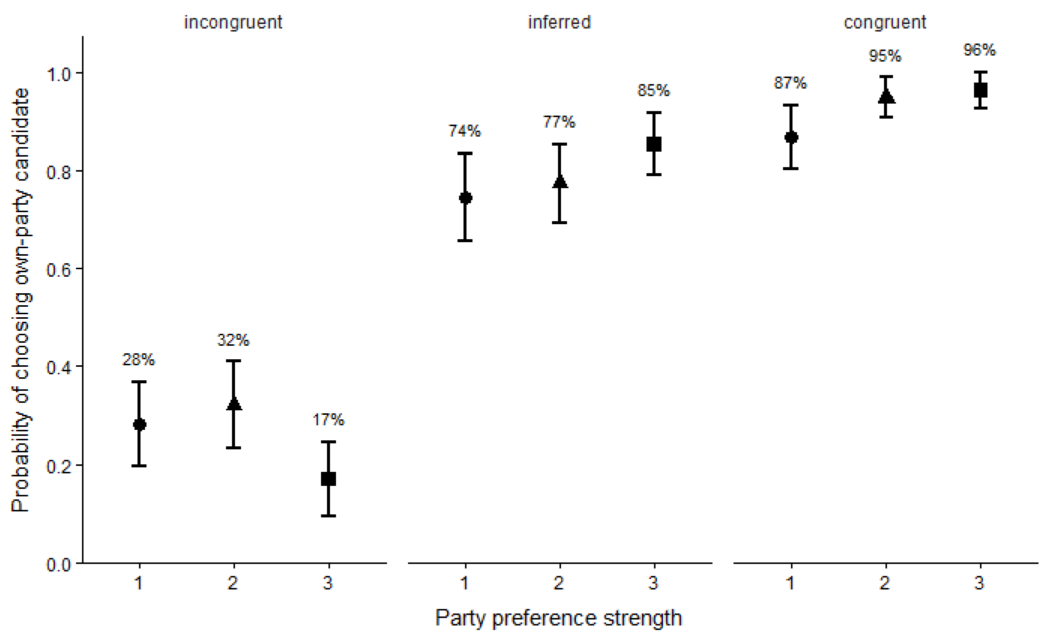
To test this prediction, we estimated a logistic regression predicting orientation-consistent choice (1 = choice aligned with participants' own partisan orientation) from a preregistered linear motivated-blindness contrast capturing increasing alignment between party-identity beliefs and ideological content (explicit–incongruent = −1, inferred = 0, explicit–congruent = +1). Analyses were restricted to participants with a non-neutral partisan orientation ($n = 475$).

The contrast was positive $b = 1.81$, $SE = 0.14$, $z = 13.15$, $p < .001$, indicating a significant increase in the probability of orientation-consistent choice as party-identity beliefs shifted from conflicting with to reinforcing ideological content. The effect accounted for a large proportion of variance in choice behavior (*pseudo-*$R^2 = .17$, 95% CI [.13, .22], $r \approx .44$), consistent with the model's prediction of motivated blindness under identity–evidence conflict.

Descriptively, orientation-consistent choices were least frequent when partisan labels contradicted ideological content (24%), more common when party affiliation had to be inferred from statement content alone (79%), and most frequent when partisan labels and ideological content were explicitly aligned (94%). Figure 5 further shows that this monotonic shift holds across levels of partisan orientation strength. In the inferred condition, participants chose the option aligned with their own orientation even when their partisan orientation strength was weak (middle panel). In contrast, explicit incongruent labels reduced orientation-consistent choices across all levels of partisan orientation strength (left panel). Finally, explicit congruent labels led to very high rates of orientation-consistent choices, reinforcing participants' preferences regardless of their partisan orientation strength (right panel).

Figure 5.

Probability of Orientation-Consistent Choice by Party-Identity Belief Manipulation and Party Preference Strength



Note. The figure displays the probability of choosing the candidate aligned with participants' own partisan orientation across the *incongruent*, *inferred*, and *congruent* conditions. In the *inferred* condition, candidates belonged to different parties but party labels were not revealed. In the *congruent* condition, party labels were consistent with the ideological content of the candidates' statements, whereas in the *incongruent* condition party labels contradicted this content. Partisan orientation strength reflects the absolute distance from the neutral midpoint of the party-identification scale. Symbols represent estimated probabilities with 95% confidence intervals; percentages indicate observed proportions. Participants reporting no partisan orientation were excluded from the analysis.

H5: Dynamic Updating Under Differential Partisan Constraint

H5 examines whether evaluative distortion reflects a purely static influence of party-identity beliefs or whether it remains sensitive to trial-by-trial changes in preference strength, as predicted by the coherence-based model. Crucially, the model implies that evaluative distortion should not only depend on the pre-existing preference state prior to encountering new information but should dynamically track preference updating induced by newly presented information. Specifically, information that shifts preference toward the trailing candidate should attenuate preference-consistent distortion, whereas information that further strengthens the leading candidate should amplify it (H5a). At the same time, the model predicts that this dynamic sensitivity should depend on the strength of party-identity belief activation, such that trial-level updating exerts a weaker influence on evaluation when belief-based constraints are strong (H5b).

To test these predictions, we separated prior preference state from within-trial preference updating. Preference strength at the end of the preceding trial (t–1) was included to capture the existing preference before evaluating the current statement. Dynamic updating was operationalized as trial-level preference change (ΔPreference) was computed as the difference between preference strength after evaluating the current statement and preference strength at the end of the preceding trial (mean-centered), isolating dynamic updating induced by newly presented information[3]. Information type was effect-coded (leader = +0.5, trailer = −0.5), and party-identity belief was modeled as a preregistered linear contrast capturing increasing effective

[3] Because preferences are reported after information evaluation within each round, trial-level preference change cannot be interpreted causally. We therefore treat preference change as an indicator of within-round preference–evaluation dynamics rather than as an exogenous cause of information distortion.

strength of belief-based constraints (inferred = −1, explicit–incongruent = 0, explicit–congruent = +1).

Consistent with H5a, trial-level preference change robustly predicted evaluative distortion beyond prior preference. The interaction between information type and ΔPreference was large and reliable, $b = 0.09$, SE = 0.04, $t(5733) = 24.30$, $p < .001$, accounting for substantial unique variance (semi-partial $R^2 = .09$, 95% CI [.08, .11], $r \approx .32$). Increases in preference strength were associated with more positive distortion for leader-consistent information and more negative distortion for trailer-consistent information, indicating ongoing coherence-based updating at the trial level.

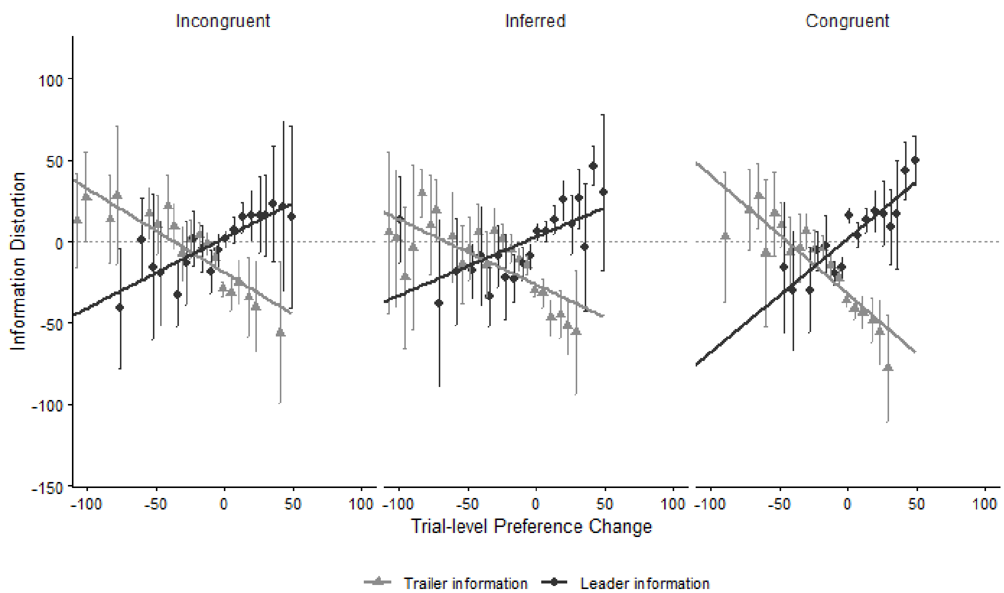
Crucially, this dynamic updating was not attenuated under explicit partisan labeling. Instead, the three-way interaction between information type, ΔPreference, and party-identity belief did not reach significance, $b = 0.07$, SE = 0.05, $t(5740) = 1.61$, $p = .108$, semi-partial $R^2 = .000$, 95% CI [.000, .002], (see Figure 6). Although we expected explicit partisan cues to reduce the influence of trial-level preference updating, the estimated effect was positive and thus opposite in direction to this prediction. Descriptively, the pattern shown in Figure 6 indicates that evaluative distortion continued to vary as a function of trial-level preference changes even under explicit partisan labeling.

Together, these findings are consistent with an additive account in which strong partisan priors elevate baseline distortion while coherence-based updating continues to shape evaluations dynamically.[4]

[4] A parallel robustness analysis using the corresponding model in Study 1 yielded the same qualitative pattern: trial-level preference change significantly interacted with information type, and this interaction was further moderated by partisan-belief condition (all *ps* < .001)

Figure 6.

Information Distortion by Trial-Level Preference Change and Party-Identity Belief Condition



Note. Plotted values show mean information distortion as a function of trial-level preference change between successive trials (ΔPreference = preference$_t$ − preference$_{t-1}$). Positive values on the x-axis indicate shifts in preference toward the currently leading candidate, whereas negative values indicate shifts toward the trailing candidate. *Leader information* refers to statements associated with the option that was preferred at the end of the preceding trial; *trailer information* refers to statements associated with the currently nonpreferred option. Preference change was binned into equally sized intervals across its observed range; points represent mean distortion within each bin. Error bars indicate ±1.96 SE. The dashed horizontal line indicates zero distortion. Panels correspond to the *incongruent*, *inferred*, and *congruent* conditions: in the *inferred* condition, candidates belonged to different parties but party labels were not revealed; in the *congruent* condition, party labels matched the ideological content of the candidates' statements; and in the *incongruent* condition, party labels contradicted this content. Lines show linear fits for descriptive purposes.

General Discussion

Research on biased reasoning has produced two influential accounts of how evaluative distortions arise in political judgment. In the political psychology literature, partisan motivated reasoning is typically conceptualized as a belief-driven process in which prior ideological commitments to one party or ideological position over others exert a predominantly top-down influence on evaluation. In contrast, research on predecisional information distortion has shown that systematic evaluative asymmetries can emerge even in the absence of prior beliefs as emerging preferences bias the evaluation of subsequently encountered information during sequential information integration. The present research integrates these perspectives within a single process framework. Drawing on coherence-based models of judgment and decision making, we proposed that partisan bias reflects the outcome of interacting top-down ideological priors and bottom-up preference construction within a coherence-seeking system. From this perspective, belief-driven and preference-driven distortions are not distinct mechanisms but arise from the same underlying process operating under different constraint structures. The present research advances this theoretical integration by formalizing partisan motivated reasoning as a political instantiation of coherence-driven information integration.

Across two experiments using a sequential information integration paradigm, the findings provide converging evidence consistent with this account. Political bias emerged from coherence-driven dynamics during sequential information integration, with ideological priors entering the coherence network as belief-based constraints that alter the constraint structure under which emerging preferences and incoming evidence shape the evolving evaluation process. This brings a more detailed understanding of the many ways in which biases can occur:

First, evaluative distortion arose as a consequence of emerging preferences during information integration. Once one of the two candidate gained a preference advantage, subsequently encountered statements were interpreted more favorably when associated with that candidate than when attributed to the competing candidate. This asymmetry increased when ideological priors were more strongly activated—both through experimentally manipulated partisan cues and as a function of participants' own partisan identification—indicating that belief-based constraints intrude top-down on coherence-driven preference dynamics.

Second, emerging preferences shaped not only evaluation but also the perceived ideological meaning of political information. Statements associated with the currently preferred candidate were perceived as more ideologically aligned with participants' own orientation, reflecting the bidirectional updating dynamics predicted by coherence-based models. When partisan cues were strongly activated, these belief-based categorizations were further amplified, suggesting that ideological priors both influence and are reinforced by coherence-driven preference formation during information integration.

Third, the findings indicate that partisan bias remains dynamically sensitive to competing constraints within the coherence network. When ideological content favored one candidate while coherence dynamics favored the other, evaluative distortion was attenuated rather than eliminated, consistent with a constraint-satisfaction process in which multiple influences are jointly weighted. At the level of choice, strong ideological priors could redirect candidate selection toward the candidate labeled as belonging to the participant's preferred party, even when the ideological content of the statements favored the opposing candidate. In these cases, belief-based constraints dominated the coherence network, producing a pattern consistent with motivated blindness.

Considered solely at the level of outcomes, such patterns could easily be interpreted as reflecting a static, belief-driven form of partisan motivated reasoning. Process-sensitive measures, however, reveal a different picture. Trial-level changes in preference strength continued to predict the evaluation of subsequently encountered statements even under explicit partisan labeling, indicating that coherence-driven updating remained operative throughout the decision process rather than giving way to a purely belief-driven mechanism.

Taken together, these findings support a unified interpretation of partisan motivated reasoning as coherence-driven processing under ideological priors. Preferences emerge bottom-up from the evaluation of statements and act as internal coherence constraints, while ideological priors intrude top-down by increasing the weight of belief-based constraints. Political bias thus reflects neither purely motivational override nor passive belief application, but the graded dominance of interacting constraints within a single coherence-seeking mechanism.

To contextualize the magnitude of the observed distortion effects, it is useful to consider the standard metric used in the information distortion literature. Following Russo (2015), distortion magnitude can be expressed as the mean deviation of information evaluations from a neutral baseline as a proportion of the maximum possible deviation on the rating scale. Studies using the stepwise evolution-of-preference paradigm typically report distortion magnitudes ranging between approximately 5% and 40% of the maximum possible scale deviation (Russo, 2015; Russo et al., 1996; Russo et al., 2000).

The effects observed in the present studies fall below the lower bound of this range but show a systematic increase across the constraint structures examined here. Under baseline coherence conditions, in which both candidates were described as belonging to the same party and no ideological cues were present, distortion averaged 0.9% of the maximum possible deviation. When ideological positions had to be inferred from the content of the statements, distortion increased to 3.6%. When explicit partisan labels conflicted with ideological content, distortion rose further to 5.0%, and reached 5.8% when ideological cues were both explicit and congruent. This systematic increase across experimentally manipulated constraint structures is consistent with the coherence-based interpretation advanced here: ideological priors do not introduce a qualitatively distinct mechanism of partisan bias but function as higher-order constraints that increase the weight of belief-consistent information within the coherence network. As partisan cues become more explicit and diagnostically aligned in the experimental conditions, these constraints increasingly shape how incoming information is evaluated.

Comparisons of absolute magnitudes across studies should, however, be interpreted cautiously because distortion magnitude partly depends on the measurement scale used to assess evaluations. Many SEP studies summarized by Russo relied on relatively coarse 7- or 9-point rating scales, for which even small discrete shifts can represent a substantial proportion of the maximum possible deviation from the control-group baseline. For example, if the control-group rating lies at the midpoint of a 7-point scale (i.e., 4), a one-point shift corresponds to 33% of the maximum possible deviation from the baseline, whereas on a −100 to +100 scale, a shift of five units corresponds to only 5%. The present studies employed a fine-grained −100 to +100 evaluation scale, allowing smaller incremental changes in evaluations to be expressed. Consequently, lower proportional magnitudes may partly reflect increased measurement precision rather than weaker coherence-driven effects.

Implications

When people's political beliefs are challenged, they often exhibit a tendency to reject information that contradicts their pre-existing partisan or ideological beliefs. Political psychology has identified a variety of mechanisms by which this process takes place. People may select information in a motivated manner, meaning that they only select those sources that they expect will deliver news that fits in their ideological narrative and thus will confirm their beliefs (Iyengar et al., 2008; Iyengar & Hahn, 2009). American conservatives listen to Fox News, while liberals turn instead to MSNBC. Furthermore, even if people stumble on information that challenges their beliefs, they may weigh information differently. When being exposed to information that challenges their beliefs, they strongly counterargue against it. As a result, they not only fail to update their beliefs, but even grow more extreme when they are challenged. In contrast, when people read information that fits or supports their beliefs, they often unquestionably accept it (Lord et al., 1979; Taber & Lodge, 2006). Consequently, exposure to information often leads to polarization – meaning that information that is consistent with prior beliefs strengthens it, but information that is inconsistent also strengthens prior beliefs – due to motivated counterarguing.

Our findings point to a more dynamic account of how this process unfolds. Following Russo's (2018) notion that information distortion does not reflect a failure to update, but a systematic departure from Bayesian independence, we find that new information continues to influence belief states, but its perceived diagnosticity is biased by the current coherence state of the system (see also Glöckner & Herbold, 2011; Holyoak & Simon, 1999). From a descriptive perspective, such bidirectional updating serves psychologically meaningful goals, including internal coherence and stability, even if it departs from Bayesian optimality (Engel & Glöckner, 2013). Information distortion thus occupies an intermediate position between normatively optimal Bayesian revision and purely belief-driven reasoning. The present findings suggest that partisan motivated reasoning operates within precisely this intermediate regime. Even under explicit partisan cues, participants' evaluations and choices remained dynamically responsive to newly encountered information, as reflected in robust trial-level updating effects. At the same time, ideological priors systematically biased how information was evaluated and integrated, redirecting judgments away from Bayesian-optimal conclusions. In line with Russo (2018), partisan bias therefore reflects coherence-preserving updating under strong ideological constraints rather than epistemic closure or the abandonment of evidence-based reasoning.

This interpretation is consistent with recent coherence-based accounts that explicitly address the relation between coherence and Bayesian inference (D. Simon & Read, 2025). Rather than viewing coherence-based reasoning as incompatible with Bayesian approaches, Simon and Read argue that coherence processes can be understood as the brain's resource-constrained implementation of inference. From this perspective, departures from Bayesian optimality arise not because evidence is ignored, but because inference is carried out through parallel constraint satisfaction under limits of time, computational resources, and biological implementation. Coherence-based reasoning thus provides a plausible account of how belief-consistent distortions can emerge from systems that remain sensitive to evidence, particularly when strong priors function as attractors that shape the interpretation of incoming information.

Within this framework, partisan motivated reasoning need not be understood as a rejection of rational updating, but as a consequence of inference under strong ideological constraints. The present findings support this view by showing that partisan cues bias evaluation and choice while leaving the underlying updating process intact. Political bias, on this account, reflects not a dichotomy between rationality and motivated distortion, but a graded trade-off between accuracy and coherence in sequential belief updating.

Theoretical Contribution: Process Specification & Parsimony

Research on biased reasoning has documented robust deviations from normative models of inference but has often lacked process-specified accounts that integrate these findings. Belief-driven accounts of partisan bias and process-driven accounts of preference construction have largely developed in parallel, leaving open whether they reflect distinct mechanisms or different operating conditions of a shared underlying process. Building on the coherence-based perspective articulated by Simon and Read (2025), the present research addresses this question by providing empirical process-level specification. Whereas Simon and Read emphasize how partisan worldviews can act as powerful coherence attractors in political reasoning, the present findings delineate how such belief-based constraints interact with ongoing preference construction during sequential information integration. Across two studies, partisan bias and information distortion arose within the same coherence-driven process, with ideological priors modulating the constraint structure under which updating unfolded rather than introducing a qualitatively distinct mechanism.

This specification clarifies how belief-driven and preference-driven influences jointly shape political judgment over time. Preference-based distortion emerged during sequential evaluation even in the absence of salient ideological cues. When partisan beliefs were activated, they entered the same coherence network and biased subsequent evaluation and choice, while sensitivity to trial-level information remained observable. In this sense, partisan motivated reasoning reflects a particular operating regime of coherence-based reasoning rather than a departure from dynamic evidence integration.

The present account also complements recent efforts toward theoretical parsimony in bias research, including belief-consistent information processing frameworks (Oeberst et al., 2025; Oeberst & Imhoff, 2023). While these approaches highlight the role of prior beliefs in shaping information processing across diverse bias phenomena, our findings specify how belief-consistent patterns can emerge from a more general, bidirectional process. From a coherence-based perspective, belief-consistent processing arises when belief-based constraints dominate an otherwise dynamically updating system.

Taken together, the present research situates partisan motivated reasoning within a unified, process-specified framework of biased reasoning. By embedding political bias in a coherence-based account of sequential decision making, it advances existing theories by clarifying how ideological beliefs and emerging preferences interact over time to shape evaluation and choice.

Boundary Conditions and Scope of the Model

Several boundary conditions of the present work warrant consideration. First, the experimental paradigm focused exclusively on positive political statements. Although this choice reflects the empirical predominance of affirmative campaign communication, it limits generalization to contexts involving negative or mixed-valence information. Negative information may engage additional affective or defensive processes that shift the balance between informational input and belief-based constraints.

Second, the model was tested in a binary choice context within a two-party political system, corresponding to the political environment of the sampled U.S. population. While this structure captures many consequential political decisions, coherence dynamics may differ in multi-option environments or multiparty systems, where competition and belief alignment are distributed across a larger set of alternatives. Extending the present framework to such contexts would help clarify how coherence-based constraint satisfaction scales with increasing structural complexity.

Third, ideological priors were operationalized through partisan identity cues that varied in clarity and diagnosticity. Although this manipulation captures a central feature of real-world political reasoning, other motivational sources—such as moral convictions, issue-specific identities, or group-based values—may impose different or additional constraints on coherence-based processing. Future work could examine whether these motivations interact with preference construction in similar or distinct ways.

Finally, although analyses of trial-level updating provide converging evidence for dynamic coherence effects, they are correlational in nature. Preferences at time *t* are assessed after evaluating information at time *t*, which limits causal inference regarding the directionality of updating. The use of preference-change measures mitigates this concern by isolating within-trial variation, but future studies employing process-tracing methods or continuous preference measures would further strengthen causal conclusions.

Taken together, these considerations clarify the scope of the present findings rather than undermining their central implication. Within the experimental contexts examined here, the results consistently indicate that partisan bias emerges from coherence-driven updating processes in which ideological priors increasingly constrain the integration of new information as preferences evolve. At the same time, the boundary conditions outlined above suggest that the relative influence of informational input and belief-based constraints may vary across informational environments, motivational contexts, and decision structures. The present work therefore provides a process-level account of partisan motivated reasoning under clearly specified conditions while identifying several avenues through which this framework can be extended.

Constraints on Generality

In line with recent calls to include explicit Constraints on Generality statements in empirical reports (Simons et al., 2017), several characteristics of the present studies delimit the scope of the conclusions. The samples used in the present research approximate the demographic composition of the U.S. population with respect to gender, age, and political identification. At the same time, the findings are situated within a specific institutional and cultural context of political decision making.

First, participants were embedded in a democratic electoral environment in which citizens regularly evaluate competing political candidates. As a result, participants are likely accustomed to weighing positive and negative information about candidates (Hibbing & Theiss-Morse, 1995; Lammers, Pauels, et al., 2022; McAllister, 2000). Second, the studies reflect the institutional structure of a two-party political system that is currently characterized by strong elite and mass polarization. In such environments, political communication is frequently organized around contrasts between opposing candidates, which may increase the salience of candidate differences when participants evaluate political statements (Druckman et al., 2013). Third, the studies were conducted in a Western cultural context that is often described as comparatively individualistic. Such contexts may further encourage attention to differences between alternatives rather than similarities when forming evaluations (Nisbett et al., 2001).

Consequently, although the coherence-based mechanisms examined here—where emerging preferences and ideological priors jointly shape the evaluation of incoming political information—are likely to extend beyond the specific context studied, the magnitude and expression of the observed effects may vary in political environments with different institutional structures, levels of polarization, or cultural orientations. Replications in multiparty systems, less polarized political contexts, and non-Western cultural settings would therefore provide important tests of the generality of the present findings.

Conclusion

Across two experiments, the present research suggests that partisan motivated reasoning is not well captured by accounts that conceptualize bias primarily as the stable influence of prior ideological commitments. Rather than reflecting a categorical suspension of evidence integration, partisan bias appears to emerge from coherence-based updating in which ideological priors increasingly constrain how new information is incorporated. The present work provides a process-level integration of two literatures that have largely developed in isolation. By embedding partisan motivated reasoning within a framework originally developed to explain predecisional information distortion, the findings demonstrate that belief-driven and preference-driven biases can arise from the same underlying coherence-seeking mechanism operating under different constraint regimes. Crucially, this integration yields explanatory leverage that neither tradition affords on its own: The results specify when ideological priors stabilize evaluations and redirect choice, and when dynamic updating remains operative despite strong partisan cues. In doing so, the present research advances a parsimonious, process-specified account of political judgment that moves beyond outcome-level descriptions of bias and clarifies how beliefs, preferences, and evidence jointly shape political reasoning over time.

Appendix 5A

Open Science Framework (OSF) Study-Specific Links

The project overview is accessible via the Open Science Framework (OSF):

https://osf.io/hdf9a/overview?view_only=e76b08bd21af43e39252048c898b04f4

Preregistrations

Study 1: https://osf.io/fj4r7/files/z2hbt

Study 2: https://osf.io/a2mxt/files/su9da

Pre-Study: https://osf.io/xnbha/files/5xuvj

Data and Analysis

Study 1: https://osf.io/rbk6s/overview?view_only=3fa9169a3f2f496ebd7935c86b2f7dc0

Study 2: https://osf.io/zvyhn/overview?view_only=618db5614eab4edaadd75c66b55b7f30

Pre-Study: https://osf.io/dxqme/overview?view_only=b9b61a04f7ec4680ae75395fd924adeb

Power Analysis Study 2:

https://osf.io/hsd5f/overview?view_only=990b7609a2944c7ba7cd0e70d2aea47a

5.3 Summary and Implications

This chapter examined partisan motivated reasoning from a process-level perspective, asking whether political bias reflects a static, belief-imposed distortion or a dynamic interaction within sequential coherence formation. Extending coherence-based models of sequential information integration, the findings indicate that partisan bias is not simply imposed prior to evaluation. Rather, ideological priors operate as top-down constraints within a single updating process in which emerging preferences and incoming evidence jointly shape evaluation and choice over time.

Across two experiments, evaluative distortion emerged as a coherence-consistent effect: once a candidate gained a preference advantage, subsequent statements were evaluated more favorably when associated with the leading candidate than with the trailing candidate. This distortion was amplified under stronger partisan cue activation, indicating that ideological beliefs function as top-down constraints that increase the weight of belief-consistent links within the coherence network.

Crucially, bias did not operate unidirectionally. In line with the bidirectional architecture of coherence-based models, emerging preferences not only biased evaluation but also fed back into belief-based categorization, shaping how the ideological meaning of statements was interpreted. Even under explicit partisan labeling, trial-level preference dynamics continued to predict both evaluative distortion and belief categorization. Political judgment thus remained dynamically responsive to incoming information rather than being exhaustively determined by static ideological priors.

Within the broader structure of the dissertation, Chapter 5 extends coherence-based models beyond their traditional focus on preference–evaluation coupling by formally incorporating stable belief systems into sequential updating dynamics. It demonstrates that belief-based and preference-based influences are not competing explanations but interacting components of a unified coherence process. In doing so, the chapter provides empirical support for integrative accounts proposing that diverse manifestations of biased reasoning can be understood as graded outcomes of a common coherence-seeking mechanism operating under varying constraint regimes.

Extending this architectural perspective to information acquisition concerns not the existence of coherence-driven search—already formalized within the iCodes framework—but the scope of coherence-based reasoning as a unifying account of multiple search biases. Chapter 6 addresses this extension by situating attraction search and confirmatory search within the same sequential coherence architecture and examining their respective contributions as preferences evolve.

Chapter 6: Extending Coherence-Based Models to Information Search

6.1 Chapter Introduction

As outlined in Section 3.2.3, research on biased information search has largely proceeded along separate theoretical lines. Attraction search has been formalized within coherence-based models, whereas confirmatory search has primarily been investigated in selective exposure paradigms emphasizing information valence and decision certainty. Consequently, these tendencies have rarely been examined within a single sequential paradigm that allows their joint operation as preferences and confidence evolve over time.

The integrated Coherence-based Decision and Search framework (iCodes; Jekel et al., 2018) specifies attraction search as a structural consequence of coherence maximization: once an option gains an activation advantage, information associated with that option becomes more likely to be selected. However, the extent to which coherence-based reasoning can function as a parsimonious framework for multiple forms of biased search remains underexamined. In particular, it is unclear whether attraction search and confirmatory search reflect distinct mechanisms or interacting expressions of a shared coherence-based architecture when tested within the same evolving decision process.

Chapter 6 develops a sequential coherence model that jointly captures attraction search and confirmatory search within a unified framework. The present paradigm allows both option identity and information valence to be available at the moment of selection, enabling the independent and additive contributions of attraction-based and confirmatory-based search to be disentangled within a single dynamic decision environment.

In addition to preference-driven attraction effects, the model incorporates subjective confidence as a process-relevant state variable that may selectively amplify confirmatory search, consistent with cognitive-economy accounts (Fischer, 2011). Across three preregistered studies employing sequential process-tracing designs, the chapter examines whether attraction search and confirmatory search constitute dissociable influences within a shared coherence architecture and how their relative contributions unfold during the stepwise evolution of preference. Chapter 6.2 is based on the following article:

Häffner, C., Jekel, M., & Lisovoj, D. (2025). *Do I search for what I like or what keeps me liking? Disentangling attraction search vs. confirmatory search in the stepwise evolution-of-preference paradigm.* Manuscript in revision.

Minor modifications were made to headings and formatting to align with the dissertation style. The content of the manuscript remains unchanged.

6.2. Manuscript III: Do I Search for What I Like or What Keeps Me Liking? Disentangling Attraction Search vs. Confirmatory Search in the Stepwise Evolution-of-preference Paradigm

Abstract

Do people search for what they already like—or for what allows them to keep liking it? This paper disentangles two accounts of biased information search. Attraction search reflects a tendency to seek additional information about the currently preferred option, regardless of whether this information is preference-confirming or preference-disconfirming. Confirmatory search, in contrast, reflects a tendency to favor preference-confirming and avoid preference-disconfirming information. Building on and extending the integrated coherence-based decision and search framework (iCodes), we develop a sequential coherence model that jointly captures these mechanisms in evolving decisions. Across three preregistered studies, we test the subtractive and additive contributions of attraction- and confirmatory-based search. We find a robust attraction search effect across all three studies. Confirmatory search, by contrast, did not appear as a general predecisional bias but showed a more selective and context-dependent pattern, with only limited evidence for moderation by subjective confidence. A consistent bias toward positive information further shaped search behavior independently of both mechanisms. These findings provide novel evidence for the distinct and interacting roles of attraction search, confirmatory search, and valence-based asymmetries in guiding information acquisition as preferences unfold during decision making.

Keywords:

attraction search, confirmatory search, parallel-constraint satisfaction

Do I Search for What I Like or What Keeps Me Liking? Disentangling Attraction Search vs. Confirmatory Search in the Stepwise Evolution-of-preference Paradigm

Emerging preferences shape how information is interpreted long before a decision is made. A large body of research demonstrates that when information about decision options is presented sequentially in the evolution-of-preference paradigm, individuals systematically distort their evaluations to align with emerging preferences (Russo, 2015). This phenomenon, known as information distortion, leads to more favorable evaluations of information supporting the preferred option and more negative evaluations of information supporting the non-preferred option (Blanchard et al., 2014). While the replicability and boundary conditions of information distortion are well established (Russo, 2015; Russo et al., 1998), research has primarily focused on how individuals evaluate and integrate information in this paradigm. Much less is known about how emerging preferences influence information search behavior, particularly in sequential decision-making tasks, where information is revealed stepwise. Yet, real-world decisions often require both: decision-makers must not only evaluate and integrate information but also choose what information to search for in the first place.

Information search, like information evaluation, can be biased. Some biases stem from cognitive constraints: In pseudodiagnosticity, people evaluate evidence only with respect to a focal hypothesis, neglecting implications for alternatives (Fischhoff & Beyth-Marom, 1983), which can bias subsequent search. In positive hypothesis testing, people preferentially sample cases that would confirm the focal hypothesis, even though sampling disconfirming cases would often be more informative (Klayman & Ha, 1987, 1989; Navarro & Perfors, 2011). Dual-process theories of reasoning suggest that focusing on a single hypothesis at a time is a general characteristic of human thinking (Evans et al., 2002), leading to systematic distortions in both search and evaluation (Mynatt et al., 1993). A more recent line of work documents a robust tendency to sample information about the currently leading option—even when the valence of that information is unknown—thereby reinforcing focus on the preferred option and producing a pattern akin to pseudodiagnosticity (Carlson & Guha, 2011; Jekel et al., 2018; Scharf et al., 2019). Importantly, it strengthens under restricted or costly search conditions, a pattern difficult to attribute to attentional limits alone and suggestive of emerging preferences guiding acquisition, potentially to consolidate the leading option when cognitive or situational resources are limited. Other search patterns have been associated more clearly with motivational factors. Research on selective exposure shows that people tend to seek attitude-consistent or decision-consistent information while avoiding contradicting evidence, especially under conditions that trigger defense motivation (Fischer & Greitemeyer, 2010; Hart et al., 2009; Nickerson, 1998). These findings suggest that both cognitive and motivational influences can shape information acquisition before integration takes place.

Despite this rich literature, biased search has rarely been studied within the stepwise evolution-of-preference (SEP) paradigm, where information unfolds trial by trial and preferences develop dynamically, allowing the link between emerging preferences, confidence, and search to be examined. Existing studies offer important but incomplete insights. Carlson and Guha (2011) contrasted leader-focused with leader-supporting search and found a robust focus on the emerging leader but measured search only once after a preliminary information phase and therefore could not examine search alongside emerging preferences. Chaxel et al. (2013) studied information search for a single decision object, which allowed them to capture confirmatory tendencies toward that object but structurally precluded leader-focused search. Because there was no competing option, search could only be consistent or inconsistent with one object's evaluation, making attraction-based tendencies unobservable in their paradigm. Mischkowski et al. (2021) applied the stepwise evolution-of-preference paradigm to legal judgments and observed strong coherence effects in evaluation of information accompanied by unexpected disconfirmatory search. Similar to Chaxel et al. (2013), their domain involved a single focal hypothesis (guilt), so all information pertained to the same target. Thus, although disconfirmatory choices allow testing for violations of confirmatory search, the single-target structure makes attraction-based search untestable because all information pertains to the same option. Consequently, neither article can disentangle confirmatory and attraction-based search or examine their interplay with dynamically evolving preferences and confidence. Addressing this limitation requires a diagnostic task with at least two distinct options and a framework that links the evolution of these biases to preferences, confidence, and search within a model, as a fully integrative account of these relations over the course of the information presentation is currently lacking.

From Descriptive Biases to a Formal Process Account

The evidence reviewed above suggests that emerging preferences bias both the integration and the acquisition of information in sequential decision making. For biased information integration, a formal account already exists: DeKay (2015) applied a coherence-based model to the evolution-of-preference paradigm, showing that information distortion emerges when early preference states feed activation back into the interpretation of subsequently presented cues. More broadly, coherence-based or parallel-constraint-satisfaction models (Glöckner & Betsch, 2008b; Holyoak & Simon, 1999; Thagard, 1989) conceptualize decisions as the gradual stabilization of a maximally coherent representation. In these models, bidirectional activation among information and options ensures that even weak initial preference signals are amplified, producing coherence effects—most notably systematic shifts that strengthen evaluations of preference-consistent information while suppressing evaluations of preference-inconsistent information, and faster responses as the network converges on a coherent interpretation (Glöckner & Betsch, 2012; Simon, 2004).

Extending these models to search, the iCodes framework (Jekel et al., 2018) generalizes coherence dynamics from evaluation to information acquisition. The model assumes that, as preferences begin to form, the options gain different activation levels in the coherence network. This imbalance increases the activation of information associated with the currently leading option, making it more likely to be selected next. Consequently, the model predicts a structural tendency to sample information about the more favorable option—the Attraction Search Effect—framing leader-focused search as a byproduct of coherence maximization rather than a motivational preference.

Existing studies testing iCodes (Jekel et al., 2018; Scharf et al., 2019), however, have relied on environments in which participants knew only whether the information they could select concerned the leader or the trailer and how valid it was. Because the valence of the information was concealed, these designs cannot distinguish coherence-driven attraction from motivational confirmatory search—the defense-oriented tendency to select preference-consistent information (Fischer, Lea, et al., 2011; Fischer & Greitemeyer, 2010). Consequently, no prior work has examined these structural and motivational forces together in a fully sequential, valenced, trial-by-trial paradigm.

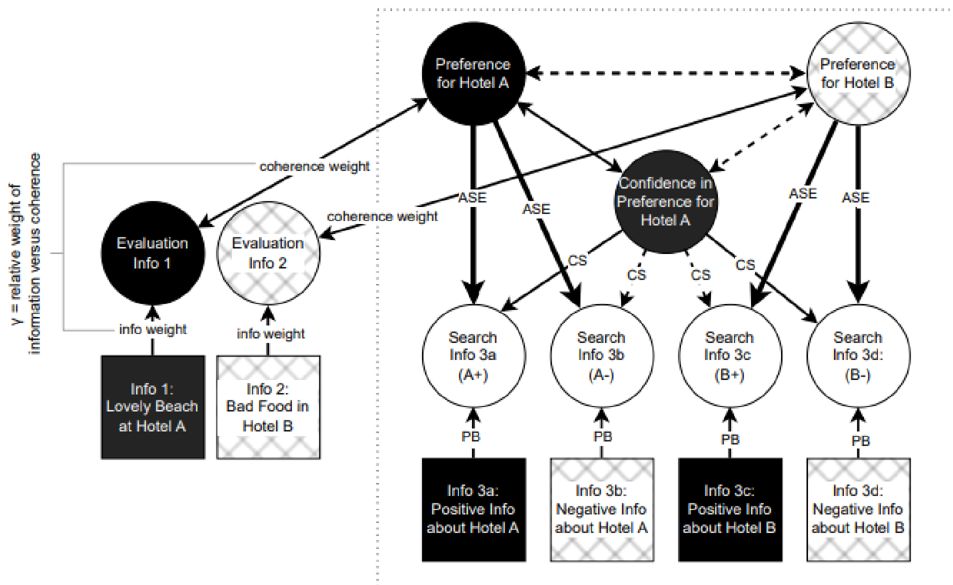
The present studies address this gap by tracing how both tendencies evolve over the course of preference formation. Importantly, prior work shows that confirmatory tendencies increase as decision certainty rises—a pattern attributed to the reduced cognitive effort required for processing preference-consistent information (Fischer, 2011). By tracking search behavior alongside evolving preferences and confidence, we provide a unified account of how these forces jointly shape information acquisition. The following section introduces the broader theoretical perspective within which we embed this account.

A Coherence-based Perspective on Attraction and Confirmatory Search

Recent theoretical work by Simon and Read (2025) argues that a wide range of reasoning and judgment biases can be understood as natural consequences of constraint-satisfaction dynamics. In line with earlier foundational models of coherence-based processing (Holyoak & Simon, 1999; Read et al., 1997; Thagard, 1989), their analysis highlights how emerging interpretations steer cognitive processing by aligning beliefs, evidence, and goals within a coherent representational structure. Complementary work on belief-consistent information processing (e.g., Oeberst et al., 2025) similarly suggests that structural and motivational influences can be captured within a common coherence framework. Prior work further indicates that coherence shifts are accompanied by high decision confidence and can reproduce classic confirmation-bias patterns (e.g., Holyoak & Simon, 1999; Simon et al., 2004; Simon & Spiller, 2016) hinting at distinct functional roles for preference, confidence, and motivational nodes within such networks.

Figure 1.

A coherence-based network model of information integration and search. Search for information types (four white nodes) is influenced by three search tendencies: attraction search effect (ASE), confirmatory search (CS), and positivity bias (PB).



Extending this line of work, we propose a network architecture in which emerging preferences and confidence jointly guide information search in sequential decision making. Specifically, we propose a coherence-based reasoning model for simple decision and search tasks with two or more options that participants choose between after an initial phase of information search and presentation (e.g., customer reviews about hotels). In this model, sequentially presented information is represented as rectangular nodes in Figure 1—for example, the first piece indicating that Hotel A has a lovely beach (a positive piece shown as a black node) and the second indicating that Hotel B has bad food (a negative piece shown as a cross-hatched node). The subjective evaluation of this information appears in the second layer, and these evaluations are linked to the option nodes in the third layer that represent emerging preferences, with a black node indicating a positive preference for Hotel A and a cross-hatched node indicating a negative preference, that is, a negative evaluation, for Hotel B. The two option nodes inhibit each other, as indicated by a dotted line, reflecting the mutually exclusive nature of the choice. Because all links are bidirectional, information influences preferences, but preferences also shape the evaluation of information, producing coherence effects that amplify preference-consistent evaluations and attenuate preference-inconsistent ones (hypothesis 1a). This part follows the core mechanisms of existing well-validated coherence-based network models of information integration and decision making (DeKay, 2015; Glöckner et al., 2014; Glöckner & Betsch, 2008b).

Our target part of the model concerns search, as shown in the search box of the network, enclosed by a dotted rectangle in Figure 1. Participants can choose to see either positive or negative information about Hotel A or Hotel B next, represented by the four rectangles in the input layer. The emerging preference for Hotel A due to the first two pieces of information increases activation of the nodes concerning information about option A (i.e., 3a and 3b), whereas the lack of preference for Hotel B reduces activation of the nodes concerning information about option B (i.e., 3c and 3d). This part of the model is identical to the well-validated iCodes model (Jekel et al., 2018; Scharf et al., 2019) and represents the attraction search effect (hypothesis 2), as indicated by the labels on the connections between option nodes and search nodes.

A second influence in the model arises from the confidence node, which reflects the momentary strength of the preference for one option over the other. As confidence increases, activation flows from the preference representation to the valenced search nodes, increasing the activation of preference-consistent information (positive information about the leader and negative information about the trailer) and reducing activation of preference-inconsistent information. This mechanism captures confirmatory search in the model: a bias toward information that aligns with one's emerging choice. Although this confidence-based influence has not yet been formalized in prior coherence network models, it provides a parsimonious implementation of two robust findings—namely, that defensive tendencies (Fischer & Greitemeyer, 2010) and reduced processing of inconsistent information under higher certainty (Fischer, 2011) both contribute to confirmatory search. In this paper, we distinguish between two predictions derived from this account. The first concerns the existence of confirmatory search when information valence is known (hypothesis 3a), whereas the second concerns the strengthening of this tendency as confidence increases (hypothesis 3b).

Our model primarily builds on attraction- and confirmation-based processes, but it also includes a positivity bias (PB) that increases the activation of positive over negative information. This is represented by positively (black rectangles) and negatively activated (cross-hatched) input nodes linking to the search nodes as a third effect. Several theoretical perspectives suggest why such a tendency could occur without being specific predictions of our framework. Regulatory focus research shows that a promotion focus encourages eager, advancement-oriented strategies and heightened attention to gain-relevant positive cues (Crowe & Higgins, 1997; Idson et al., 2000), which aligns with the goal of identifying the best option in our task. In addition, positive information is typically more fluent (Unkelbach & Greifeneder, 2013) and slightly more pleasant to process (Knobloch-Westerwick & Jingbo Meng, 2009), whereas negative information often increases perceived conflict and evaluation effort, consistent with conflict-based accounts of decision difficulty (Tversky & Shafir, 1992). These perspectives may help explain why a valence tendency can emerge, but we consider this component exploratory because we added it after observing the positivity bias in Study 2 (post-hoc hypothesis 4) and replicating it in Study 3 (a-priori hypothesis 4), and as potentially context-dependent rather than a generalizable feature of the model, as a bias for negative information in other contexts is plausible (e.g., search for exculpatory evidence in legal or medical contexts). Together, these three effects—attraction search, confirmatory search, and positivity bias—jointly shape search as preferences evolve, enabling the model-based predictions tested in the subsequent studies.

Relation Between Search and Information Distortion

Scharf et al. (2022) recently showed that the attraction search effect was stronger for participants who indicated their current preference before information search. They attributed this effect by heightened preference awareness, which strengthened the impact of preferences on search as predicted by the preference-search relation in iCodes. To account for this, they extended the iCodes model by setting the weight linking information to its evaluation (info weight in Figure 1) relative to the weight linking evaluation to preferences (coherence weight in Figure 1), with this ratio treated as a free γ-parameter (Scharf et al., 2022). We predict that increasing awareness of pre-search preferences will also strengthen information distortion (hypothesis 1b), a prediction that follows directly from the model via the γ-parameter but was not tested in Scharf et al.'s study because their focus was on search rather than evaluation. Because engaging with additional information during search draws attention back to the options and their developing evaluations, which may increase preference awareness and potentially strengthen preferences (Wolf et al., 2019), the model therefore implies stronger information distortion when people search rather than when they receive information without search.

Overview of Studies

We test five hypotheses derived from our model (Figure 1) across three studies: preference-consistent evaluation (H1a), stronger information distortion under heightened preference awareness (H1b), attraction-based search (H2), the existence of confirmatory search (H3a) and its predicted confidence dependence (H3b), and positivity bias (H4). Study 1 examines H1a, H1b, and H2 in the stepwise evolution-of-preference paradigm under concealed information valence. Studies 2 and 3 disentangle attraction-based from confirmatory search by introducing explicit valence cues. Study 2 tests whether confirmatory search emerges when the valence of all upcoming information is known (H3a), using a ranking-based selection task across the four information types (i.e., positive or negative information about the leading or trailing option, where leader and trailer are defined by individuals' emerging preferences). In Study 3, participants decide which information to search next in all possible pairwise comparisons of the four types of information, making search decisions more consequential and forcing direct competition between information types. Because Study 3 additionally includes trialwise confidence assessments, it provides a direct test of the predicted confidence–confirmatory search relation (H3b).

Open Materials, Data, and Analysis Code

All materials, anonymized data, analysis scripts, and power analyses for Studies 1–3 are available on OSF:

https://osf.io/58rw2/overview?view_only=e3462bad156f4046826a89f0ebee4d90

The repository contains three study-specific components ("Study 1", "Study 2", "Study 3"), each providing all materials, data files, analysis code, and the corresponding power analysis. Materials and procedures for the two pre-studies used to validate the information pool are also available on OSF: (1) https://osf.io/xpnqs/?view_only=1e3e91d19f5f4b89a0ec7f89ed151343 , (2) https://osf.io/f9rn5/?view_only=279620ab5ac0456394e7f01c8d755498. Materials and data from the name pretest used to establish neutral option labels are available at https://osf.io/fm295/?view_only=1faba23cf110478a836f1d949f46ea20

Preregistrations

All studies were preregistered prior to data collection. The preregistration documents for each study are available via OSF at the following project pages: (1) Study 1: https://osf.io/cazm8/?view_only=b071a88ff4ce465eae02c3775602c0b4, (2) Study 2: https://osf.io/y3t6c/?view_only=7ed0b5bf2ee347938cccbf21d72ec4af, (3) Study 3: https://osf.io/yugrz/?view_only=e76035f489164a4e85e0442bdde4580a

Study 1: Attraction Search under Concealed Information Valence

Study 1 offers an initial test of the model in a setting where participants can choose only whether to receive information about the leading or the trailing option, without knowing the valence of that information. This design isolates attraction-based search: if coherence-driven preference activation guides information acquisition, participants should preferentially search the currently leading option even when information valence is concealed. Study 1 also examines whether active selection of information amplifies information distortion. Because choosing which option to search directs additional attention to the evolving preference state, search should heighten coherence pressures during evaluation. We therefore compare information distortion under search with distortion under passive exposure, where participants receive the same information as in the search condition but without searching for it.

Method

Participants

The study was conducted as an online experiment programmed in *lab.js* (Henninger et al., 2022) and hosted via *JATOS* (Lange et al., 2015). Participants were recruited via the German online panel provider Toluna (https://www.toluna-group.com/) and received standard compensation for a study of approximately 15 minutes. Eligibility required participants to be at least 18 years old and to have German language proficiency of at least C1 level.

To determine an adequate sample size for Study 1, we conducted a bootstrap-based power analysis using data from a pilot study employing the same paradigm. In each of 1,000 bootstraps, participants were resampled with replacement and full trial-level datasets were reconstructed, preserving within-participant structure. For each resampled dataset, we refit the preregistered mixed model on information distortion and tested whether leader-directed search significantly exceeded chance-level. A sample of $N = 200$ yielded statistically significant effects in at least 88% of cases for both outcomes. To ensure balanced cells and sufficient power for between-condition comparisons, we therefore targeted a total sample size of $N = 400$ (approximately 200 in the search condition and 100 in each control condition). The simulation script is available on the OSF project page (see links above).

A total of $N = 848$ individuals participated. In line with preregistered exclusion criteria, participants who failed both attention checks were excluded ($n = 353$), resulting in a final sample of $N = 495$ participants ($M_{age} = 49.55$, $SD = 16.02$; age data missing for two participants; 47.9% female). Recruitment exceeded the preregistered target to compensate for higher-than-expected exclusion rates, ensuring that the planned number of analyzable cases was reached. The final sample therefore included $n = 222$ participants in the active search condition, $n = 144$ in the random-order control, and $n = 129$ in the fixed-order control condition.

Design

The experiment used a fully crossed 3 (search condition: active vs. passive fixed-order vs. passive random-order; between-subjects) × 2 (information type: leader vs. trailer; within-subjects) × 2 (decision context: hotels vs. dentists; within-subjects) mixed design. Participants decided between two hotels and two dentists based on information they could either search for or were presented with in the passive conditions. Depending on their search choices or the presentation order, information belonged to either the leading or trailing option as defined by participants' emerging preferences for the options. Participants were randomly assigned to one of three between-subjects conditions. In the active search condition, participants selected which of the two options they wished to receive the next piece of information about. We included two control groups to separate effects of search from effects of information content and order. Each control participant was matched to a unique experimental participant and shown the same information that this participant had selected, keeping information content constant across groups. The fixed-order control group preserved the experimental participant's search sequence, whereas the random-order control group broke this sequence by presenting the same information in random order. Including both control groups allowed us to test whether the stronger information distortion predicted for search requires the specific order generated by selection and to assess whether order itself influences distortion. Control group assessments were conducted immediately after data collection from the corresponding participants in the active search condition. The order of decision contexts, option side, and labeling was randomized for participants in the active search condition and mirrored for their matched control participants in the passive conditions to ensure structural equivalence across conditions.

Materials and Procedure

Decision Contexts and Information Pool

Participants completed two decision tasks—one about hotels and one about dentists—each involving a comparison between two unfamiliar options described by eight sequentially presented pieces of information (four per option). All information was selected from a validated pool of 124 pre-rated customer reviews originally sourced from real online platforms. The stimulus material was developed and validated in two independent online pre-studies (total N = 424; 293 female; M_{age} = 28.52, SD = 8.97).

Information reflected typical online reviews and referred to concrete, decision-relevant aspects. For example, hotel-related information described features such as limited breakfast options, room size, or access to nearby leisure facilities, whereas dentist-related information addressed issues such as appointment flexibility, practice atmosphere, or clarity of cost communication.

In the pre-studies, participants rated each information independently on a continuous scale (−100 = strongly negative to +100 = strongly positive), indicating how strongly the information would influence a decision for or against an option. No choice or preference formation was involved, yielding preference-independent estimates of information diagnosticity and valence, consistent with standard procedures for baseline measurement in the stepwise evolution-of-preference paradigm (Russo, 2015)

For the present study, highly negative items were excluded to prevent early disconfirmation. From the remaining pool, items with comparable diagnosticity were paired and assigned to opposing options, ensuring balance in overall evaluative strength while preserving variability for preference formation. The final stimulus set comprised 16 items per decision context (eight per option), with diagnosticity values ranging from −33 to +62.

Labels for Decision Options

Because the formation of preferences requires identifiable options, a pre-test was conducted to screen for neutral fictitious names used in the experiment to label options. In a pilot study, thirty-three participants (21 female; $M_{\text{age}} = 28.87$, SD = 8.69) rated 20 randomly generated, pronounceable nonwords following German phonotactic rules from three content domains (hotels, dentists, bluetooth speakers) on a −100 to +100 scale, indicating "how strongly this name would influence [them] personally in a potential decision for or against the option." Only names with mean ratings not significantly different from zero and narrow confidence intervals were retained. For the current study, four names (two per context) were selected to ensure neutrality.

Procedure

Participants made two sequential decisions between two hotels and two dentists. Following the stepwise-evolution-of-preference paradigm, information was presented one piece at a time, and each item was followed by two ratings: (a) participants rated its influence on their decision on a continuous scale from −100 (*strong negative influence*) to 100 (*strong positive influence*), and (b) indicated their current preference between the two options using a continuous slider (−100 strong preference for A to +100 strong preference for B, with 0 = no preference). Preference ratings were framed as interim judgments to capture the dynamic formation of a leading option. Depending on their assigned condition, participants either determined which option to search next (active search condition) or were presented with the same set of information determined by a matched participant (control conditions). After viewing all eight information items, participants made a final decision between the two options and reported their confidence in the decision. Two slider-based attention checks were included in which participants were asked to move the slider to a specific position; only participants who passed both were retained for analysis.

Results

All analyses were conducted using R (R Core Team, 2019). Visualizations were generated with the ggplot2 package (Wickham, 2016a), and all mixed model analyses were carried out using the lme4 (Bates et al., 2015b) and lmerTest (Kuznetsova et al., 2017) packages.

Information Distortion

Information distortion was computed as the deviation between a participant's evaluation of each item and its baseline evaluation obtained from the pre-study sample. Distortion was classified as leader or trailer distortion based on the preferences participants indicated in the preceding round of information presentation; thus, information distortion was calculated from the second round of information presentation onward. A linear mixed-effects model with a random intercept for participants and effect-coded information type (leader = +1, trailer = −1) showed a robust information distortion effect: leader information received more positive evaluations than trailer information, $b = 5.72$, $SE = 0.49$, $t(5825.11) = 11.63$, $p < .001$.[5] To assess whether information distortion differed between search and passive exposure, we added search condition

[5] Because standardized effect size metrics for mixed-effects models remain heterogeneous across the literature, we report all analyses in their unstandardized model metrics (logits). For transparency, we additionally computed marginal and conditional R^2 values using the r2beta() procedure (package performance). The full code and effect size outputs for all models are provided on OSF.

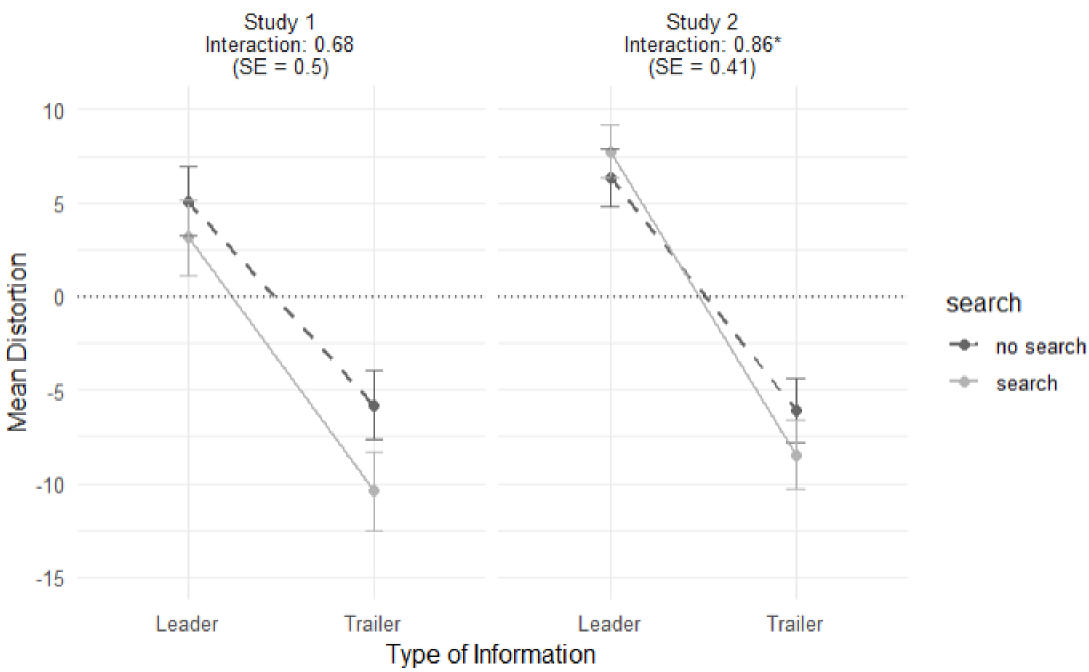
as a predictor (effect-coded as search = +1, no search = −1). The interaction between information type and search condition was not significant, $b = 0.68$, $SE = 0.50$, $t(5833.95) = 1.38$, $p = .168$, suggesting comparable distortion magnitudes in search and no-search conditions. Figure 2 (left panel) displays mean distortion scores by condition and information type.

Because the preregistration specified separate tests for the two passive control conditions, we next estimated a model with a three-level condition factor represented by two orthogonal contrasts. One contrast compared active search with the two passive conditions combined (coded as 1, −.5, −.5), and a second contrast compared the two passive conditions directly (coded as 0, +1, −1). Both contrasts were interacted with information type (leader vs. trailer) to test whether the magnitude of information distortion differed across conditions. Neither interaction was significant (both $ps > .17$), indicating no evidence that the leader–trailer distortion effect differed between active search and passive exposure or between the two passive control conditions.

In an exploratory analysis restricted to trials, in which an information about the non-preferred option was evaluated (i.e., trailer information), a follow-up mixed model suggested that participants in the search condition evaluated information about the non-preferred option more negatively than those in the control groups, $b = 2.32$, $SE = 0.93$, $t(430.64) = 2.49$, $p = .013$.

Figure 2.

Mean information distortion by condition (search vs. no search) and information type (leader vs. trailer) in Study 1 and Study 2.



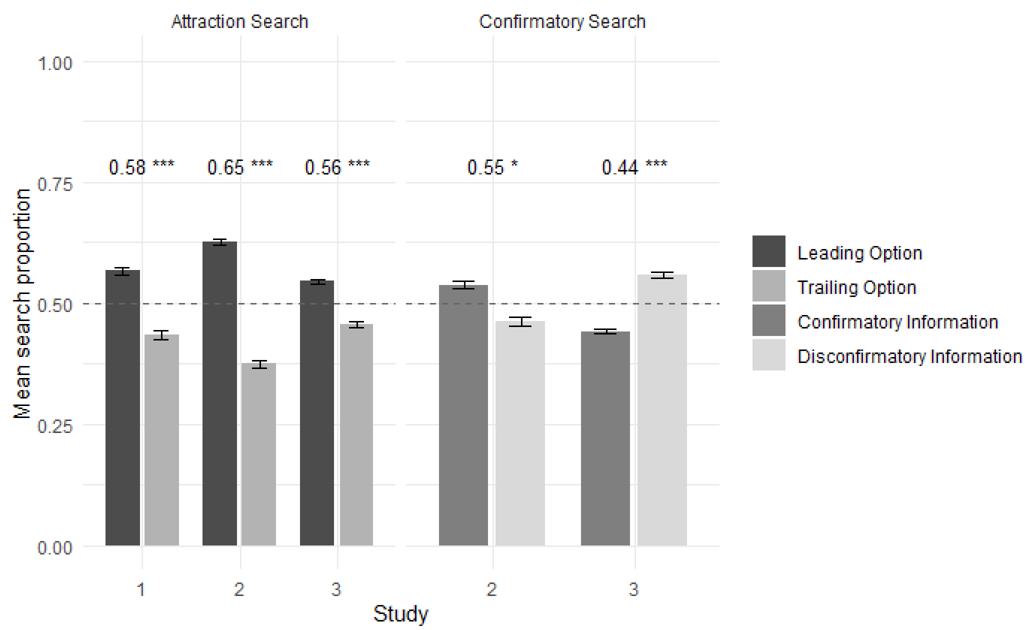
Note. Mean distortion for leader vs. trailer information in the search and no-search conditions. Solid lines indicate search; dashed lines indicate no search. Error bars show ±1 SE. Facet labels report the Information type × Search interaction from mixed-effects models with random intercepts for participants: $p < .05$ (*), $p < .01$ (**), $p < .001$ (***).

Attraction Search

Participants in the search condition showed a reliable attraction search effect. On average, participants selected information about their currently preferred option on 56.09% of trials (see Figure 3, left panel). A generalized linear mixed-effects model with a random intercept for participants yielded a positive intercept, $b = 0.30$, $SE = 0.07$, $z = 4.38$, $p < .001$, indicating a systematic tendency to sample information from the leading rather than the trailing option. To assess whether attraction search increased with preference strength, standardized preference strength from the preceding trial was added as a predictor to the regression model. The coefficient was positive but not statistically significant, $b = 0.09$, $SE = 0.05$, $z = 1.71$, $p = .087$, suggesting no reliable evidence that attraction-based search increased with preference strength.

Figure 3.

Information search proportions across information types and search strategies in Studies 1–3.



Note. Mean search proportions by information type across studies. The left panel depicts attraction search—the probability of selecting information about the currently leading versus trailing option. The right panel depicts confirmatory search—the probability of selecting preference-consistent (confirmatory) versus preference-inconsistent (disconfirmatory) information; this measure was assessed only in Studies 2 and 3. Error bars represent 95% confidence intervals. The dashed line at .50 indicates chance-level search. Numerical values above bars denote model-based predicted probabilities derived from intercept-only mixed-effects logistic models (inverse-logit transformed). Asterisks indicate significance relative to chance ($p < .05$, *$p < .01$, **$p < .001$)

Discussion

Study 1 provided an initial test of the evaluation and search components of our sequential coherence model. First, we replicated the predicted information distortion effect (H1a): leader information was evaluated more positively than trailer information, consistent with coherence-driven bidirectional links between evaluations and emerging preferences.

The predicted increase in distortion under search (H1b) was not supported. Although search was expected to increase preference awareness—and thereby strengthen coherence-driven distortion—distortion magnitudes were comparable across groups. An exploratory analysis restricted to trailer information suggested more negative evaluations in the search condition than in the passive conditions, but because this analysis was exploratory, the finding requires replication and does not provide compelling evidence for H1b. Concerning search, Study 1 provides clear support for the attraction search effect (H2): participants selectively sampled information about the leading option. This aligns with the model's prediction that coherence-driven activation of the preferred option node increases activation of search nodes associated with information about that option. Preference strength did not reliably predict this bias, suggesting that even weak emerging preferences are sufficient to trigger attraction-based search. Study 2 builds on these findings by introducing explicit valence cues, allowing a more direct test of how attraction- and confirmation-based search emerge within the full sequential coherence architecture.

Study 2: Disentangling Attraction Search and Confirmatory Search

Study 1 demonstrated attraction-based search under concealed information valence, whereas Study 2 extends this test to conditions in which the valence of all upcoming information is explicit, allowing attraction- and confirmation-based search to diverge. This allows us to assess whether confirmatory search emerges when participants can anticipate whether information will support or contradict their current preference (H3a), while also examining whether attraction search persists when valence is known (H2).

Method

Participants

The study was conducted as an online experiment using *lab.js* (Henninger et al., 2022) and hosted via *JATOS* (K. Lange et al., 2015). Recruitment and compensation were organized through the panel provider Bilendi (https://www.bilendi.de/). The total duration was approximately 15 minutes. Participants were required to be at least 18 years old and to have German language proficiency of at least C1 level.

To ensure adequate statistical power, a bootstrap-based power analysis was conducted using the combined data from the pilot study (previously used for the Study 1 power analysis) and the data from Study 1. The analysis focused on detecting between-condition differences in negative information distortion about the trailer and the attraction search effect. Simulations indicated that with n = 600 participants (approximately 300 per condition), more than 80% of bootstraps yielded statistically significant regression weights for the target effects. Accordingly, the sample size was set to N = 600 participants.

A total of N = 878 individuals participated in the study. After applying preregistered exclusion criteria—failure to answer both attention checks correctly or completion times below 5 minutes (search) or 3 minutes (passive control)—the final sample comprised N = 618 participants (n = 303 = 49.0% female, (M_{age} = 45.84, SD = 15.56), with n = 302 in the active search condition and n = 316 in the control condition.

Materials and Procedure

Study 2 closely followed the structure of Study 1, using the same sequential decision paradigm. Participants again completed two decision tasks (hotels and dentists), each involving eight pre-rated information (four per option). The main change concerned the information selection process: In each round, participants were provided with a list of four information types (positive vs. negative x leader vs. trailer) and asked to rank them from 1 (most preferred to search next) to 4 (least preferred). The information ranked 1 was then presented.

Because the ranking-based procedure allowed participants to repeatedly select positive or negative information, the stimulus set from Study 1—balanced across options rather than valence—was no longer suitable. To provide a neutral basis for all possible search paths, we created a new valence-balanced pool consisting of eight positive and eight negative items for each decision context. These items were drawn from the validated pre-study pool but were not identical to those used in Study 1. Highly negative items were again excluded to support early preference formation, and the final pool covered a diagnosticity range from −53 to +65. When an information was chosen, it was displayed as information about whichever option (A or B) the participant selected in that round.

Identical to Study 1, participants were randomly assigned to a search condition or a control condition. In the active condition, participants ranked the four available information types on each trial, and only the top-ranked information was revealed and evaluated. Because Study 1 showed no reliable differences between the two control groups (fixed vs. randomized order), only the fixed order control was included in Study 2. In the passive condition, participants were presented with the same top-ranked information as a randomly matched participant from the active condition, in the exact order that participant had selected.

Results

Information Distortion

A linear mixed-effects model with a random intercept for participants and effect-coded information type (leader = +1, trailer = −1) again revealed a robust information distortion effect: leader information was evaluated more positively than trailer information, $b = 5.91$, $SE = 0.41$, $t(7347.76) = 14.49$, $p < .001$. Including search condition (effect-coded as search = +1, no search = −1) yielded a significant interaction with information type, $b = 0.83$, $SE = 0.41$, $t(7346.58) = 2.02$, $p = .044$, indicating stronger distortion under search. Leader information was evaluated more positively—and trailer information more negatively—when participants selected the information (see Figure 2, right panel).

Separate Tests of Search Biases and Preference-Related Factors

Participants in the search condition again showed a reliable attraction search effect. A generalized linear mixed-effects model with a random intercept for participants yielded a significant positive intercept, $b = 0.61$, $SE = 0.07$, $z = 8.86$, $p < .001$, indicating a systematic tendency to search information from the leading rather than the trailing option. On average, participants selected information about their currently preferred option on 61.40 % of trials (see Figure 3, middle panel). To examine whether attraction search varied with preference strength, standardized preference strength from the preceding trial was added as a predictor. In contrast to Study 1, participants were significantly more likely to sample leader information when their preference for the leader had been stronger on the previous trial, $b = 0.16$, $SE = 0.05$, $z = 3.40$, $p < .001$, suggesting that the absence of this effect in Study 1 may have been due to limited statistical power.

To examine whether information about the preferred option was also prioritized in the ranking task, we estimated a cumulative link model predicting the ranked order of the four information types from a binary predictor coding whether the item referred to the preferred versus the non-preferred option (1 = preferred option, 0 = non-preferred option). In the ranking task, lower numeric ranks indicated higher priority (1 = most preferred to see next; 4 = least preferred). The model yielded a robust effect, $b = –0.63$, $SE = 0.03$, $z = –20.59$, $p < .001$, indicating a consistent tendency to prioritize information about the currently preferred option.[6]

Confirmatory search as the dependent variable was coded analogously to attraction search (1 = preference-consistent, 0 = preference-inconsistent). A binomial mixed-effects model with a random intercept for participants yielded a positive intercept, $b = 0.20$, $SE = 0.08$, $z = 2.38$, $p = .017$, indicating a tendency to select preference-consistent information. Aggregated across trials, confirmatory information was chosen on 52.31% of trials (see Figure 3, right panel). Exploratory analyses further showed that confirmatory search increased across trials, $b = 0.05$, $SE = 0.02$, $z = 2.36$, $p = .018$, whereas standardized preference strength from the preceding trial was not significantly associated with confirmatory search, $b = 0.095$, $SE = 0.051$, $z = 1.85$, $p = .065$.

To examine whether confirmatory information was also prioritized in the ranking task, we estimated a cumulative link model predicting the ranked order of the four information types from a binary predictor coding whether the information was confirmatory versus disconfirmatory (1 = confirmatory, 0 = disconfirmatory). The model showed, contrary to our hypothesis, that

---

[6] The preregistration incorrectly specified random intercepts for trial number in the ranking models. Because trial number reflects a fixed sequence rather than a randomly sampled factor and random-effects models did not converge, ranking results are reported from cumulative link models without random effects; fixed-effect estimates were virtually identical. All search models in Studies 2 and 3 included participant-level random intercepts.

confirmatory information was assigned less favorable ranks than disconfirmatory information, $b = 0.09$, $SE = 0.03$, $z = 3.09$, $p = .002$, indicating that participants gave confirmatory information lower priority.

Joint and Dissociable Effects of Attraction and Confirmatory Search

To evaluate whether attraction- and confirmation-search tendencies independently and additively contributed to information selection, we modeled both search decisions and rank-order preferences using two binary contrasts that captured the theoretically relevant dimensions: an attraction search contrast (leader = 1, trailer = 0) and a confirmatory search contrast (preference-consistent = 1, inconsistent = 0). Their interaction identified the joint case in which both tendencies aligned (i.e., positive leader information) relative to baseline (i.e., positive trailer information). Search probabilities were analyzed using a linear mixed-effects model, and ranking priorities (lower ranks indicating higher preference) were analyzed using an ordinal mixed model. Both models included random intercepts for participants (and for trials in the ranking model). Descriptive patterns are shown in Figure 4, and full estimates appear in Table 1.

When information aligned with both tendencies (i.e., positive leader information) was contrasted with baseline positive trailer information, participants selected it substantially more often ($b = 0.29$, $SE = 0.03$, $t = 10.97$, $p < .001$), indicating the expected joint influence of attraction and confirmation processes. This pattern was mirrored in the ranking task: positive leader information was assigned significantly higher priority (lower numeric ranks), $b = –1.26$, SE = 0.06, $z = –20.67$, $p < .001$. Thus, information consistent with both tendencies was favored both in immediate search choices and in overall prioritization.

Attraction-only information (i.e., negative leader information) did not differ reliably from baseline in search frequency, $b = –0.03$, SE = 0.02, $t = –1.66$, $p = .098$. In the ranking task, negative leader information likewise did not receive higher priority than baseline, $b = -0.02$, SE $= 0.04$, $z = -0.57$, $p = .571$. Thus, across both outcome measures, there was no evidence that negative leader information was favored relative to positive trailer information.

By contrast, confirmation-only information (i.e., negative trailer information) was selected less frequently than baseline positive trailer information, $b = -0.12$, SE $= 0.02$, $t = -6.51$, $p < .001$. In the ranking task, negative trailer information was also assigned reliably lower priority (i.e., higher rank values), $b = 0.70$, SE $= 0.04$, $z = 16.45$, $p < .001$. Thus, across both search choices and ranking priorities, negative information about the trailing option was consistently disfavored rather than, as predicted, preferentially selected.

Table 1.

Mixed-effect regression estimates for search choices and ranking priorities across comparisons.

Comparison	Predictor	Model	*b*	*SE*	*t/z*	*p*
T+ vs. 0	Intercept	Search	0.25	0.01	19.21	< .001
L- vs. T+	AS	Search	–0.03	0.02	–1.66	.098
		Ranking	–0.02	0.04	–0.57	.571
T- vs. T+	CS	Search	–0.12	0.02	–6.52	< .001
		Ranking	0.70	0.04	16.45	< .001
L+ vs. T+	AS x CS	Search	0.29	0.03	10.97	< .001
		Ranking	–1.26	0.06	–20.67	< .001

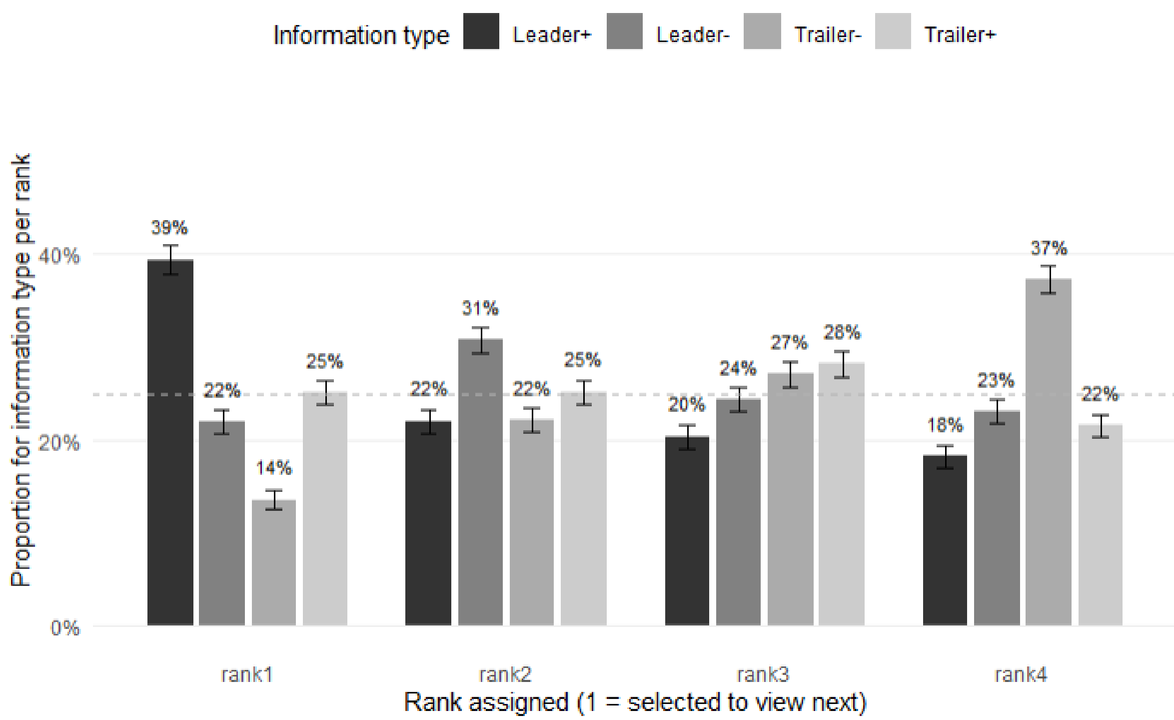
Note. Comparison lists the contrast between each information type and the baseline condition (T+ = positive trailer). Predictor refers to the contrast coding applied in the regression models: AS = attraction contrast (leader = 1, trailer = 0), CS = confirmation contrast (preference-consistent = 1, inconsistent = 0), AS × CS = product of the two contrasts. Model indicates whether effects were estimated in the linear mixed-effects search model or the cumulative link mixed ranking model (lower ranks = higher priority). The table reports unstandardized coefficients (*b*), standard errors (*SE*), and the corresponding test statistics (*t* values for Search; *z* values for ranking). All *p*-values are two-tailed.

Two Bonferroni-corrected contrasts were added to test (a) whether positive information about the currently leading option—where attraction- and confirmation-based tendencies converge—was favored over the remaining three information types, and (b) whether the attraction-based tendency (reflected in negative information about the leader) exceeded the confirmation-based tendency (negative information about the trailer). The first contrast assigned weights (–1, –1, –1, +3) to the four information types (negative leader, positive trailer, negative trailer, positive leader) to compare positive leader information against the average of all alternatives. The second contrast used weights (0, –1, +1, 0) to directly compare negative leader information (+1) with negative trailer information (–1).

In the analysis of search choices, positive leader information was selected substantially more often than the other types ($b = 0.56$, $SE = 0.05$, $t(900) = 12.32$, p < .001), whereas its ranking priority did not differ reliably in the ordinal analysis ($b = –0.19$, $SE = 0.10$, $z = –1.87$, $p = .123$). The second contrast showed that negative leader information was chosen more often than negative trailer information ($b = 0.09$, $SE = 0.02$, $t(900) = 4.86$, $p < .001$). This asymmetry was mirrored in the ranking analysis, where negative leader information was also assigned markedly higher priority than negative trailer information ($b = –0.58$, $SE = 0.04$, $z = –13.38$, $p < .001$), indicating that attraction-based tendencies outweighed confirmation-based tendencies across both outcome measures.

Figure 4.

Percentage of each information type assigned to the four ranking positions.



Note. Ranks range from 1 = highest preference to be selected next to 4 = lowest preference. Information types are defined by target (leader vs. trailer) and valence ("+" = positive, "–" = negative). Percentages are calculated within each rank so that values sum to 100 % per rank. The dashed horizontal line marks random ranking.

Discussion

Study 2 provided the first test of information search under conditions in which the valence of hidden information was explicitly available within the sequential coherence framework and replicated the information distortion effect (H1a). Critically—and in contrast to Study 1—the predicted increase in distortion under search (H1b) emerged: leader information was evaluated more positively and trailer information more negatively when participants selected the information, consistent with the model's assumption that search heightens preference awareness and thereby strengthens coherence-driven distortions. Attraction search (H2) again appeared robust: participants preferentially sampled preference-consistent information even when the valence of upcoming information was explicit. Unlike in Study 1, exploratory analyses indicated that stronger preferences were associated with a modest increase in sampling leader-related information. Because the presence of explicit valence cues in Study 2 introduces a competing confirmatory mechanism, this pattern suggests that preference strength can bolster attraction-based sampling even when confirmatory search is also available as an alternative biasing process.

For confirmatory search, participants showed a small overall tendency to select preference-consistent over preference-inconsistent information (H3a). However, contrasts revealed that this effect was driven primarily by positive leader information—a condition in which attraction-based and confirmatory predictions align. The information type most diagnostic of confirmatory search (negative information about the trailer) was selected least often, indicating that confirmatory tendencies were weak and did not show the valence-specific avoidance predicted by defense-motivated accounts. Confirmatory search also increased slightly across trials, whereas preference strength was not reliably associated with confirmatory sampling. Finally, participants displayed a consistent positivity bias: positive information was favored over negative information even when it concerned the non-preferred option.

Study 3: Testing the Relative Influence of Attraction and Confirmatory Search Under Forced-Choice Conditions

Study 3 builds on Study 2 by implementing a pairwise forced-choice paradigm that puts attraction- and confirmation-based search tendencies into direct competition. In Study 2, all four information types —positive and negative information about the leader or the trailer (L+, L–, T+, T–)—were presented simultaneously, allowing participants to choose positive leader information that fulfills both leader-focused and confirmatory search, thereby not requiring a choice between the two motives. Study 3 overcomes this limitation by requiring participants to choose between all possible pairwise comparisons between these information types. Our study design creates trials in which attraction and confirmation make either aligned (e.g., L+ vs. T+) or conflicting predictions (e.g., L- vs. T-), enabling us to test whether their effects are additive when aligned and which tendency dominates when they diverge. Trialwise confidence ratings in current preferences for options further allow a direct test of the predicted confidence dependence of confirmatory search (H3b), and the full crossing of valence and preference alignment permits a robust assessment of the positivity bias observed in Study 2 (H4).

Method

Participants

Study 3 used the same online procedure, recruitment strategy, and exclusion criteria as Study 2, but did only include the search condition because the study focused exclusively on search behavior. Participants completed the experiment online via *lab.js* (Henninger et al., 2022) and *JATOS* (Lange et al., 2015). To determine the required sample size, a bootstrap-based power analysis was conducted using data from Study 2. Specifically, in each bootstrap iteration (N = 350 resampled participants), we estimated a linear mixed-effects model predicting aggregated search probabilities from the attraction-search contrast, the confirmatory-search contrast, and their interaction, with random intercepts for participants. Power was evaluated for the attraction search main effect, the interaction between attraction and confirmatory search, and two theoretically diagnostic planned contrasts based on model-estimated marginal means: positive leader information versus all other information types, and negative leader versus negative trailer information. Across 1,000 bootstrap iterations, statistical power exceeded 80% for all target effects. Accordingly, the planned sample size for Study 3 was set to $N = 350$ participants.

A total of $N = 463$ individuals completed the study. After applying preregistered exclusion criteria (failed attention checks or implausible short completion times below 5 minutes), $N = 348$ participants remained for analysis ($M_{age} = 51.02$, $SD = 14.92$; $n = 150 = 43.1\%$ female). The sample was drawn from the German population via Bilendi's online panel and stratified by age and gender.

Materials and Procedure

Study 3 adopted the same general setup as Studies 1 and 2. As in Study 2, each decision context (i.e., hotel and dentist) again contained a pool of balanced positive and negative information that could be assigned to either option depending on participants' search choices. The key procedural change concerned the search task. Instead of ranking all four information types (L+, L–, T+, T–) as in Study 2, participants made a forced choice between two types of information. The pair presented on each trial depended on the participant's current preference, recorded at the end of the preceding trial. For any given preference state, the 2 (positive vs. negative) × 2 (leader vs. trailer) valence-by-target structure yields six possible pairwise comparisons. Across seven search rounds, each participant encountered all six comparisons at least once. The theoretically most informative comparison between negative leader (attraction-consistent) and negative trailer (confirmation-consistent) information was presented twice to increase reliability. After each preference indication, participants reported their subjective confidence on a continuous scale ranging from 50 (*absolutely a toss-up*) to 100 (*absolutely confident*), indicating how certain they were that their current leading option would ultimately become their final choice. Confidence ratings were included to test the preregistered prediction that confirmatory search increases with increasing preference certainty.

Results

Information Distortion

Replicating the information-distortion effect observed in Studies 1 and 2, information about the currently leading option was evaluated more positively than information about the trailing option. A linear mixed-effects model with random intercepts for participants and effect-coded information type (leader = +1, trailer = −1) revealed a significant main effect of information type, $b = 7.18$, $SE = 0.57$, $t(3781) = 12.62$, $p < .001$. Thus, coherence-consistent evaluation biases were again observed within the search condition in Study 3.

Separate Tests of Search Biases and Preference and Confidence-Related Factors

As expected, participants showed a reliable attraction search effect (see Figure 3, right panel): They chose leader information on 54.69% of all trials in which leader and trailer information was presented, reflected in a significant positive intercept from a logistic mixed-effects model (leader info = 1, trailer info = 0; random intercepts for participants and trials; $b = 0.27$, $SE = 0.06$, $z = 4.37$, $p < .001$), indicating that participants systematically selected information about the currently preferred option more often than the trailing option. Extending this model with standardized preference strength from the preceding trial revealed a small but significant positive association with leader-consistent search, $b = 0.14$, $SE = 0.04$, $z = 3.41$, $p < .001$, suggesting that stronger preferences were accompanied by a slightly increased likelihood of selecting information about the preferred option, replicating results from Study 2.

Unexpectedly, participants searched confirmatory information only in 44% of all trials in which confirmatory and non-confirmatory information was presented (see Figure 3, right panel) and therefore more often searched disconfirming information, reflected in a significant negative intercept from a logistic mixed-effects model (confirming info = 1, disconfirming info = 0; random intercepts for participants; $b = -0.34$, $SE = 0.05$, $z = -7.34$, $p < .001$). Exploratory analyses showed no evidence for systematic changes in confirmatory search across trials, $b = 0.01$, $SE = 0.02$, $z = 0.73$, $p = .464$, nor a reliable association with preference strength from the preceding trial, $b = -0.07$, SE = 0.04, $z = -1.53$, $p = .126$.

When confidence (z-standardized) was entered alongside preference strength in the model, higher confidence was associated with a positive but only marginal effect on confirmatory search, $b = 0.10$, SE = 0.05, $z = 1.85$, $p = .065$. Preference strength from the preceding trial showed a negative association with confirmatory selection, $b = -0.12$, SE = 0.05, $z = -2.31$, $p = .021$. Because confidence and preference strength were substantially correlated ($r = .61$, $p < .001$), preference strength was retained in the model to separate confidence-related variance from overall preference strength. Together, the results indicate that confidence and preference strength exert opposing influences on confirmatory search, with confidence showing at most a weak positive association when both predictors are considered jointly. Notably, confidence itself increased reliably across trials, $b = 0.03$, $SE = 0.01$, $t(3558) = 5.93$, $p < .001$, indicating a gradual consolidation of confidence over the decision sequence.

Beyond the overall frequency of confirmatory search, we tested whether information that jointly aligns with both attraction-based and confirmatory search tendencies (i.e., positive information about the currently leading option) was selected more often than expected by chance. Participants reliably preferred positive leader information in 53.71 % of all trials involving this type of information, consistent with the convergence of attraction-based and confirmatory search on the same information type. A logistic mixed-effects model predicting the selection of positive leader information (L+ = 1, other information types = 0), with random intercepts for participants, revealed a significant positive intercept, $b = 0.20$, SE = 0.06, $z = 3.59$, $p < .001$.

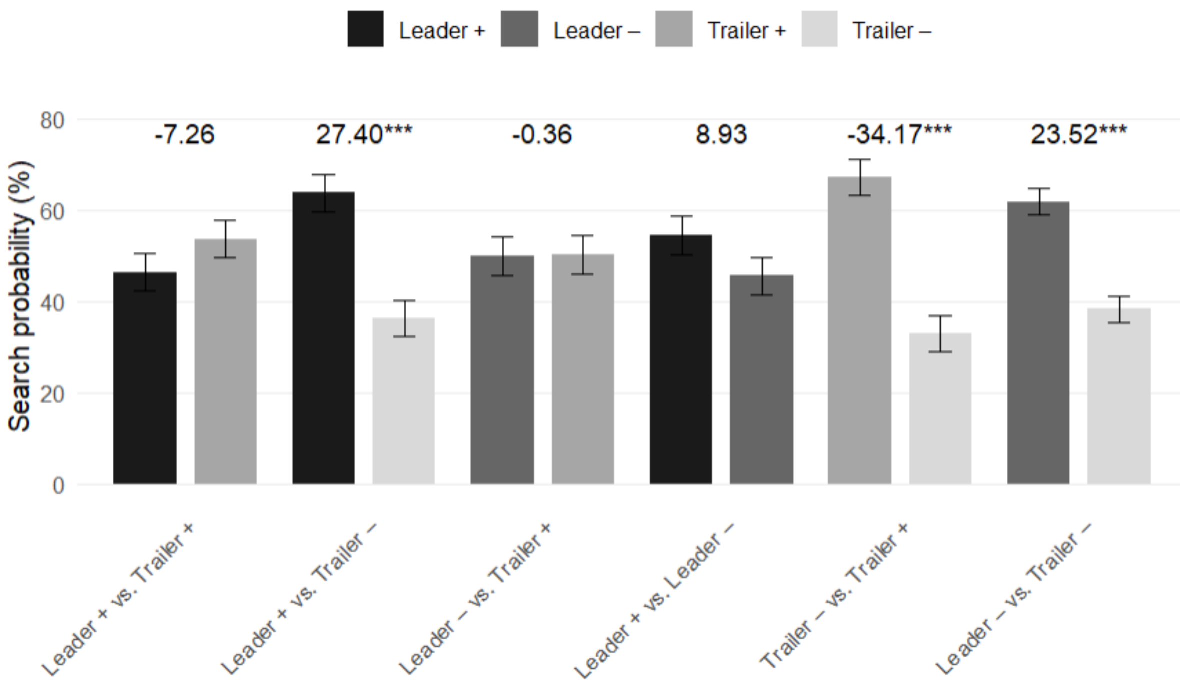
Consistent with a positivity bias, participants selected positive information in 58.8% of trials in which a positive and a negative information option were directly contrasted. This tendency was reflected in a significant positive intercept from a logistic mixed-effects model (positive = 1, negative = 0; random intercepts for participants), $b = 0.39$, $SE = 0.06$, $z = 6.90$, $p < .001$. Exploratory analyses including trial number and standardized preference strength from the preceding trial revealed no systematic change in the positivity bias across trials, $b = 0.00$, $SE = 0.02$, $z = 0.20$, $p = .84$, and no reliable association with preference strength, $b = –0.03$, $SE = 0.05$, $z = –0.55$, $p = .59$. Although these tests establish the basic empirical tendencies in search behavior, the design includes six pairwise information comparisons in which valence and option-related information combine in different ways. We therefore next examine search choices across these comparisons.

Search Patterns Across Pairwise Information Comparisons

Across the six pairwise information comparisons (Figure 5), we analyzed search choices using logistic mixed-effects models with random intercepts for participants, estimating the log-odds of selecting one information item over another while accounting for repeated observations. In the theoretically diagnostic comparison between attraction-consistent and confirmation-consistent information, participants selected negative leader information over negative trailer information on 61.76% of trials, $b = 0.54$, $SE = 0.15$, $z = 3.49$, $p < .001$, consistent with attraction-based search predictions. Participants also selected positive leader over negative trailer information on 63.70% of trials, $b = 0.60$, $SE = 0.10$, $z = 5.84$, $p < .001$. In addition, positive trailer information was preferred over negative trailer information on 67% of trials, $b = -0.87$, $SE = 0.14$, $z = -6.46$, $p < .001$, contrary to the pattern predicted by confirmatory-search accounts. No significant differences emerged for the remaining comparisons—positive leader vs. positive trailer, negative leader vs. positive trailer, and positive vs. negative leader information (all *ps* > .05). Together, these results show robust attraction- and valence-based tendencies but no consistent evidence for confirmatory search across the pairwise comparisons. Because the six comparison types reflect heterogeneous combinations of leader/trailer identity, preference consistency, and valence, leading to numerous statistical tests on different subsets of trials, we next integrate these patterns using a contrast-based decomposition aligned with the coherence model.

Figure 5.

Search probability across all pairwise comparisons of information types.



Note. The plot shows the proportion of trials in which participants selected a given information type (y-axis), separately for each pairwise comparison of information type (x-axis). Each bar reflects the relative frequency of selecting a specific type of information—Leader +, Trailer +, Leader –, or Trailer – and is coded by different shades of gray accordingly. The x-axis labels denote the specific pairwise comparison presented in each trial (e.g., Leader + vs. Trailer + for the first two bars). The overlaid values represent the percentage-point difference in selection probability between the two options in each comparison (first-listed minus second-listed information type), with significance indicated by asterisks (**p* < .05, ***p* < .01, ****p* < .001).

Integrative Contrast Analysis of Search Mechanisms

To integrate the diverse pairwise comparisons into a single model-based assessment of search, we mapped each comparison onto three theory-derived contrasts representing attraction search, confirmatory search, and positivity bias. Unlike the preregistered joint contrast provided in the Appendix, which combined attraction and confirmatory search into one contrast and was therefore not informative in the absence of confirmatory-search variation, this decomposition is less restrictive and preserves the separability of the three mechanisms implied by the coherence model. The contrasts in Table 3 map directly onto the three activation pathways specified in the sequential coherence model (Figure 1): a coherence pathway, in which activation from the preference node increases search for leader information (i.e., attraction search); a confidence pathway, in which rising activation of the confidence node selectively increases search for confirmatory information; and a valence pathway, in which activation due to the valence of information increases search for positive information.

Table 2.

Contrasts for Attraction-based Search, Confirmatory Search, and Positivity Bias for All Six Types of Information Comparisons Derived from the Coherence Model.

No.	Target Info (coded as 1)	Non-target Info (coded as 0)	Contrasts		
			Attraction Search (biased)	Confirmatory Search (biased)	Positivity Bias
1	Leader+	Trailer+	0.55 (0.5)	1.09 (1)	0
2	Leader+	Trailer−	0.55 (0.5)	0.00 (0)	1
3	Leader −	Trailer+	0.55 (0.5)	0.00 (0)	-1
4	Leader +	Leader−	-1.09 (-1)	1.09 (1)	1
5	Trailer−	Trailer+	-1.09 (-1)	1.09 (1)	-1
6	Leader−	Trailer−	0.27 (0.5)	-1.64 (-3)	0

Note. Due to unequal trial frequencies (e.g., doubled exposure for Leader– vs. Trailer-), contrast weights were adjusted for the analysis using harmonic means (Rosenthal & Rosnow, 1985). Values in parentheses indicate the unadjusted weights not used in the analysis; these make the intended contrast structure easier to see. Weights are coded in the direction of the target option, with positive values indicating that the target information is more likely to be searched for and negative values indicating search for the non-target information.

More specifically, the contrast for attraction-based search predicts that in all four comparisons that contain both leader and trailer information (i.e., no. 1, 2, 3, and 6), participants search the target information (i.e., leader information) more often than in the two comparisons where only leader or only trailer information is presented (i.e., no. 4 and 5). Similarly, the confirmatory search contrast predicts that participants are more likely to choose the target information when it is confirmatory and the non-target information is disconfirmatory (i.e., no. 1, 4, and 5) than in comparison 6, where the non-target information is confirmatory and the target information is disconfirmatory. Finally, the positivity bias contrast predicts that the target option is more likely chosen when it contains positive information and the non-target information is negative (i.e., no. 2 and 4), and that the non-target information is more likely chosen when it contains positive information and the target information is negative (i.e., no. 3 and 5). To account for unequal frequencies of the six comparison types (i.e., more frequent L– vs. T– pairings in trial no. 6), all analyses used unbiased, weighted versions of these contrasts derived from harmonic means (Rosenthal & Rosnow, 1985).

To jointly evaluate the three contrasts, we fit a logistic mixed-effects model with random intercept on participant level that included all three theory-derived contrasts, standardized preference and confidence ratings, and the interactions of the attraction and confirmatory search contrasts with both ratings. The attraction search contrast significantly predicted search choices, $b = 0.14$, $SE = 0.05$, $z = 2.59$, $p = .010$, indicating a reliable tendency to select information about the currently leading option. This effect increased with preference strength, $b = 0.20$, $SE = 0.07$, $z = 2.97$, $p = .003$, but decreased with confidence, $b = -0.15$, $SE = 0.07$, $z = -2.20$, $p = .028$. In contrast, the confirmatory-search contrast showed a robust main effect in the opposite direction, $b = -0.22$, $SE = 0.03$, $z = -6.63$, $p < .001$, and was not reliably moderated by confidence or preference strength in the integrative model (both *ps* > .41). Thus, when modeled jointly, attraction-based search accounted for a larger share of systematic variance in search behavior.

Because the attraction-based and confirmatory-search contrasts were not orthogonal, $r = -.48$, $p < .001$, and yielded partially overlapping predictions—most notably converging on increased selection of positive leader information (L+)—we conducted complementary follow-up models examining each mechanism in isolation. In the attraction-only model, attraction-based search remained highly robust, $b = 0.31$, $SE = 0.05$, $z = 6.53$, $p < .001$, again increasing with preference strength, $b = 0.23$, $SE = 0.06$, $z = 3.78$, $p < .001$, and decreasing with confidence, $b = -0.18$, $SE = 0.06$, $z = -2.90$, $p = .004$. In a separate confirmatory-search model, confirmatory search also showed a strong negative main effect, $b = -0.26$, $SE = 0.03$, $z = -8.89$, $p < .001$, but was positively associated with confidence, $b = 0.07$, $SE = 0.04$, $z = 2.02$, $p = .043$, and negatively associated with preference strength, $b = -0.09$, $SE = 0.04$, $z = -2.46$, $p = .014$. Notably, the two mechanisms exhibited opposing associations with confidence and preference strength: Attraction-based search increased with stronger preferences but decreased with confidence, whereas confirmatory search showed the reverse pattern. Taken together, the results indicate that biased information search is predominantly governed by a coherence-based attraction mechanism, whereas confirmatory search reflects a more selective, confidence-dependent modulation that becomes apparent only when examined in isolation.

Discussion

Study 3 provided a stringent test of information search when outcome valence was explicit by forcing direct choices between competing information types. Replicating the previous studies, attraction search (H2) emerged robustly: participants preferentially selected information about the currently leading option, and this tendency increased with preference strength. These findings are consistent with coherence-based accounts in which stronger preference activation biases search toward information associated with the favored option.

In contrast, no general confirmatory-search tendency was observed (H3a). Across trials in which confirmatory and disconfirmatory information competed directly, participants more often selected disconfirming information. Thus, confirmatory search did not manifest as a standalone bias in this paradigm.

Exploratory analyses nevertheless yielded a fragile indication that confirmatory influences may depend on subjective confidence (H3b). When confidence and preference strength were modeled jointly, confidence showed a numerically positive but only marginal association with confirmatory search, whereas preference strength was negatively related. In contrast, in exploratory contrast-based models that did not simultaneously include attraction-based search, confidence reliably predicted increased confirmatory selection. This pattern suggests that confidence-related confirmatory influences, if present, are comparatively small and partially masked when attraction-based variance is modeled concurrently.

Importantly, the coherence model (Figure 1) makes a more specific prediction about where confirmatory influences should be most apparent if they are operative. Rather than producing a global confirmatory bias, confirmatory and attraction-based processes converge on positive leader information. Consistent with this convergence—and replicating Study 2—positive leader information was selected more frequently than any other information type, suggesting that biased search is strongest when multiple mechanisms favor the same option.

Taken together, Study 3 shows that attraction-based search dominates biased information acquisition, whereas confirmatory influences appear weaker and conditional. Rather than operating as an independent driver of search, confirmatory tendencies manifest most clearly in conjunction with attraction-based processes, amplifying search for information on which both mechanisms converge.

General Discussion

The present studies aimed to clarify how information distortion and biased information search jointly emerge in sequential decision making by testing the key components of a sequential coherence model. Across three preregistered experiments, we examined how information distortion and biased information search arise within a sequential coherence framework. Consistent with H1a, information supporting the currently preferred option was evaluated more positively than information supporting the trailing option in all studies, replicating the information distortion effect. Study 2 further provided first evidence for H1b: distortion was amplified under active search, consistent with the assumption that search heightens preference awareness and thereby strengthens coherence-driven updating.

Attraction search (H2) emerged robustly across all studies. Participants preferentially searched information about their currently favored option when valence information was concealed (Study 1), fully visible (Study 2), or directly contrasted with competing information in forced-choice trials (Study 3). In Studies 2 and 3, this tendency increased with preference strength, aligning with the model prediction that larger preference asymmetries magnify coherence pressures. Across studies, attraction search showed medium-sized effects (*ds* ≈ .27–.48), comparable to those reported in earlier demonstrations of attraction search, despite substantial differences in task structure and measurement approach (Jekel et al., 2018; Scharf et al., 2019).

Confirmatory search (H3a) exhibited a selective pattern across studies. Whereas Study 2 provided evidence for an overall confirmatory tendency, Study 3 did not reveal a general bias toward confirmatory information. Instead, confirmatory influences were restricted to specific informational contexts. Participants preferentially selected positive leader information—where confirmatory and attraction-based predictions aligned—while consistently avoiding negative trailer information, despite its confirmatory status. Exploratory contrast-based analyses further showed that confirmatory tendencies increased with decision confidence (H3b) when attraction-based variance was not modeled concurrently, suggesting that confidence-related confirmatory influences are comparatively weak and partly obscured when attraction-based processes are accounted for.

Across Studies 2 and 3, we also observed a robust positivity bias (H4): participants systematically preferred positive over negative cues, including positive trailer information that was neither coherent nor confirmatory. Although this effect is captured in the network architecture as an additional valence node (see Figure 1), its theoretical grounding remains tentative. The valence component was introduced post hoc in Study 2 and tested a priori in Study 3 to accommodate the empirical pattern and may reflect an affective influence that extends beyond the core coherence dynamics. To clarify the origins and implications of this positivity bias, we next turn to theoretical accounts that may shed further light on this pattern.

Taken together, the results point to two aspects of biased information search that warrant closer theoretical scrutiny: the conditional and comparatively weak expression of confirmatory search, and a robust positivity bias that extends beyond coherence-based predictions. We address these two patterns in separate sections below.

Theoretical Accounts of Predecisional Confirmatory Search

The present findings align with a growing body of work suggesting that confirmatory search is not a pervasive feature of predecisional information acquisition, even when preferences evolve dynamically over time. Across studies, confirmatory tendencies were weak or absent at the aggregate level, despite continuous trial-by-trial tracking of emerging preferences in consumer choice contexts.

This pattern is broadly consistent with prior findings using related paradigms. Carlson and Guha (2011) documented a robust tendency to focus search on the emerging leader, reflecting attraction-based rather than confirmatory search. In contrast, Chaxel et al. (2013) reported confirmatory information acquisition primarily in postdecisional settings, whereas predecisional search did not show a consistent confirmatory pattern. Similarly, Mischkowski et al. (2021) observed predominantly disconfirmatory predecisional search despite strong coherence effects in judgment. Together, these findings indicate that predecisional information search often tracks emerging option dominance rather than selectively avoiding disconfirming evidence.

The present findings provide at most tentative support for the idea that confirmatory influences in information search depend on subjective confidence. Across studies, no general confirmatory-search tendency emerged. Exploratory analyses suggested that confirmatory selection increased with confidence only under restricted modeling conditions—specifically, when attraction-based variance was not modeled simultaneously—indicating that any confidence-related confirmatory effects are small and easily masked by stronger attraction-based processes.

This pattern is consistent with cognitive-economy accounts of confirmatory information processing, which propose that rising subjective certainty reduces the perceived utility of further information search and promotes more selective, preference-consistent sampling (Fischer, 2011). Relatedly, dual-process models such as the Heuristic–Systematic Model predict that increasing confidence narrows information processing once the perceived confidence gap is sufficiently reduced (Chaiken, 1980; Chen et al., 1999).

At the same time, features of the present paradigm likely constrained the expression of confidence-dependent confirmatory search. Pairwise forced-choice trials sometimes required the selection of disconfirming information even under high confidence, and relatively short search sequences may have limited the emergence of strongly consolidated certainty states. Accordingly, the absence of a robust confirmatory bias should be interpreted as reflecting boundary conditions of the present design rather than as evidence against confidence-based confirmatory mechanisms per se.

More broadly, the present results speak to general issues of theory specification in the study of biased information search (Glöckner & Fiedler, 2024). Confirmatory search is often defined in broad terms, such as the avoidance of disconfirming evidence, which leaves open how it should manifest in concrete choice situations. The pattern observed here—selective prioritization of positive leader information rather than uniform avoidance of disconfirming information—aligns more closely with the logic of positive hypothesis testing, whereby individuals preferentially sample information expected to confirm a focal hypothesis, even when disconfirming information would be more diagnostic (Klayman & Ha, 1987, 1989; Navarro & Perfors, 2011). Future research should therefore move beyond coarse dichotomies of confirmatory versus disconfirmatory search and instead develop more precise, process-level predictions that distinguish between attraction-based mechanisms, confidence-dependent selectivity, and hypothesis-testing strategies.

Having delineated the scope and limitations of predecisional confirmatory search, we now turn to a second, conceptually distinct pattern observed in the present studies: the systematic preference for positive information.

Theoretical Accounts of the Positivity Bias

Several broader perspectives offer potential explanations for why positive information was consistently prioritized—even when it neither pertained to the leader option nor was confirmatory. Affective accounts highlight that positive cues are often more fluent, less conflict-inducing, and processed with greater ease than negative cues (Knobloch-Westerwick & Jingbo Meng, 2009; Unkelbach & Greifeneder, 2013). Regulatory-focus theory similarly predicts a preferential engagement with gain-related information under promotion-oriented goals (Crowe & Higgins, 1997; Idson et al., 2000), which aligns with the advancement-oriented structure of the present task. Learning-based accounts provide another source of asymmetry: empirical and computational work suggests that people exhibit higher learning rates for positive than for negative feedback (Palminteri & Lebreton, 2022), which may render positive information subjectively more diagnostic or rewarding. From an information-theoretic standpoint, Optimal Experimental Design models argue that individuals often value cues they (accurately or not) perceive as improving choice accuracy (Nelson et al., 2010), potentially inflating the perceived utility of positive trailer information. Finally, discovery-oriented frameworks propose that information search sometimes serves to reduce ambiguity rather than confirm existing beliefs (Spektor & Wulff, 2024). Within such accounts, the appeal of positive trailer information may reflect an exploratory tendency to sample pleasant or non-threatening information, thereby broadening the representational space of the decision. Together, these perspectives suggest that the positivity bias observed across Studies 2 and 3 potentially reflects an interplay of affective fluency, perceived utility, and exploratory motivation—processes that extend beyond the structural and motivational dynamics formalized in the coherence model.

Boundary Conditions and Theoretical Integration

The present findings also clarify several boundary conditions of the sequential coherence account. First, the reliable positivity bias observed in Studies 2 and 3 indicates that valence exerts an influence on search. Because this component was added post hoc and lacks a principled derivation from coherence dynamics, it should be viewed as an additional process that operates alongside, rather than within, the structural and motivational mechanisms formalized in the model. Second, all studies employed binary choice environments. Although parallel constraint satisfaction models can be extended to multi-option decisions, the dynamics observed here—especially the convergence of coherence and confidence on positive leader information—may behave differently when several options compete for activation. Whether the same interaction between coherence, confidence, and valence generalizes to richer choice sets remains an open question. Third, the predicted effect of preference awareness on information distortion (H1b) emerged in Study 2 but not in Study 1. This difference is consistent with the task structure: repeated ranking of all options in Study 2 likely heightened and sustained awareness of emerging preferences, thereby strengthening coherence-driven distortion. This interpretation aligns with extensions of the iCodes framework, which attribute stronger preference effects under active engagement to increased preference awareness (Scharf et al., 2022). The absence of H1b in Study 1 suggests that weaker or less sustained manipulations of preference awareness may be insufficient to reliably modulate information distortion.

These boundary conditions allow a cautious connection to coherence-based theories of biased reasoning. Simon and Read (2025) argue that confirmation bias need not reflect a distinct or ubiquitous motivational tendency but can emerge as a byproduct of coherence-maximizing dynamics when a belief state becomes sufficiently stabilized and functions as an attractor. From this perspective, confirmation-like search should not be expected as a general predecisional bias, but rather as a contingent pattern that depends on the strength and stability of the underlying coherence structure.

The present findings are broadly compatible with this account. Attraction-based search emerged as a robust coherence-driven tendency, whereas confirmatory search did not appear as a stable or pervasive pattern. Exploratory analyses yielded at most weak and model-dependent indications that confirmation-like search may increase under conditions of elevated subjective certainty. Importantly, these effects were neither consistent nor robust across modeling approaches, underscoring that confirmatory influences—if present—are fragile and easily overshadowed by attraction-based processes.

Rather than providing evidence for confidence-driven confirmatory search, the present results therefore delineate boundary conditions under which confirmation-like patterns may fail to emerge. Future research should test coherence-based predictions more directly by experimentally manipulating the stability of belief states—for example, by inducing strongly diagnostic evidence, systematically varying pre- versus post decisional contexts, or explicitly strengthening subjective certainty—to examine whether and when coherence-driven attractor dynamics give rise to selective confirmation-like information search.

Together, these considerations highlight how structural coherence, confidence, and valence jointly shape sequential information acquisition, while clarifying the scope of the present model and motivating future work toward a more integrative account of biased reasoning.

Constraints on Generality

In line with recent calls to include explicit *Constraints on Generality* (COG) statements in empirical reports (Simons et al., 2017), it is important to note that all stimulus materials in the present work were pretested within German-speaking, WEIRD samples, and that all reported studies were likewise conducted in such populations. In the present studies, we took care to balance gender and to approximate the age distribution of the German population. Moreover, search behavior was examined exclusively within health- and consumer-related domains. Although exploratory analyses revealed no meaningful differences between these two domains, no domain-independent conclusions can be drawn. As such, while the mechanisms discussed—positivity bias, probability gain, and discovery-driven search—may extend beyond the tested context, such claims remain to be established through targeted replications in more diverse populations, cultural settings, and decision domains.

Conclusion

Across three preregistered studies, the present research provides a process-level account of how information distortion and biased information search unfold in sequential decision making. Attraction-based search emerged as a robust tendency across paradigms, whereas confirmatory search was weak, selective, and contingent on specific conditions rather than a general predecisional bias. In addition, a robust positivity bias shaped information selection independently of emerging preferences. Together, these findings indicate that biased information search reflects the interplay of multiple mechanisms, with attraction-based processes exerting a dominant influence. By integrating information distortion and biased search within a unified sequential framework, the present work refines coherence-based theories of decision making and clarifies how emerging preferences guide information acquisition over time.

Appendix 6A: Pre-registered Contrast for Study 3

The preregistered joint contrast did not adequately capture the observed search pattern including (a) the absence of a general confirmatory-search tendency, (b) the strong positivity asymmetry, or (c) the confidence dependence of confirmatory search. For transparency, we therefore report the preregistered analysis here and present the primary analyses in the main text using separate predictors for attraction search and confirmatory search.

Rank-Ordered Search Preferences

To integrate search behavior across all six pairwise information comparisons, we fit the preregistered contrast model specified for Study 3. The model assigned theory-derived contrast weights to each pairing, reflecting the predicted relative support for selecting the target information based on attraction-based (leader vs. trailer) and confirmatory (preference-consistent vs. inconsistent) considerations (Table 3). Pairings that represented the same combination of attraction and confirmaory components were assigned identical weights.

Because the six pair types occurred with unequal frequency, we implemented unbiased contrasts following Rosenthal and Rosnow (1985), using harmonic mean cell sizes and rescaling the contrast vectors to preserve their sum of squares. Given consistent evidence for a general preference for positive over negative information, the model additionally included a valence-asymmetry covariate coded as +1 (target positive, comparison negative), 0 (equal valence), or –1 (target negative, comparison positive).

A generalized linear mixed-effects model with random intercepts for participants and trials tested whether the probability of selecting the target cue increased with the global contrast values. Contrary to the preregistered prediction, the contrast term was significantly negative, b =

–0.06, *SE* = 0.02, *z* = –2.41, *p* = .016. In contrast, the valence-asymmetry covariate showed the expected positive effect, *b* = 0.36, *SE* = 0.04, *z* = 8.33, *p* < .001.

Table 3.

Preregistered Contrasts for the Attraction Search (AS), Confirmatory Search (CS) and the Positivity Bias.

No.	F1: Target Info (coded as 1 when selected)	F2: Competitor Info (coded as 0 when selected)	F1 – F2	Contrast AS and CS	Contrast Positivity Bias
1	L+ = AS + CS	T+ = 0	AS + CS	1.87 (2)	0
2	L+ = AS + CS	T– = CS	AS	0.94 (1)	1
3	L– = AS	T+ = 0	AS	0.94 (1)	-1
4	L+ = AS + CS	L– = AS	CS	-0.94 (-1)	1
5	T– = CS	T+ = 0	CS	-0.94 (-1)	-1
6	L– = AS	T– = CS	AS - CS	-1.74 (-2)	0

Note. Each column represents a planned contrast tested in the model predicting the likelihood of selecting the target information. Positivity bias is coded as +1 (target positive, comparison negative), 0 (equal valence), or –1 (target negative, comparison positive). Contrast of ASE and CS is based on the difference (column F1 – F2) between the two biases for each comparison (columns F1 and F2) between target and competitor information. The contrast for ASE and CS imply for search that AS + CS > AS > CS > AS – CS.

Appendix 6B: Open Science Framework (OSF) Resources

Open Science Framework (OSF) Study-Specific Links

The project overview is accessible via the Open Science Framework (OSF): https://osf.io/58rw2/overview?view_only=e3462bad156f4046826a89f0ebee4d90.

The repository is organized into three study-specific components ("Study 1," "Study 2," and "Study 3"), each containing the full set of experimental materials, data files, analysis code, and the corresponding power analysis.

Preregistrations

Study 1: https://osf.io/cazm8/?view_only=b071a88ff4ce465eae02c3775602c0b4

Study 2: https://osf.io/y3t6c/?view_only=7ed0b5bf2ee347938cccbf21d72ec4af

Study 3: https://osf.io/yugrz/?view_only=e76035f489164a4e85e0442bdde4580a

Materials

Pre-Study 1: https://osf.io/xpnqs/?view_only=1e3e91d19f5f4b89a0ec7f89ed151343

Pre-Study 2: https://osf.io/f9rn5/?view_only=279620ab5ac0456394e7f01c8d755498

Names Pretest: https://osf.io/fm295/?view_only=1faba23cf110478a836f1d949f46ea20

6.3 Summary and Implications

This chapter examined biased information search within a fully sequential decision framework, extending coherence-based accounts from information evaluation to active information acquisition. Addressing the limitations identified in Section 3.2.3, it investigated how attraction search and confirmatory search operate when evaluation and search co-evolve over time.

Across three preregistered studies, attraction search emerged as a robust and general tendency: once an option gained a preference advantage, additional information about that option was preferentially sampled, independent of whether it was preference-confirming or preference-disconfirming. This pattern aligns with the core prediction of coherence-based models that activation asymmetries systematically bias information acquisition as coherence develops.

In contrast, confirmatory search did not manifest as a pervasive predecisional bias. Confirmatory tendencies were selective and context-dependent, with only limited evidence that subjective confidence reliably amplified preference-consistent selection. Rather than constituting a dominant mechanism during early preference formation, confirmatory selectivity appears to operate under more restricted psychological or informational conditions.

Beyond these mechanisms, a consistent positivity bias shaped information acquisition independently of both attraction-based and confirmatory influences. Although not directly derived from the core coherence architecture, this effect underscores that sequential search reflects the interplay of coherence-driven dynamics with additional evaluative asymmetries.

Taken together, the findings indicate that attraction search and confirmatory search are not competing explanatory accounts but distinct influences operating within the same sequential coherence process. By embedding information acquisition within the architecture that also governs evaluative distortion and belief-based constraint, Chapter 6 broadens the empirical scope of coherence-based models in sequential decision making.

The General Discussion integrates these findings with those of Chapters 4 and 5 to clarify their joint theoretical implications for coherence-based accounts of judgment and decision making.

Chapter 7: General Discussion

7.1 Summary of the Main Findings

The present dissertation examined how coherence-seeking processes shape decision making when information is encountered sequentially. Across three empirical projects, it investigated how emerging preferences interact with incoming evidence, stable beliefs, and information acquisition as decisions unfold over time. Rather than focusing on isolated bias effects, the dissertation examined the process-level dynamics implied by coherence-based models of reasoning in sequential environments.

Chapter 4 addressed a foundational issue in the empirical study of predecisional information distortion: the incomplete translation of coherence-based process assumptions into statistical predictors. Although conceptual and connectionist models assume continuous bidirectional updating between preferences and information evaluation, empirical tests in the SEP paradigm have typically relied on serial predictors that condition distortion on prior preference states. By deriving and testing model-implied predictors at the trial level, Chapter 4 showed that key coherence dynamics—such as amplification, attenuation, and reversals of distortion as preferences change—remain systematically underestimated under standard analytic practices. These findings demonstrate how misalignment between process models and statistical operationalization can obscure theoretically central dynamics even in well-established paradigms.

Chapter 5 extended the sequential coherence framework to political judgment by examining how stable ideological beliefs interact with dynamically evolving preferences during information integration. The results indicate that partisan motivated reasoning does not operate as a fixed belief-imposed distortion. Instead, ideological priors function as higher-order constraints within the same coherence-driven updating process that governs evaluative distortion. Preference dynamics and belief-based categorization influenced one another bidirectionally: emerging preferences shaped both evaluation and the perceived ideological meaning of information, while partisan cues amplified these effects by strengthening belief-consistent constraints. Political bias thus emerged as a graded outcome of coherence-based integration rather than as a static endpoint of deliberation.

Chapter 6 further extended coherence-based models from information evaluation to active information search. Building on the iCodes framework, the chapter examined how attraction-based and confirmatory search unfold alongside evolving preferences and decision confidence in sequential decision tasks. Across three preregistered studies, attraction search emerged as a robust and general tendency, indicating that coherence dynamics systematically bias which information is sampled as preferences develop. Confirmatory search, by contrast, appeared as a weaker and context-dependent influence, with only limited evidence that subjective confidence reliably amplified preference-consistent selection. In addition, a consistent positivity bias shaped information acquisition independently of both attraction-based and confirmatory tendencies.

Taken together, the three projects converge on a central conclusion: coherence-based reasoning provides a unifying process-level account of biased judgment and decision making in sequential environments. Across evaluation, belief formation, and information search, biases emerged as dynamic consequences of parallel constraint satisfaction operating over time, rather than as fixed tendencies or categorical failures of evidence integration. By specifying and empirically testing the dynamic implications of coherence-based models across domains, the dissertation advances a process-oriented understanding of biased reasoning and search that highlights how preferences, beliefs, and selective information search jointly shape decisions as they unfold.

7.2 Information Distortion under Varying Constraint Structures

To illustrate how the coherence dynamics examined in the preceding chapters manifest across different sequential decision contexts, Table 1 summarizes descriptive indicators of information distortion across the paradigms investigated in this dissertation. Two complementary indicators are reported. The prevalence of hypothesis-consistent distortion reflects the proportion of trials in which preference-consistent information was evaluated more positively than preference-inconsistent information. Distortion magnitude reflects the mean deviation of information evaluations from a control-group baseline, expressed as a proportion of the maximum possible deviation on the rating scale (Russo, 2015), corresponding to the standard operationalization used in the stepwise evolution-of-preference literature.

Viewed across paradigms, the pattern summarized in Table 1 illustrates how the observable magnitude of information distortion varies with the informational and structural characteristics of the decision environment. Under baseline sequential integration conditions—such as those examined in Chapter 4—distortion effects are comparatively small.

Across all studies, distortion magnitudes range from approximately 0.9% to 11% of the maximum possible deviation. Relative to prior SEP research, which typically reports values between approximately 5% and 40% (Russo, 2015; see also Russo et al., 1996, 2000; Russo & Yong, 2011), many of the observed effects—particularly in the baseline and weak-constraint conditions—fall at the lower bound of this range or even below it. At the same time, distortion levels in the information search paradigm (Chapter 6) reach values that overlap with the lower bound of previously reported effects.

However, comparisons of absolute effect sizes across studies should be interpreted cautiously because distortion magnitude partly depends on the measurement scale used to assess evaluations. Many SEP studies summarized by Russo relied on relatively coarse 7- or 9-point rating scales, for which a one-point shift already constitutes a substantial proportion of the maximum possible deviation from the scale midpoint. In contrast, the present studies employed a fine-grained −100 to +100 evaluation scale, allowing smaller incremental changes in evaluations to be expressed. Consequently, lower proportional magnitudes may partly reflect differences in measurement rather than weaker coherence-driven effects.

Table 1.

Information Distortion Metrics across Sequential Decision Contexts and Constraint Structures.[7]

Study context	Constraint structure	*k*	*N*	Prevalence of Hypothesis-Consistent Information Distortion (% of Trials)	Hypothesis-Consistent Information Distortion (Mean % of Maximum Possible Deviation)
		studies	trials	range	range (mean)
Paradigm Blindness (Chapter 4)	Baseline coherence	4	26008	51.3–54.1	2.6–5.9 (4.2)
Political decision making (Chapter 5)	Baseline coherence	1	3877	52.9	0.9 (0.9)
	Weak belief constraint	2	5710	54.3–58.3	2.1–5.1 (3.6)
	Competing constraints	1	1859	56.8	5.0 (5.0)
	Strong belief constraint	2	6089	54.7–61.9	3.9–7.7(5.8)
Information search (Chapter 6)	Baseline coherence	2	7053	56.5–57.7	8.5–11.6 (10.0)
	Search constraints	3	8804	59.1–61.7	7.3–11.1 (9.0)

Note. The term *baseline coherence* refers to conditions in which no additional constraints were introduced into the coherence network beyond those inherent to the sequential evaluation task. In Chapter 4 (Paradigm Blindness), this corresponds to the standard stepwise evolution-of-

[7] The distortion estimates reported here are based on conventional operationalizations used in the SEP literature, which model distortion as a function of prior preference states. As shown in Chapter 4, this approach does not fully capture the parallel, bidirectional updating dynamics implied by coherence-based process models. Analyses based on model-implied predictors yield systematically different estimates of distortion. For comparisons between conventional and model-implied estimates, see Chapter 4.1, Table 1.

preference paradigm without additional belief or search manipulations. In Chapter 5 (Political decision making), baseline coherence refers to conditions in which no ideological cues were provided (i.e. same party condition). In Chapter 6 (Information search), it denotes passive information presentation without active information search. Additional constraint structures varied across projects. In the political judgment paradigm, weak belief constraints arise when partisan positions must be inferred from contextual cues. Strong belief constraints occur when political candidates are explicitly labeled by party affiliation. In these conditions, statement content may either align with the labeled source or vary independently of it. Competing constraints refer to conditions in which explicit partisan labels are provided but the ideological content of the statements conflicts with the labeled source, creating tension between source-based and content-based cues. In the information search studies, search constraints refer to conditions in which participants actively choose which information to acquire next.

The political decision studies reported in Chapter 5 illustrate how the informational structure of the environment can further shape these dynamics. When ideological affiliations could only be inferred indirectly from statements, distortion levels remained moderate. When partisan labels were made explicit, however, distortion effects increased systematically. This pattern is consistent with the interpretation advanced in Chapter 5 that ideological beliefs function as higher-order constraints within the coherence process: stronger belief cues increase the activation of belief-consistent links in the decision network, thereby amplifying the influence of emerging preferences on the evaluation of incoming information.

A somewhat different pattern emerges in the information search paradigm examined in Chapter 6. Distortion levels observed in these studies are generally higher than those in the sequential integration paradigms. One possible interpretation is consistent with the theoretical relation between preference awareness and coherence dynamics derived from the iCodes framework. Scharf et al. (2022) showed that attraction-based search becomes stronger when participants indicate their current preference before selecting information, which they attributed to increased preference awareness strengthening the influence of preferences on search via the γ-parameter of the model. Building on this logic, Chapter 6 derived the prediction that increased preference awareness during active information acquisition may also amplify the influence of preferences on evaluation.

The present studies did not reveal a clear difference between active and passive search conditions in terms of distortion magnitude. Nevertheless, the overall distortion levels observed in the search paradigm are broadly compatible with this theoretical implication. One possible explanation is related to the structure of the control condition: participants in the passive condition were presented with the same information previously selected by matched participants in the active condition. Consequently, the informational environment itself remained partially shaped by preference-guided search processes. Under such conditions, even participants who did not actively choose information may have encountered an information sequence that reflected preference-guided sampling, potentially reducing differences between active and passive search conditions.

Taken together, the descriptive synthesis presented in Table 1 highlights a broader implication of coherence-based models of reasoning. Observable biases in evaluation do not appear as fixed properties of a single experimental paradigm but vary systematically with the informational and structural characteristics of the sequential decision environment. Differences between paradigms in how information is presented, structured, and acquired shape the dynamics of coherence formation, and these differences are reflected in corresponding variation in evaluative bias. In this sense, variation in distortion across paradigms can be understood not merely as methodological noise but as a theoretically meaningful consequence of how sequential decision environments structure the process of coherence formation.

7.3 Theoretical Integration and Contributions

The present dissertation makes two major interrelated theoretical contributions. First, it advances theory construction in judgment and decision making by addressing a systematic bottleneck between conceptual process models and statistical testing. Building on recent debates about theory specification and formalization, it proposes and exemplifies an intermediate step that strengthens the translation from conceptual models to statistical tests without requiring full formalization.

Second, the dissertation contributes to the integration of reasoning biases by extending coherence-based models to new domains and dependent variables. Across Chapters 5 and 6, it shows how seemingly distinct biases—such as partisan motivated reasoning, information distortion, attraction search, and confirmatory search—can be understood as context-dependent manifestations of a common coherence-seeking process operating in sequential decision environments.

7.3.1 Contribution to Theory Construction

One central theoretical contribution of this dissertation lies in identifying a specific failure mode in psychological theorizing: the incomplete translation of conceptual models into statistical tests. This failure concerns neither the absence of theory nor insufficient formal rigor per se, but a breakdown at a distinct stage of theory development—namely, the step from conceptual (or verbal) assumptions to precise hypotheses and statistical models.

This contribution is strongly related to recent debates about theory construction and diagnoses of a "theory crisis" in psychology, which converge on the observation that empirical findings often accumulate without corresponding gains in theoretical constraint, integration, or explanatory power (Borsboom et al., 2021; Eronen & Bringmann, 2021; Gigerenzer, 2010; Oberauer & Lewandowsky, 2019). Importantly, several authors emphasize that this problem persists even when theories are explicitly invoked, because theoretical assumptions are frequently too weakly specified to strongly constrain hypothesis formulation and statistical testing (Fiedler, 2017; Fiedler & Trafimow, 2024; Kruglanski, 2001).

In response, substantial effort has been devoted to developing methods for formal theory specification, including computational modeling (e.g., Farrell & Lewandowsky, 2010), equation-based frameworks such as Theory Construction Methodology (Borsboom et al., 2021), network approaches (e.g., Dalege et al., 2016), agent-based simulations (e.g., Jackson et al., 2017), and ontology-based approaches (e.g., Hale et al., 2020; Leising et al., 2023; for a review, see Glöckner & Fiedler, 2025). These approaches substantially increase theoretical precision and are indispensable for mature theory development.

At the same time, it has been repeatedly noted that many formal specification methods are technically demanding and require specialized training. Past attempts at highly rigorous formal reconstructions—such as structuralist approaches to classic psychological theories (e.g., Moulines, 1996; Westermann, 2000)—achieved impressive formal precision but have been criticized for their complexity and limited accessibility. In particular, structuralist reconstructions have been described as too complex to be readily accessible to a broader range of researchers (Westmeyer, 2004). More recent discussions have reiterated this concern and emphasized the importance of accessibility for successful theory specification (Glöckner & Fiedler, 2025b; Koch & Wendt, 2025).

The present dissertation explicitly endorses this accessibility-oriented perspective and aligns closely with recent proposals by Glöckner and Fiedler (2025b), who emphasize verbal theory specification through explicit propositions as a viable and necessary component of theory development, as reflected in the Proposition-Based Theory Specification Method (PBTS). Crucially, the contribution advanced here does not compete with this approach. Rather, it addresses a complementary bottleneck that arises *after* verbal specification: how verbally articulated process assumptions can and should constrain empirical testing.

To clarify the contribution, the dissertation proposes a three-milestone framework of theory development: (1) establishing a robust phenomenon, (2) articulating a conceptual or proto-theoretical model, and (3) translating this model into hypotheses and statistical tests. The first two milestones closely follow Borsboom et al.'s (2021) theory construction methodology, in which Step 2 explicitly involves the formulation of a verbally articulated proto-theory—a set of general principles specifying how the phenomenon would arise if the theory were true (Haig, 2005, 2014).

In line with this usage, we refer to such verbally specified explanatory frameworks as *conceptual models*, and to *conceptual process models* when these frameworks additionally specify assumptions about the generative and temporal structure of the phenomenon (Borsboom et al., 2021). Importantly, recent proposals for verbal theory specification do not require theories to be process models in this sense. For example, Glöckner and Fiedler (2025) explicitly emphasize the value of making theoretical assumptions explicit at the verbal level in the first step, without necessarily committing to assumptions about underlying cognitive processes.

The present dissertation adopts this minimal requirement for theory specification but takes a further step by focusing on cases in which a conceptual model *does* make process-related commitments. As argued by Glöckner and Betsch (2011), process models have the potential to increase the empirical content of a theory by generating predictions not only about choices or judgments, but also about additional dependent variables such as time, confidence (Glöckner et al., 2014; Glöckner & Betsch, 2008a) and information search (Jekel et al., 2018; Payne et al., 1988). In this sense, process models are best understood as an optional extension of verbal theories that can increase their empirical precision.

The empirical case studies examined in this dissertation are based on such a conceptual process model. Coherence-based accounts do not merely predict decision outcomes, but posit bidirectional and temporally local interactions between preferences, beliefs, evaluations, and information acquisition. Once such process assumptions are articulated—whether verbally or formally—they necessarily imply constraints on admissible hypotheses and statistical models.

The central contribution of the present dissertation lies in making these constraints explicit. Milestone 3—the translation of conceptual process models into statistical tests—does not presuppose computational formalization, nor does it extend the scope of verbal theory specification. Rather, it addresses a subsequent and logically distinct question: how verbal propositions, which may also involve process assumptions, once posited, can be empirically tested in a way that is diagnostic of the theory's core commitments.

While the first two milestones are frequently achieved in psychological research, the third often remains underspecified. Instead, hypothesis formulation and statistical modeling can be guided by paradigm-specific conventions that reflect procedural features of the task rather than the theory's process assumptions. This systematic mismatch—referred to here as *paradigm* *blindness*—has direct implications for theory evaluation. When statistical tests fail to instantiate the propositions of a conceptual (process) model, empirical results cannot meaningfully inform theory refinement or adjudication, even if effects replicate robustly. Under these conditions, empirical success does not translate into theoretical progress.

The contribution of the present dissertation is not only conceptual but demonstrative. Chapter 4 provides the core illustration of the bottleneck by showing that standard analyses of information distortion test only a restricted subset of the predictions implied by coherence-based process models. By conditioning distortion on prior preference states, these analyses implicitly treat preferences as fixed inputs and thereby fail to instantiate the bidirectional updating dynamics assumed by the model. By translating process assumptions into statistical predictors, Chapter 4 reveals theoretically central dynamics—such as amplification, attenuation, and reversal of distortion—that are systematically underestimated when Milestone 3 is left incomplete.

Chapters 5 and 6 consequently apply the same three milestone framework to substantively different domains and processes. In Chapter 5, research on motivated reasoning in political judgment is shown to rely predominantly on unidirectional tests from prior beliefs to distorted evaluations. By contrast, coherence-based process models imply bidirectional relations in which emerging preferences also shape belief-based categorization. By proposing an extended coherence-based model of political reasoning and explicitly testing its implications, Chapter 5 demonstrates that emerging preferences do not only affect the evaluation on new information but also systematically influence the perceived ideological meaning of information, while the activation of a priori partisan beliefs modulates the strength of those effects rather than imposing a fixed belief-driven distortion.

Chapter 6 extends the milestone logic to active information search. Rather than treating confirmatory search as a static preference for supportive information, it derives predictions about how emerging preferences and confidence jointly constrain information acquisition over time. By translating these assumptions into statistical tests, Chapter 6 shows that attraction-based search emerges as a robust consequence of coherence dynamics, whereas confirmatory search appears only under more restricted conditions. Across the three empirical chapters, we demonstrate how systematically completing Milestone 3 yields more discriminating evidence about when, how, and why reasoning and search biases emerge.

Finally, the proposed approach satisfies central criteria for strong theories articulated in van Lange's (2013) TAPAS framework. It increases precision by specifying which effects must occur under which process conditions, rather than merely predicting outcome biases. It enhances testability by ruling out broad classes of statistical models that are incompatible with the process assumptions, thereby increasing the theory's diagnostic power. It supports applicability by demonstrating that the same translation principles generalize across domains, tasks, and dependent measures. By sharpening the link between process assumptions and empirical tests, it also facilitates parsimony and accumulation, as findings can more directly inform theory refinement rather than motivating ad hoc extensions. From a Popperian perspective, the contribution of the present approach does not lie in increasing falsifiability per se. Paradigm-guided analyses have typically specified clear and empirically falsifiable statistical models. However, these models are often tied to specific operationalizations within a given paradigm and therefore do not directly test the core process assumptions of the theory.

By translating conceptual process assumptions into constraints on admissible hypotheses and statistical models, the present approach increases a theory's empirical content in the

Popperian sense: it restricts the space of outcomes that are compatible with the theory itself, rather than merely with a task or analytic convention.

In this sense, the systematic translation of conceptual models into statistical tests constitutes a theory-critical intermediate step. It neither replaces verbal theory specification nor formal modeling but connects the two by ensuring that empirical tests genuinely probe the core commitments of conceptual models. By making theory testing more diagnostic while remaining broadly applicable, the present dissertation advances a central step toward cumulative, process-oriented theorizing in judgment and decision making.

7.3.2 Contribution to Theory Integration

A central motivation of the present dissertation is to contribute not only to theory construction but also to theory integration in research on reasoning and information search biases. Concerns about theoretical fragmentation in judgment and decision making have long been raised, with numerous bias phenomena coexisting under partially overlapping labels, paradigms, and explanatory narratives (Eronen & Bringmann, 2021; Gigerenzer, 2010, 2017). Classic critiques have emphasized that the field often accumulates robust effects without corresponding gains in theoretical constraint or integration, resulting in theory sprawl—an increasing number of theories that are redundant, weakly specified, or insufficiently connected (Gigerenzer, 2010; Glöckner & Betsch, 2011; Glöckner & Fiedler, 2025a).

In response to this fragmentation, Gigerenzer (2010, 2017) has argued that one of psychology's most pressing tasks is the integration of existing theoretical "patchworks" into more coherent explanatory frameworks. Rather than relying exclusively on competitive theory testing (Popper, 1970) or on the construction of unified theories (Anderson, 1983, 1990; A. Newell, 1990), Gigerenzer (2017) proposes theory integration as a complementary route to scientific progress—one that seeks to relate phenomena and concepts by identifying shared explanatory principles. As discussed in Section 7.2.1, recent proposition-based approaches to theory construction can be understood as addressing a key prerequisite for such integration. In particular, Glöckner and Fiedler (2025b) emphasize that theoretical progress requires making assumptions explicit at the level of verbal propositions, thereby clarifying what exactly is claimed and what follows from these claims. This approach sharpens the conceptual content of theories and provides a foundation for identifying what is to be integrated, but it does not by itself resolve how integrative claims should be empirically tested.

The present dissertation builds on this foundation by addressing a subsequent challenge that corresponds to Gigerenzer's (2017) second integrative step: the integration of concepts through shared explanatory mechanisms. Recent work has independently proposed integrative frameworks that aim to unify multiple reasoning biases under common principles. Simon and Read (2025) advance a coherence-based framework of biased reasoning, conceptualizing a wide range of biases as the outcome of parallel constraint satisfaction among beliefs, preferences, and evaluations. Importantly, coherence-based models have already been proposed as accounts of biased sequential information integration (DeKay, 2015). The present dissertation builds on this work by using coherence-based models of sequential integration as a foundation for empirically testing the broader coherence-based framework of biased reasoning proposed by Simon and Read (2025).

In parallel, Oeberst and Imhoff (2023) propose belief-consistent information processing (BCIP) as a common mechanism underlying a broad set of biases, emphasizing unidirectional distortions of information processing toward prior beliefs. While this framework highlights important regularities and promotes parsimony at the level of belief consistency, it differs from coherence-based accounts in its process commitments. Crucially, recent work by Oeberst et al. (2025) directly addresses the relation between BCIP and coherence-based reasoning, arguing that BCIP can be understood as a special case of coherence-based reasoning under conditions of strongly held or situationally dominant prior beliefs. They further emphasize that progress in this area requires empirical tests that jointly assess belief change and belief-consistent information processing in order to distinguish unidirectional from bidirectional coherence shifts—an aspect that has been largely neglected in prior research.

Chapter 5 of the present dissertation directly responds to this call. Focusing on partisan motivated reasoning in political judgment—a domain characterized by strong and stable ideological priors—it provides a process-level test of how belief-based biases relate to coherence-based reasoning. Rather than treating belief-consistent processing as a separate mechanism, the findings show that BCIP-like effects can be understood as arising within a dynamic coherence system in which emerging preferences, ideological priors, and incoming evidence interact over time. Importantly, the results demonstrate that belief-based influences do not operate in a purely unidirectional manner but are embedded in bidirectional coherence dynamics. In this way, Chapter 5 empirically substantiates and refines recent integrative proposals by specifying the conditions under which belief-consistent processing emerges as a special case of broader coherence-based dynamics, rather than constituting a distinct or competing mechanism.

Building on this account of how belief-based biases arise within coherence dynamics, the integrative contribution of the present dissertation extends beyond reasoning and evaluation to the domain of information search. Chapter 6 builds directly on the iCodes framework, which already formalizes predecisional information search and predicts attraction-based sampling as a consequence of coherence-seeking during preference formation. The theoretical contribution of Chapter 6 lies not in introducing information search as a new component, but in extending this existing coherence-based framework to a decision environment in which the valence of information is known at the time of selection.

Within this extended framework, attraction-based search, confirmatory search, and a general positivity bias can be jointly accommodated within a single sequential coherence model. Rather than positing separate mechanisms, these tendencies emerge from the same underlying coherence dynamics under different constraints of the decision environment, allowing previously separate lines of research to be integrated within a common explanatory framework while preserving their theoretical distinctiveness. Importantly, this framework also enables the joint analysis of information distortion and information search, allowing for the derivation and empirical testing of hypotheses about their interaction over the course of sequential decision making.

Together with the contributions of Chapter 5, this extension demonstrates how coherence-based process models can integrate biases in evaluation, belief categorization, and information acquisition across different stages of sequential decision making, while preserving meaningful distinctions between these phenomena. Clarifying the integrative contribution, however, also requires specifying its conceptual scope and limits.

7.4 Conceptual Scope and Integrative Constraints

The present dissertation advances theoretical precision and cross-phenomenon integration within coherence-based models of sequential decision making. However, its contribution is deliberately limited to the integration and refinement of a specific theoretical framework rather than a comprehensive comparison of competing mechanisms. Recent proposals for theory development have called for coordinated competitive benchmarking of foundational mechanisms within shared paradigmatic structures, accompanied by systematic variance partitioning and cumulative multi-lab evaluation (Betsch & Bröder, 2025). The present work does not implement such a large-scale integrative architecture. Rather than adjudicating among multiple foundational principles within a unified benchmarking framework, it consolidates and refines a specific theoretical approach by clarifying its internal commitments and integrating related bias phenomena within a single paradigm. This reflects a limited but deliberate form of theoretical integration at the level of framework refinement rather than macro-level mechanism competition.

This positioning is particularly relevant for Chapter 4. The analysis of "Paradigm Blindness" addresses a stage of theory development that concerns the translation of conceptual process commitments into empirically testable implications. As emphasized in discussions of the theory crisis (Eronen & Bringmann, 2021), cumulative progress depends on robust phenomena that meaningfully constrain theoretical claims. However, such constraints presuppose analytic alignment between theoretical assumptions and statistical operationalization. Chapter 4 demonstrates that central implications of coherence-based models—specifically amplification, attenuation, and reversal patterns implied by parallel preference–evaluation coupling—are systematically underestimated under conventional serial analyses. The contribution therefore lies not in introducing additional constructs but in strengthening the mapping between conceptual assumptions and the statistical procedure used to evaluate them.

At the same time, this approach entails limitations. The present analyses operate at a stage of theory development that precedes full formal implementation (Borsboom et al., 2021). As argued by Farrell and Lewandowsky (2010), verbal theorizing may obscure the implications of complex process assumptions, and formal computational modeling can reveal constraints that are not transparent at the conceptual level. In the present case, direct formalization would likely have avoided the analytic mismatch discussed in this article. Coherence-based network models have already been implemented in formal and computational terms (Glöckner & Betsch, 2008a; McClelland et al., 2014; Read et al., 1997; Thagard, 1998) and directly adapting and testing such a model would have made the implications of parallel processing explicit. The present analyses therefore do not question the value of formalization. Instead, they target a different link in the explanatory chain. In the terminology of the productive explanation framework (van Dongen et al., 2025), the focus lies on whether central, model-defining implications of the conceptual framework are adequately reflected in the statistical tests used to evaluate it, independent of a particular functional form or parameterization.

Formal models introduce auxiliary assumptions, such as specific update rules, weighting schemes, noise structures, and parameter constraints. When empirical misfit occurs, it can therefore remain unclear whether a breakdown reflects a failure of a core theoretical principle (e.g., the assumption of parallel preference–evaluation coupling) or the consequences of a particular model instantiation. By testing implications that must hold across admissible implementations of coherence-based models, the present analyses aim to clarify whether the statistical representation of the phenomenon is aligned with the theory's defining commitments. In this sense, the findings strengthen the empirical anchor of the framework without committing to a specific computational realization.

Taken together, the present dissertation contributes to mid-range theoretical consolidation. It increases the empirical specificity of coherence-based process assumptions, improves alignment between conceptual commitments and statistical tests, and integrates related bias phenomena within a shared framework. It does not, however, constitute a comprehensive competitive benchmarking program across foundational mechanisms. Future work may embed the refined process specifications developed here within broader integrative architectures that systematically compare alternative principles under shared structural conditions. Building on these considerations, the following section outlines several directions for future research that extend the present approach and further develop its potential.

7.5 Future Directions

While the present dissertation refines and empirically specifies coherence-based accounts of sequential decision making, further work is needed to extend and test these insights more broadly. The following sections outline three promising directions that may further strengthen the explanatory and integrative potential of coherence-based models.

7.5.1 Formal Modeling of Sequential Integration and Search Biases

As outlined earlier, the present analyses operate at a stage of theory development that precedes full formal implementation. In the terminology of the Theory Construction Methodology (Borsboom et al., 2021), the empirical projects presented in this dissertation extend the transition from conceptual prototheory to empirically testable implications by introducing an intermediate step that systematically translates process assumptions into statistical tests. This step clarifies how conceptual process models constrain empirical predictions, but it does not yet constitute a formal model implementation.

Within the theory construction framework, the natural next step therefore concerns formal model development. According to Borsboom et al. (2021), formal modeling translates theoretical principles into explicit model equations that encode the assumed mechanisms and generate quantitative predictions. Such formalization can substantially strengthen theory development by clarifying theoretical assumptions (Fiedler, 2018), enabling more precise tests of proposed mechanisms (Farrell & Lewandowsky, 2010; Glöckner & Betsch, 2011), and generating novel predictions about process dynamics that can be empirically evaluated (Marewski & Olsson, 2009).

This need for formalization is particularly evident in research on predecisional information distortion under sequential information presentation. Although empirical evidence for such effects is extensive, theoretical accounts of the underlying coherence-based mechanisms have largely remained conceptual. While some formal models have been developed to account for information distortion effects (e.g., Hagmayer & Kostopoulou, 2013; Mehlhorn & Jahn, 2009), these models primarily focus on evaluation dynamics and do not incorporate information search within the same process architecture. As a result, formal models that jointly account for sequential information integration and active information search within a unified framework remain comparatively rare in sequential decision settings that involve stepwise information presentation.

One promising step in this direction is the Sequential-iCodes framework proposed in Chapter 6, which generalizes the iCodes model of coherence-based decision making to sequentially presented information (Jekel et al., 2018; Scharf et al., 2019). The model integrates processes of preference formation, information evaluation, and information search within a common network-based architecture. Because preferences emerge through coherence-maximizing distortions of sequentially presented information, Sequential-iCodes provides a framework that can, in principle, account for both information distortion and preference-guided information search within a unified architecture.

However, while many of these process-level predictions can be derived conceptually and have partly been examined empirically, they have not yet been formally specified and tested within a computational modeling framework. Moreover, coherence-based models of sequential decision making have rarely been compared directly to alternative modeling approaches that rely on different theoretical principles or architectures (e.g., Hogarth & Einhorn, 1992). Formal model comparison therefore represents an important direction for future research.

Developing and empirically testing such formal models would allow researchers to derive quantitative predictions about the joint dynamics of information integration, confidence formation, and information search. This step could substantially strengthen the explanatory power of coherence-based accounts of sequential decision making by enabling stronger tests of their underlying mechanisms and by clarifying how different process assumptions generate observable patterns in sequential decision tasks.

Taken together, the present dissertation contributes to the conceptual and empirical foundations required for such formal developments. By clarifying the process implications of coherence-based reasoning and translating them into empirically testable predictions, the present work helps prepare the ground for future formal modeling efforts that integrate information distortion, confidence dynamics, and information search within a unified computational framework. Complementary to these formal modeling efforts, an equally important challenge concerns how these processes are jointly examined at the empirical level.

7.5.2 Integrating Information Evaluation and Information Search

A closely related direction concerns the closer empirical integration of information evaluation and information search within sequential decision paradigms. Importantly, coherence-based models already provide a unified theoretical architecture for both processes. The iCodes framework represents an extension of PCS models in which the same network dynamics that govern preference formation also guide the selection of subsequent information (Jekel et al., 2018; Scharf et al., 2019). In principle, coherence-based models therefore offer a common process framework for understanding both how information is evaluated and which information is sampled during decision making.

Despite this shared theoretical foundation, empirical research has often examined these processes in separate research traditions. Work based on PCS models has primarily investigated the dynamics of information integration and preference formation (Glöckner & Betsch, 2008a; Holyoak & Simon, 1999; Simon et al., 2001). Research on information search, in contrast, has largely developed within the iCodes framework, which extends coherence-based models to account for preference-guided information search (Jekel et al., 2018; Scharf et al., 2019). As a consequence, empirical studies have often focused either on how information is evaluated or on which information is selected, even though both processes are assumed to arise from the same underlying coherence-based mechanisms.

The present dissertation provides an initial step toward bridging this gap. Across the empirical projects presented here, emerging preferences were shown to influence both how information is evaluated (Chapters 4 and 5) and which information is selected (Chapter 6). In particular, the experimental paradigm developed in Chapter 6 allows the simultaneous examination of preference-guided information search and distortion within the same sequential decision environment. This design illustrates how preference-guided search can shape the information available to the decision maker, thereby influencing the evaluative processes through which preferences subsequently evolve.

Future research could build on this approach by developing experimental paradigms that allow researchers to examine the reciprocal interaction between information integration and information search over time. Such work would allow researchers to examine how preference-guided search shapes which information decision makers encounter and how evolving evaluations influence subsequent information acquisition. Studying both processes within the same sequential framework may help clarify how early preferences become amplified over successive decision steps as coherence builds.

An additional direction concerns the extension of such paradigms to richer search environments. The studies in the present dissertation focused on binary decisions. Although parallel constraint satisfaction models can in principle be extended to multi-option choice situations, the dynamics observed here—particularly the convergence of coherence and confidence on leader-consistent information and the interaction between top-down constraints and bottom-up coherence dynamics—may behave differently when several options compete for activation. Whether the same interaction between coherence, confidence, and information valence generalizes to richer choice sets therefore remains an open question for future research.

Taken together, developing paradigms that jointly examine information integration and information search represents a promising direction for future research. Because coherence-based theories assume a shared mechanism underlying both processes, studying them within the same sequential decision environments may help clarify how preferences, evaluations, and information acquisition co-evolve during complex decisions.

7.5.3 Confidence in Sequential Bias Dynamics

A further promising direction for future research concerns the role of confidence in shaping biases in sequential decision making. Confidence is particularly relevant in this context because it appears to influence both how information is evaluated and which information is selected, suggesting that it may play an important role in linking biased information integration and biased information search within sequential decision processes.

Research on coherence-based decision processes has long emphasized the close relationship between confidence and biased information evaluation. In particular, the relationship between confidence and information distortion is understood to be bidirectional (Glöckner et al., 2010b; Holyoak & Simon, 1999; Simon et al., 2001; Simon, Krawczyk, et al., 2004). Empirical studies of predecisional information distortion consistently show that confidence in the current leader biases the evaluation of subsequent information: the higher the confidence at a given point in the decision process, the more likely it is that newly encountered information will be interpreted as supportive of the leading option (e.g., Chaxel et al., 2013; DeKay et al., 2009, 2011; Polman & Russo, 2012). In turn, such preference-consistent evaluations can further increase confidence in the leader, creating a feedback loop in which distortion and confidence reinforce one another over the course of the decision (DeKay et al., 2009).

Importantly, however, the two directions of this feedback loop have received unequal attention in the empirical literature. Most research has focused on confidence-driven distortion, whereas the reverse pathway—distortion-driven confidence—has been examined less frequently. Notable exceptions include DeKay et al. (2009, 2011) and Boyle et al. (2025), who showed that the degree to which newly encountered information is distorted predicts concurrent increases in confidence during the unfolding decision process.

As discussed in Chapter 4, this imbalance reflects a broader pattern in which theoretically bidirectional process assumptions are often translated into unidirectional statistical tests. In studies of predecisional information distortion, distortion is typically modeled as a function of prior preference or confidence, whereas the reverse pathway—distortion contributing to the development of confidence—has rarely been modeled explicitly. Chapter 4 demonstrated that such analytic simplifications can obscure theoretically central dynamics implied by coherence-based models. Future research could therefore benefit from more systematic investigation of the reciprocal connections of confidence and evaluative bias within sequential decision processes.

The present dissertation further suggests that confidence may also play an important role in understanding sequential information search. Across the studies reported in Chapter 6, confirmatory search did not emerge as a robust general tendency during early stages of preference formation. Instead, evidence for confirmatory search appeared primarily at higher levels of confidence. This pattern suggests that confirmatory search may not represent a ubiquitous predecisional bias but rather a later-stage search dynamic that becomes more likely once an emerging preference has reached a sufficient degree of stabilization. In this sense, confidence may mark the transition from attraction-based sampling toward a more defensive form of preference-consistent information acquisition.

Notably, the present results also indicate that the roles of preference strength and confidence may not be identical. In the search studies reported in Chapter 6, confirmatory search was associated with higher levels of confidence, but not with stronger preferences per se. This pattern suggests that confidence may capture an additional aspect of decision stabilization that is not fully reflected in measures of preference strength alone.

This distinction may also help clarify inconsistent findings in the literature on predecisional search biases. Several studies report a tendency to preferentially search for preference-consistent information (e.g., Fischer, Lea, et al., 2011; Fraser-Mackenzie & Dror, 2009; Sawicki et al., 2011; Young et al., 2011), whereas other studies examining strictly predecisional search have failed to observe such effects (Carlson & Guha, 2011; Mischkowski et al., 2021). As Russo (2015) noted, these differences appear to coincide with an important procedural distinction: studies that observe confirmatory search typically require participants to indicate a preliminary decision, whereas those studies that investigate information search prior to any expression of a tentative decision seemingly do not find such a bias. Russo suggested that expressing a preliminary choice may create a sense of commitment that motivates decision makers to defend the emerging leader, even when the choice remains formally reversible (Fischer & Greitemeyer, 2010).

From this perspective, confidence may provide a more precise process-level indicator of such commitment states. Rather than treating confirmatory search as emerging only after a discrete transition from predecisional to postdecisional processing, confidence may capture gradual differences in the stability of emerging preferences. Future research could therefore examine whether confirmatory search is better predicted by levels of confidence than by categorical distinctions between predecisional and postdecisional stages of decision making.

Finally, this line of reasoning connects to a broader limitation of coherence-based modeling of information search. As noted by Bröder and Scharf (2025), the iCodes framework currently lacks a psychologically plausible stopping rule. Whereas evidence accumulation models incorporate explicit decision thresholds, coherence-based models do not yet specify when sequential information acquisition should terminate. As potential candidates for such a mechanism, Bröder and Scharf (2025) discuss activation differences between competing options—corresponding to choice confidence—or minimum activation levels of hidden attribute values.

The present findings provide a first empirical indication that confidence may be informative in this regard, as higher confidence was associated with shifts in search behavior toward confirmatory sampling. Future research could therefore examine whether confidence predicts both changes in search strategies and the termination of information acquisition, thereby contributing to the development of psychologically plausible stopping rules for coherence-based decision models.

This perspective also suggests a connection to the belief-based dynamics examined in Chapter 5. If confidence reflects activation differences between options within the coherence network, top-down constraints such as prior beliefs or ideological labels may contribute to earlier activation differences between options. Such early activation differences would be expected to manifest as early confidence in the emerging leader, potentially stabilizing preference-consistent processing at earlier stages of the decision process.

Future research could therefore investigate whether early confidence—reflecting early activation differences between options—predicts more stable and potentially more biased sequential processing, particularly in contexts where strong prior beliefs provide top-down constraints on the coherence network. If these effects are primarily expressed through changes in confidence dynamics, the influence of prior beliefs may operate through the same activation mechanisms that govern coherence formation more generally. Under this interpretation, it may not always be necessary to introduce additional belief nodes into coherence networks to model decisions under prior constraints; instead, such effects may already be captured through their impact on activation differences between competing options.

Taken together, these considerations highlight the potential of confidence as a process variable in sequential decision biases. Because confidence appears to influence both information evaluation and information search, it may help bridge research on biased information integration and search. Examining how confidence co-evolves with information distortion, preference formation, and information search may therefore help clarify how coherence-based decision processes stabilize over time and how top-down constraints influence these dynamics.

7.6. Conclusion

Many real-world decisions unfold sequentially rather than through the simultaneous evaluation of all available information. As individuals encounter information step by step, emerging preferences shape both how new information is evaluated and which information is considered next. Understanding biased reasoning under such conditions therefore requires theoretical frameworks that capture the dynamic interplay between preference formation, belief based constraints, and information acquisition. The present dissertation contributes to this goal by examining coherence-based models in sequential decision environments. Across the empirical projects presented here, biases in information evaluation, belief-based reasoning, and information search emerged as interconnected consequences of coherence-seeking dynamics unfolding over time. Rather than reflecting independent cognitive failures, these biases can be understood as the outcome of a common process architecture in which beliefs, preferences, and evidence mutually constrain one another during the construction of a coherent interpretation of the decision environment. By clarifying the empirical implications of coherence-based process models and extending them across multiple domains of biased reasoning, the present work contributes to strengthening both theory construction and theory integration in research on judgment and decision making. More broadly, the present work shows how coherence-based dynamics unfolding in sequential decisions provide a unifying perspective on how biases in evaluation, belief-based reasoning, and information search emerge and interact over time.

References

Alford, J. R., Funk, C. L., & Hibbing, J. R. (2005). Are Political Orientations Genetically Transmitted? *American Political Science Review*, *99*(2), 153–167. https://doi.org/10.1017/S0003055405051579

Anderson, J. R. (1983). *The architecture of cognition*. Harvard University Press.

Anderson, J. R. (1990). *The adaptive character of thought*. Erlbaum. https://doi.org/10.4324/9780203771730

Ansolabehere, S., & Iyengar, S. (1994). Of horseshoes and horse races: Experimental studies of the impact of poll results on electoral behavior. *Political Communication*, *11*(4), 413–430. https://doi.org/10.1080/10584609.1994.9963048

Arkes, H. R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin*, *110*(3), 486–498.

Barnfield, M. (2020). Think Twice before Jumping on the Bandwagon: Clarifying Concepts in Research on the Bandwagon Effect. *Political Studies Review*, *18*(4), 553–574. https://doi.org/10.1177/1478929919870691

Baron, J. (2008). *Thinking and deciding (4th ed.).* Cambridge University Press.

Bartels, L. M. (1988). *Presidential Primaries and the Dynamics of Public Choice*. Princeton University Press. https://doi.org/10.2307/j.ctv173dzv9

Bartels, L. M. (1996). Uninformed Votes: Information Effects in Presidential Elections. *American Journal of Political Science*, *40*(1), 194–230. https://doi.org/10.2307/2111700

Bartels, L. M. (2002). Beyond the running tally: Partisan bias in political perceptions. *Political Behavior*, *24*(2), 117–150.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015a). Fitting Linear Mixed-Effects Models

Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

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

Beach, L., & Mitchell, T. R. (1978). A contingency model for the selection of decision strategies. *Academy of Management Review*, *3*, 439–449.

Beeler, J. D., & Hunton, J. E. (2002). Contingent economic rents: Insidious threats to audit independence. *Advances in Accounting Behavioral Research*, *5*, 21–50. https://doi.org/10.1016/S1474-7979(02)05036-6

Betsch, T., & Bröder, A. (2025). The Wheels of Scientific Progress: Integrative Theory Building in Decision Science. *Zeitschrift Für Psychologie*, *233*(4), 251–260. https://doi.org/10.1027/2151-2604/a000605

Bettman, J. R. (1979). *An information processing theory of consumer choice*. Addison-Wesley.

Blanchard, S. J., Carlson, K. A., & Meloy, M. G. (2014). Biased Predecisional Processing of Leading and Nonleading Alternatives. *Psychological Science*, *25*(3), 812–816. https://doi.org/10.1177/0956797613512663

Bock, O., Baetge, I., & Nicklisch, A. (2014). hroot: Hamburg registration and organization online tool. *European Economic Review*, *71*, 117–120. https://doi.org/10.1016/j.euroecorev.2014.07.003

Böckenholt, U., Albert, D., Aschenbrenner, M., & Schmalhofer, F. (1991). The effects of attractiveness, dominance, and attribute differences on information acquisition in multiattribute binary choice. *Organizational Behavior and Human Decision Processes*, *49*, 258–281.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*(4), 1061–1071. https://doi.org/10.1037/0033-295X.111.4.1061

Borsboom, D., van der Maas, H. L. J., Dalege, J., Kievit, R. A., & Haig, B. D. (2021). Theory Construction Methodology: A Practical Framework for Building Theories in Psychology. *Perspectives on Psychological Science*, *16*(4), 756–766. https://doi.org/10.1177/1745691620969647

Boyle, P. J., Hanlon, D., & Russo, J. E. (2012). The Value of Task Conflict to Group Decisions. *Journal of Behavioral Decision Making*, *25*(3), 217–227. https://doi.org/10.1002/bdm.725

Boyle, P. J., Russo, J. E., & Kim, J. (2025). When deciding creates overconfidence. *Judgment and Decision Making*, *20*, e15. https://doi.org/10.1017/jdm.2024.39

Bröder, A. (2003). Decision making with the "adaptive toolbox": Influence of environmental structure, intelligence, and working memory load. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *29*, 611–625. http://dx.doi.org/10.1037/0278-7393.29.4.611

Bröder, A., & Newell, B. R. (2008). Challenging some common beliefs: Empirical work within the adaptive toolbox metaphor. *Judgment and Decision Making*, *3*, 205–214.

Bröder, A., & Scharf, S. E. (2025). Testing a Formal Model of Coherence-Based Information Search in Decision-Making. In U. Ettinger, B. Heinrichs, & C. Murawski (Eds.), *Decision Making: Fundamentals and Applications* (pp. 97–114). Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-00880-010.1007/978-3-032-00880-0_6

Bröder, A., Scharf, S., Jekel, M., Glöckner, A., & Franke, N. (2021). Salience effects in information acquisition: No evidence for a top-down coherence influence. *Memory & Cognition*, *49*(8), 1537–1554. https://doi.org/10.3758/s13421-021-01188-9

Bröder, A., & Schiffer, S. (2006). Adaptive flexibility and maladaptive routines in selecting fast and frugal decision strategies. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *32*, 904–918.

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, *57*(3), 153–178. https://doi.org/10.1016/j.cogpsych.2007.12.002

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

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*, 432–459.

Campbell, A., Converse, P., Miller, W., & Stokes, D. (1960). *The American Voter*. John Wiley & Sons.

Campbell, D. T., Kenny, D. A., & Reichardt, C. S. (2003). *A primer on regression artifacts*. Guilford Press.

Carlson, K. A., & Guha, A. (2011). Leader-focused search: The impact of an emerging preference on information search. *Organizational Behavior and Human Decision Processes*, *115*(1), 133–141. https://doi.org/10.1016/j. obhdp.2010.12.002.

Carlson, K. A., Meloy, M. G., & Russo, J. E. (2006). Leader-Driven Primacy: Using Attribute Order to Affect Consumer Choice. *Journal of Consumer Research*, *32*(4), 513–518. https://doi.org/10.1086/500481

Carlson, K. A., & Pearo, L. K. (2004). Limiting predecisional distortion by prior valuation of attribute components. *Organizational Behavior and Human Decision Processes*, *94*(1), 48–

59. https://doi.org/10.1016/j.obhdp.2004.02.001

Carlson, K. A., & Russo, J. E. (2001). Biased interpretation of evidence by mock jurors. *Journal of Experimental Psychology: Applied*, *7*(2), 91–103. https://doi.org/10.1037/1076-898X.7.2.91

Ceci, S. J., & Kain, E. L. (1982). Jumping on the Bandwagon With the Underdog: The Impact of Attitude Polls on Polling Behavior. *Public Opinion Quarterly*, *46*(2), 228. https://doi.org/10.1086/268715

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). https://doi.org/10.1037//0022-3514.39.5.752

Chaxel, A. S., Edward Russo, J., & Kerimi, N. (2013). Preference-driven biases in decision makers' information search and evaluation. *Judgment and Decision Making*, *8*(5). https://doi.org/10.1017/s1930297500003661

Chen, S., Duckworth, K., & Chaiken, S. (1999). Motivated Heuristic and Systematic Processing. *Psychological Inquiry*, *10*(1). https://doi.org/10.1207/s15327965pli1001_6

Cohen, G. L. (2003). Party Over Policy: The Dominating Impact of Group Influence on Political Beliefs. *Journal of Personality and Social Psychology*, *85*(5), 808–822.

Crowe, E., & Higgins, E. T. (1997). Regulatory Focus and Strategic Inclinations: Promotion and Prevention in Decision-Making. *Organizational Behavior and Human Decision Processes*, *69*(2), 117–132. https://doi.org/10.1006/obhd.1996.2675

Dalege, J., Borsboom, D., Van Harreveld, F., Van den Berg, H., Conner, M., & Van der Maas, H. L. J. (2016). Toward a formalized account of attitudes: The Causal Attitude Network (CAN) Model. *Psychological Review*, *123*(1), 2–22. https://doi.org/10.1037/a0039802

Darley, J. M., & Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology*, *44*(1), 20–33. https://doi.org/10.1037/0022-3514.44.1.20

DeKay, M. L. (2015). Predecisional Information Distortion and the Self-Fulfilling Prophecy of Early Preferences in Choice. *Current Directions in Psychological Science*, *24*(5), 405–411. https://doi.org/10.1177/0963721415587876

DeKay, M. L., Miller, S. A., Schley, D. R., & Erford, B. M. (2014). Proleader and antitrailer information distortion and their effects on choice and postchoice memory. *Organizational Behavior and Human Decision Processes*, *125*(2), 134–150. https://doi.org/10.1016/j.obhdp.2014.07.003

DeKay, M. L., Patiño-Echeverri, D., & Fischbeck, P. S. (2009). Distortion of probability and outcome information in risky decisions. *Organizational Behavior and Human Decision Processes*, *109*(1), 79–92. https://doi.org/10.1016/j.obhdp.2008.12.001

DeKay, M. L., Stone, E. R., & Miller, S. A. (2011). Leader-driven distortion of probability and payoff information affects choices between risky prospects. *Journal of Behavioral Decision Making*, *24*(4), 394–411. https://doi.org/10.1002/bdm.699

DeKay, M. L., Stone, E. R., & Sorenson, C. M. (2012). Sizing up information distortion: Quantifying its effect on the subjective values of choice options. *Psychonomic Bulletin & Review*, *19*(2), 349–356. https://doi.org/10.3758/s13423-011-0184-8

Ditto, P. H., Celniker, J. B., Siddiqi, S. S., Güngör, M., & Relihan, D. P. (2025). Partisan Bias in Political Judgment. *Annual Review of Psychology*, *76*(1), 717–740. https://doi.org/10.1146/annurev-psych-030424-122723

Ditto, P. H., Liu, B. S., Clark, C. J., Wojcik, S. P., Chen, E. E., Grady, R. H., Celniker, J. B., &

Zinger, J. F. (2019). At Least Bias Is Bipartisan: A Meta-Analytic Comparison of Partisan Bias in Liberals and Conservatives. *Perspectives on Psychological Science*, *14*(2), 273–291. https://doi.org/10.1177/1745691617746796

Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology*, *63*(4), 568–584. https://doi.org/10.1037/0022-3514.63.4.568

Doh, E. Y., Rim, H. B., Kim, K., & Lee, B.-K. (2025). The effect of regulatory focus on predecisional information distortion. *Current Psychology*, *44*(3), 1765–1779. https://doi.org/10.1007/s12144-024-07154-3

Doherty, M. E., Mynatt, C. R., Tweney, R. D., & Schiavo, M. D. (1979). Pseudodiagnosticity. *Acta Psychologica*, *43*(2), 111–121. https://doi.org/10.1016/0001- 6918(79)90017-9.

Dorrough, A. R., Glöckner, A., Betsch, T., & Wille, A. (2017). When knowledge activated from memory intrudes on probabilistic inferences from description - the case of stereotypes. *Acta Psychologica*, *180*, 64–78. https://doi.org/10.1016/j.actpsy.2017.08.006

Drobner, C., & Goerg, S. J. (2024). Motivated Belief Updating and Rationalization of Information. *Management Science*, *70*(7), 4583–4592. https://doi.org/10.1287/mnsc.2023.02537

Druckman, J. N., Peterson, E., & Slothuus, R. (2013). How Elite Partisan Polarization Affects Public Opinion Formation. *American Political Science Review*, *107*(1), 57–79. https://doi.org/10.1017/S0003055412000500

Edland, A. (1994). Time pressure and the application of decision rules: Choices and judgements among multiattribute alternatives. *Scandinavian Journal of Psychology*, *35*, 281–291. http://dx.doi.org/10.1111/j.1467-%0A9450.1994.tb00952.x

Edland, A., & Svenson, O. (1993). Judgement and decision making under time pressure. In O. Svenson & A. J. Maule (Eds.), *Time pressure and stress in human judgement and decision making* (pp. 27–40). Springer. http://dx.doi.org/10.1007/978-1-4757-6846-6_2

Engel, C., & Glöckner, A. (2013). Role-Induced Bias in Court: An Experimental Analysis. *Journal of Behavioral Decision Making*, *26*(3), 272–284. https://doi.org/10.1002/bdm.1761

Engel, C., Timme, S., & Glöckner, A. (2020). Coherence-based reasoning and order effects in legal judgments. *Psychology, Public Policy, and Law*, *26*(3), 333–352. https://doi.org/10.1037/law0000257

Eronen, M. I., & Bringmann, L. F. (2021). The Theory Crisis in Psychology: How to Move Forward. *Perspectives on Psychological Science*, *16*(4), 779–788. https://doi.org/10.1177/1745691620970586

Evans, J. S. B. T., Venn, S., & Feeney, A. (2002). Implicit and explicit processes in a hypothesis testing task. *British Journal of Psychology*, *93*(1), 31–46. https://doi.org/10.1348/000712602162436

Evans, N. J., & Wagenmakers, E.-J. (2020). Evidence Accumulation Models: Current Limitations and Future Directions. *The Quantitative Methods for Psychology*, *16*(2), 73–90. https://doi.org/10.20982/tqmp.16.2.p073

Farrell, S., & Lewandowsky, S. (2010). Computational Models as Aids to Better Reasoning in Psychology. *Current Directions in Psychological Science*, *19*(5), 329–335. https://doi.org/10.1177/0963721410386677

Festinger, L. (1957a). *A theory of cognitive dissonance*. Stanford University Press.

Festinger, L. (1957b). *A Theory of Cognitive Dissonancec*. Row, Peterson.

Fiedler, K. (2017). What Constitutes Strong Psychological Science? The (Neglected) Role of

Diagnosticity and A Priori Theorizing. *Perspectives on Psychological Science*, *12*(1), 46–61. https://doi.org/10.1177/1745691616654458

Fiedler, K. (2018). The Creative Cycle and the Growth of Psychological Science. *Perspectives on Psychological Science*, *13*(4), 433–438. https://doi.org/10.1177/1745691617745651

Fiedler, K., & Trafimow, D. (2024). Using theoretical constraints and the TASI taxonomy to delineate predictably replicable findings. *Psychonomic Bulletin & Review*, *31*(6), 2581–2598. https://doi.org/10.3758/s13423-024-02521-4

Fischer, P. (2011). Selective Exposure, Decision Uncertainty, and Cognitive Economy: A New Theoretical Perspective on Confirmatory Information Search. *Social and Personality Psychology Compass*, *5*(10), 751–762. https://doi.org/10.1111/j.1751-9004.2011.00386.x

Fischer, P., & Greitemeyer, T. (2010). A New Look at Selective-Exposure Effects: An Integrative Model. *Current Directions in Psychological Science*, *19*(6), 384–389. https://doi.org/10.1177/0963721410391246.

Fischer, P., Greitemeyer, T., & Frey, D. (2008). Self-regulation and selective exposure: The impact of depleted self-regulation resources on confirmatory information processing. *Journal of Personality and Social Psychology*, *94*(3), 382–395. https://doi.org/10.1037/0022-3514.94.3.382

Fischer, P., Jonas, E., Frey, D., & Kastenmüller, A. (2008). Selective exposure and decision framing: The impact of gain and loss framing on confirmatory information search after decisions. *Journal of Experimental Social Psychology*, *44*(2), 312–320. https://doi.org/10.1016/j.jesp.2007.06.001

Fischer, P., Jonas, E., Frey, D., & Schulz-Hardt, S. (2005). Selective exposure to information: the impact of information limits. *European Journal of Social Psychology*, *35*(4), 469–492.

https://doi.org/10.1002/ejsp.264

Fischer, P., Kastenmüller, A., Greitemeyer, T., Fischer, J., Frey, D., & Crelley, D. (2011). Threat and selective exposure: The moderating role of threat and decision context on confirmatory information search after decisions. *Journal of Experimental Psychology: General*, *140*(1), 51–62. https://doi.org/10.1037/a0021595

Fischer, P., Lea, S., Kastenmüller, A., Greitemeyer, T., Fischer, J., & Frey, D. (2011). The process of selective exposure: Why confirmatory information search weakens over time. *Organizational Behavior and Human Decision Processes*, *114*(1), 37–48.

Fischhoff, B. (1977). Perceived informativeness of facts. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(2), 349–358. https://doi.org/10.1037/0096-1523.3.2.349

Fischhoff, B., & Beyth-Marom, R. (1983). Hypothesis evaluation from a Bayesian perspective. *Psychological Review*, *90*(3), 239–260. https://doi.org/10.1037/0033-295X.90.3.239

Fleitas, D. W. (1971). Bandwagon and Underdog Effects in Minimal-Information Elections. *American Political Science Review*, *65*(2), 434–438. https://doi.org/10.2307/1954459

Fraser-Mackenzie, P. A. F., & Dror, I. E. (2009). Selective information sampling: Cognitive coherence in evaluation of a novel item. *Judgment and Decision Making*, *4*(4), 307–316. https://doi.org/10.1017/S1930297500003880

Freund, T., Kruglanski, A. W., & Shpitzajzen, A. (1985). The Freezing and Unfreezing of Impressional Primacy. *Personality and Social Psychology Bulletin*, *11*(4), 479–487. https://doi.org/10.1177/0146167285114013

Frey, D. (1986). Recent Research on Selective Exposure to Information. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (19th ed., pp. 41–80). Academic Press.

Funder, D. C. (1995). 1995, Funder - personality judgments.pdf. *Psychological Review*, *102*(4), 652–670.

Gerber, A. S., Huber, G. A., Doherty, D., Dowling, C. M., & Ha, S. E. (2010). Personality and Political Attitudes: Relationships across Issue Domains and Political Contexts. *American Political Science Review*, *104*(1), 111–133. https://doi.org/10.1017/S0003055410000031

Gigerenzer, G. (2010). Personal Reflections on Theory and Psychology. *Theory & Psychology*, *20*(6), 733–743. https://doi.org/10.1177/0959354310378184

Gigerenzer, G. (2017). A theory integration program. *Decision*, *4*(3), 133–145. https://doi.org/10.1037/dec0000082

Gigerenzer, G., & Gaissmaier, W. (2010). Heuristic decision making. *Annual Review of Psychology*, *62*, 451–482. http://dx.doi.org/10.1146/%0Aannurev-psych-120709-145346.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*, 650–669.

Gigerenzer, G., Todd, P. M., & The ABC Research Group. (1999). *Simple heuristics that make us smart*. Oxford University Press.

Gigerenzer, G., Todd, P. M., & The ABC Research Group. (2000). *Simple heuristics that make us smart*. Oxford University Press.

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, *102*, 74–78. https://doi.org/10.1016/j.paid.2016.06.069

Glaholt, M. G., & Reingold, E. M. (2011). Eye movement monitoring as a process tracing methodology in decision making research. *Journal of Neuroscience, Psychology, and Economics*, *4*(2), 125–146. https://doi.org/10.1037/a0020692

Glöckner, A., & Betsch, T. (2008a). Modeling option and strategy choices with connectionist networks: Towards an integrative model of automatic and deliberate decision making. *Judgment and Decision Making*, *3*(3), 215–228. https://doi.org/10.1017/S1930297500002424

Glöckner, A., & Betsch, T. (2008b). Multiple-reason decision making based on automatic processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(5), 1055–1075. https://doi.org/10.1037/0278-7393.34.5.1055

Glöckner, A., & Betsch, T. (2010). Accounting for critical evidence while being precise and avoiding the strategy selection problem in a parallel constraint satisfaction approach: A reply to Marewski (2010. *Journal of Behavioral Decision Making*, *23*, 468–472.

Glöckner, A., & Betsch, T. (2011). The empirical content of theories in judgment and decision making: Shortcomings and remedies. *Judgment and Decision Making*, *6*(8), 711–721.

Glöckner, A., & Betsch, T. (2012). Decisions beyond boundaries: When more information is processed faster than less. *Acta Psychologica*, *139*, 532–542.

Glöckner, A., Betsch, T., & Schindler, N. (2010a). Coherence shifts in probabilistic inference tasks. *Journal of Behavioral Decision Making*, *23*(5), 439–462. https://doi.org/10.1002/bdm.668

Glöckner, A., Betsch, T., & Schindler, N. (2010b). Coherence shifts in probabilistic inference tasks. *Journal of Behavioral Decision Making*, *23*(5), 439–462. https://doi.org/10.1002/bdm.668

Glöckner, A., & Fiedler, S. (2024). "Theory Specification and Theory Building in Psychology." *Zeitschrift Für Psychologie*, *232*(1), 68–70. https://doi.org/10.1027/2151-2604/a000541

Glöckner, A., & Fiedler, S. (2025a). Addressing Psychology's Theory Crisis. *Zeitschrift Für*

Psychologie, *233*(4), 219–223. https://doi.org/10.1027/2151-2604/a000608

Glöckner, A., & Fiedler, S. (2025b). How to Improve Theory Specification in Psychology. *Zeitschrift Für Psychologie*, *233*(4), 224–239. https://doi.org/10.1027/2151-2604/a000599

Glöckner, A., Heinen, T., Johnson, J., & Raab, M. (2012). Network approaches for expert decisions in sports. *Human Movement Science*, *31*, 318–333.

Glöckner, A., & Herbold, A. (2011). An eye-tracking study on information processing in risky decisions: Evidence for compensatory strategies based on automatic processes. *Journal of Behavioral Decision Making*, *24*(1), 71–98. https://doi.org/10.1002/bdm.684

Glöckner, A., Hilbig, B. E., & Jekel, M. (2014). What is adaptive about adaptive decision making? A parallel constraint satisfaction account. *Cognition*, *133*(3), 641–666. https://doi.org/10.1016/j.cognition.2014.08.017

Glöckner, A., & Hodges, S. D. (2011). Parallel constraint satisfaction in memory-based decisions. *Experimental Psychology*, *58*, 180–195.

Glöckner, A., & Moritz, S. (2008). A fine-grained analysis of the jumping to conclusions bias in schizophrenia: Data-gathering, response confidence, and information integration. *Judgment and Decision Making*, *4*, 587–600. http://dx.doi.org/10.2139/ssrn.1313623

Goren, P., Federico, C. M., & Kittilson, M. C. (2009). Source cues, partisan identities, and political value expression. *American Journal of Political Science*, *53*(4), 805–820. https://doi.org/10.1111/j.1540-5907.2009.00402.x

Greenwald, A. G. (1980). The totalitarian ego: Fabrication and revision of personal history. *American Psychologist*, *35*(7), 603–618. https://doi.org/10.1037/0003-066X.35.7.603

Hagmayer, Y., & Kostopoulou, O. (2013). A probabilistic constraint satisfaction model of information distortion in diagnostic reasoning. *Proceedings of the 35th Annual Conference* *of the Cognitive Science Society*, *i*, 531–536.

Haig, B. D. (2005). An abductive theory of scientific method. *Psychological Methods*, *10*, 371–388.

Haig, B. D. (2014). *Investigating the psychological world: Scientific method in the behavioral science*. MIT Press.

Hale, J., Hastings, J., West, R., Lefevre, C. E., Direito, A., Bohlen, L. C., Godinho, C., Anderson, N., Zink, S., Groarke, H., & Michie, S. (2020). An ontology-based modelling system (OBMS) for representing behaviour change theories applied to 76 theories. *Wellcome Open Research*, *5*, 177. https://doi.org/10.12688/wellcomeopenres.16121.1

Hart, W., Albarracìn, D., Eagly, A. H., Brechan, L., Lindberg, M. J., & Merrill, L. (2009). Feeling validated versus being correct: A meta-analysis of selective exposure to information. *Psychological Bulletin*, *135*, 555–588. https://doi.org/http://dx.doi.org/10.1037/a0015701

Hastie, R., & Dawes, R. M. (2001). *Rational choice in an uncertain world*. Sage.

Hatemi, P. K., Gillespie, N. A., Eaves, L. J., Maher, B. S., Webb, B. T., Heath, A. C., Medland, S. E., Smyth, D. C., Beeby, H. N., Gordon, S. D., Montgomery, G. W., Zhu, G., Byrne, E. M., & Martin, N. G. (2011). A Genome-Wide Analysis of Liberal and Conservative Political Attitudes. *The Journal of Politics*, *73*(1), 271–285. https://doi.org/10.1017/S0022381610001015

Hatemi, P. K., Hibbing, J. R., Medland, S. E., Keller, M. C., Alford, J. R., Smith, K. B., Martin, N. G., & Eaves, L. J. (2010). Not by Twins Alone: Using the Extended Family Design to Investigate Genetic Influence on Political Beliefs. *American Journal of Political Science*, *54*(3), 798–814. https://doi.org/10.1111/j.1540-5907.2010.00461.x

Hausmann, D., & Läge, D. (2008). Sequential evidence accumulation in decision making: The individual desired level of confidence can explain the extent of information acquisition. *Judgment and Decision Making*, *3*, 229–243.

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs*, *76*(4), 408–420. https://doi.org/10.1080/03637750903310360

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.* John Wiley & Sons.

Henninger, F., Shevchenko, Y., Mertens, U. K., Kieslich, P. J., & Hilbig, B. E. (2022). lab.js: A free, open, online study builder. *Behavior Research Methods*, *54*(2), 556–573. https://doi.org/10.3758/s13428-019-01283-5

Hibbing, J. R., & Theiss-Morse, E. (1995). *Congress as public enemy: Public attitudes toward American political institutions.* Cambridge University Press.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief adjustment model. *Cognitive Psychology*, *24*, 1–55.

Holyoak, K. J., & Simon, D. (1999). Bidirectional reasoning in decision making by constraint satisfaction. *Journal of Experimental Psychology: General*, *128*(1), 3–31. https://doi.org/10.1037/0096-3445.128.1.3

Holyoak, K. J., & Thagard, P. (1989). Analogical Mapping by Constraint Satisfaction. *Cognitive Science*, *13*(3), 295–355. https://doi.org/10.1207/s15516709cog1303_1

Hornsey, M. J., Finlayson, M., Chatwood, G., & Begeny, C. T. (2020). Donald Trump and vaccination: The effect of political identity, conspiracist ideation and presidential tweets on vaccine hesitancy. *Journal of Experimental Social Psychology*, *88*, 103947. https://doi.org/10.1016/j.jesp.2019.103947

Idson, L. C., Liberman, N., & Higgins, E. T. (2000). Distinguishing Gains from Nonlosses and Losses from Nongains: A Regulatory Focus Perspective on Hedonic Intensity. *Journal of Experimental Social Psychology*, *36*(3), 252–274. https://doi.org/10.1006/jesp.1999.1402

Iyengar, S., & Hahn, K. S. (2009). Red Media, Blue Media: Evidence of Ideological Selectivity in Media Use. *Journal of Communication*, *59*(1), 19–39. https://doi.org/10.1111/j.1460-2466.2008.01402.x

Iyengar, S., Hahn, K. S., Krosnick, J. A., & Walker, J. (2008). Selective Exposure to Campaign Communication: The Role of Anticipated Agreement and Issue Public Membership. *The Journal of Politics*, *70*(1), 186–200. https://doi.org/10.1017/S0022381607080139

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The Origins and Consequences of Affective Polarization in the United States. *Annual Review of Political Science*, *22*(1), 129–146. https://doi.org/10.1146/annurev-polisci-051117-073034

Jackson, J. C., Rand, D., Lewis, K., Norton, M. I., & Gray, K. (2017). Agent-Based Modeling. *Social Psychological and Personality Science*, *8*(4), 387–395. https://doi.org/10.1177/1948550617691100

Jaeger, B. C., Edwards, L. J., Das, K., & Sen, P. K. (2017). An R 2 statistic for fixed effects in the generalized linear mixed model. *Journal of Applied Statistics*, *44*(6), 1086–1105. https://doi.org/10.1080/02664763.2016.1193725

Jekel, M., Glöckner, A., & Bröder, A. (2018). A new and unique prediction for cue-search in a parallel-constraint satisfaction network model: The attraction search effect. *Psychological Review*, *125*(5), 744–768. https://doi.org/10.1037/rev0000107

Johnson-Laird, P. N., & Wason, P. C. (1970). A theoretical analysis of insight into a reasoning task. *Cognitive Psychology*, *1*(2), 134–148. https://doi.org/10.1016/0010-0285(70)90009-5

Johnson, E. J., Häubl, G., & Keinan, A. (2007). Aspects of endowment: A query theory of value construction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 461–474. https://doi.org/10.1037/0278-7393.33.3.461

Johnson, E. J., & Payne, J. W. (1985). Effort and accuracy in choice. *Management Science*, *30*, 1213–1231.

Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D., & Mandel, G. (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature Climate Change*, *2*(10), 732–735. https://doi.org/10.1038/nclimate1547

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(47), 263–291.

Kam, C. D. (2005). Who Toes the Party Line? Cues, Values, and Individual Differences. *Political Behavior*, *27*(2), 163–182. https://doi.org/10.1007/s11109-005-1764-y

Kanai, R., Feilden, T., Firth, C., & Rees, G. (2011). Political Orientations Are Correlated with Brain Structure in Young Adults. *Current Biology*, *21*(8), 677–680. https://doi.org/10.1016/j.cub.2011.03.017

Kim, S., Taber, C. S., & Lodge, M. (2010). A Computational Model of the Citizen as Motivated Reasoner: Modeling the Dynamics of the 2000 Presidential Election. *Political Behavior*, *32*(1), 1–28. https://doi.org/10.1007/s11109-009-9099-8

Klayman, J. (1995). Varieties of confirmation bias. *Psychology of Learning and Motivation*, *32*, 385–418.

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

Klayman, J., & Ha, Y. W. (1989). Hypothesis Testing in Rule Discovery: Strategy, Structure, and Content. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(4). https://doi.org/10.1037/0278-7393.15.4.596

Klein, S. B. (2014). What can recent replication failures tell us about the theoretical commitments of psychology? *Theory & Psychology*, *24*(3), 326–338. https://doi.org/10.1177/0959354314529616

Knobloch-Westerwick, S., & Jingbo Meng. (2009). Looking the Other Way. *Communication Research*, *36*(3), 426–448. https://doi.org/10.1177/0093650209333030

Koch, T., & Wendt, A. (2025). Theory Crises in Psychology – What Can We Learn From Structuralism? *Zeitschrift Für Psychologie*, *233*(4), 240–250. https://doi.org/10.1027/2151-2604/a000600

Kohler, W., & Emery, D. A. (1947). Figural After-Effects in the Third Dimension of Visual Space. *The American Journal of Psychology*, *60*(2), 159. https://doi.org/10.2307/1417870

Koopman, J., Howe, M., Hollenbeck, J. R., & Sin, H.-P. (2015). Small sample mediation testing: Misplaced confidence in bootstrapped confidence intervals. *Journal of Applied Psychology*, *100*(1), 194–202. https://doi.org/10.1037/a0036635

Kostopoulou, O., Russo, J. E., Keenan, G., Delaney, B. C., & Douiri, A. (2012). Information distortion in physicians' diagnostic judgments. *Medical Decision Making*, *32*(6), 831–839. https://doi.org/10.1177/0272989X12447241

Kostopoulou, O., Sirota, M., Round, T., Samaranayaka, S., & Delaney, B. C. (2017). The Role of Physicians' First Impressions in the Diagnosis of Possible Cancers without Alarm

Symptoms. *Medical Decision Making*, *37*(1), 9–16. https://doi.org/10.1177/0272989X16644563

Kruglanski, A. W. (1980). Lay epistemo-logic—process and contents: Another look at attribution theory. *Psychological Review*, *87*(1), 70–87. https://doi.org/10.1037/0033-295X.87.1.70

Kruglanski, A. W. (2001). That “vision thing”: The state of theory in social and personality psychology at the edge of the new millennium. *Journal of Personality and Social Psychology*, *80*(6), 871–875. https://doi.org/10.1037/0022-3514.80.6.871

Kruglanski, A. W. (2013). Only one? The default interventionist perspective as a unimodel—Commentary on Evans & Stanovich (2013). *Perspectives on Psychological Science*, *8*(3), 242–247. https://doi.org/10.1177/1745691613483477

Kruglanski, A. W., & Ajzen, I. (1983). Bias and error in human judgment. *European Journal of Social Psychology*, *13*(1), 1–44. https://doi.org/10.1002/ejsp.2420130102

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., & Klar, Y. (1987). A view from a bridge: Synthesizing the consistency and attribution paradigms from a lay epistemic perspective. *European Journal of Social Psychology*, *17*(2), 211–241. https://doi.org/10.1002/ejsp.2420170208

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498. https://doi.org/10.1037/0033-2909.108.3.480

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes,traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review*, *103*, 284–308. http://dx.doi.org/10.1037/0033-295x.103%0A.2.284

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

Lammers, J., Bukowski, M., Potoczek, A., Fleischmann, A., & Hofmann, W. (2022). Disentangling the factors behind shifting voting intentions: The bandwagon effect reflects heuristic processing, while the underdog effect reflects fairness concerns. *Journal of Social and Political Psychology*, *10*(2), 676–692. https://doi.org/10.5964/jspp.9241

Lammers, J., Pauels, E., Fleischmann, A., & Galinsky, A. D. (2022). Why People Hate Congress but Love Their Own Congressperson: An Information Processing Explanation. *Personality and Social Psychology Bulletin*, *48*(3), 412–425. https://doi.org/10.1177/01461672211002336

Lange, K., Kühn, S., & Filevich, E. (2015). Correction: "Just Another Tool for Online Studies" (JATOS): An Easy Solution for Setup and Management of Web Servers Supporting Online Studies. *PLOS ONE*, *10*(7), e0134073. https://doi.org/10.1371/journal.pone.0134073

Laponce, J. A. (1966). An Experimental Method to Measure the Tendency to Equibalance in a Political System. *American Political Science Review*, *60*(4), 982–993. https://doi.org/10.2307/1953771

Lee, D. G., & Holyoak, K. J. (2021). Coherence shifts in attribute evaluations. *Decision*, *8*(4), 257–276. https://doi.org/10.1037/dec0000151

Lee, M. D., & Cummins, T. D. R. (2004). Evidence accumulation in decision making: Unifying the "take the best" and the "rational" models. *Psychonomic Bulletin & Review*, *11*(2), 343–

352. https://doi.org/10.3758/BF03196581

Leising, D., Grenke, O., & Cramer, M. (2023). Visual Argument Structure Tool (VAST) Version 1.0. *Meta-Psychology*, *7*. https://doi.org/10.15626/MP.2021.2911

Lodge, M., & Taber, C. S. (2013). The rationalizing voter. In *The Rationalizing Voter*. Cambridge University Press. https://doi.org/10.1017/CBO9781139032490

Lord, C. G., Lepper, M. R., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology*, *47*(6), 1231–1243. https://doi.org/10.1037/0022-3514.47.6.1231

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, *37*(11), 2098–2109. https://doi.org/10.1037/0022-3514.37.11.2098

Lord, C. G., & Taylor, C. A. (2009). Biased Assimilation: Effects of Assumptions and Expectations on the Interpretation of New Evidence. *Social and Personality Psychology Compass*, *3*(5), 827–841. https://doi.org/10.1111/j.1751-9004.2009.00203.x

Luce, R. D. (2000). *Utility of gains and losses: measurement-theoretical and experimental approaches*. Erlbaum.

Luce, R. D., & Raiffa, H. (1957). *Games and decisions: Introduction and critical survey*. Wiley.

Lupia, A. (1994). The effect of information on voting behavior and electoral outcomes: An experimental study of direct legislation. *Public Choice*, *78*(1), 65–86.

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology*, *58*(1), 593–614. https://doi.org/10.1146/annurev.psych.58.110405.085542

Mahoney, M. J. (1977). Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research*, *1*(2), 161–175. https://doi.org/10.1007/BF01173636

Marewski, J. N. (2010). On the theoretical precision and strategy selection problem of a single-strategy approach: A comment on Glöckner, Betsch, and Schindler (2010)/dx.doi.org/10.1002/bdm.680. *Journal of Behavioral Decision Making*, *23*, 463–467. http://dx.doi.org/10.1002/bdm.680

Marewski, J. N., & Olsson, H. (2009). Beyond the Null Ritual. *Zeitschrift Für Psychologie / Journal of Psychology*, *217*(1), 49–60. https://doi.org/10.1027/0044-3409.217.1.49

Martignon, L., & Hoffrage, U. (1999). Why does one-reason decision making work? A case study in ecological rationality. In G. Gigerenzer, P. M. Todd, & A. R. Group (Eds.), *Simple heuristics that make us smart* (pp. 119–140). Oxford University Press.

Mata, R., van Helversen, B., & Rieskamp, J. (2007). The aging decision maker: Cognitive aging and the adaptive selction of decision strategies. *Psychology and Aging*, *22*, 796–810. http://dx.doi.org/10.1093/acprof:%0Aoso/9780199744282.003.0022

Mata, R., van Helversen, B., & Rieskamp, J. (2011). When easy comes hard: The development of adaptive strategy selection. *Child Development*, *82*, 687–700. http://dx.doi.org/10.1111/j.1467-8624.2010.01535.x

McAllister, I. (2000). Keeping Them Honest: Public and Elite Perceptions of Ethical Conduct among Australian Legislators. *Political Studies*, *48*(1), 22–37. https://doi.org/10.1111/1467-9248.00248

McClelland, J. L., Mirman, D., Bolger, D. J., & Khaitan, P. (2014). Interactive Activation and Mutual Constraint Satisfaction in Perception and Cognition. *Cognitive Science*, *38*(6), 1139–1189. https://doi.org/10.1111/cogs.12146

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, *88*(5), 375–407. https://doi.org/10.1037/0033-295X.88.5.375

McClelland, J. L., Rumelhart, D. E., & Hinton, G. E. (1986). The appeal of parallel distributed processing. In D. E. Rumelhart, G. McClelland, & P. R. Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition Vol. 1* (pp. 3–44). MIT Press.

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, *46*(4), 806–834. https://doi.org/10.1037/0022-006X.46.4.806

Meehl, P. E. (1967). Theory-Testing in Psychology and Physics: A Methodological Paradox. *Philosophy of Science*, *34*(2), 103–115. https://doi.org/10.1086/288135

Meffert, M. F., Chung, S., Joiner, A. J., Waks, L., & Garst, J. (2006). The Effects of Negativity and Motivated Information Processing During a Political Campaign. *Journal of Communication*, *56*(1), 27–51. https://doi.org/10.1111/j.1460-2466.2006.00003.x

Mehlhorn, K., & Jahn, G. (2009). Modeling sequential information integration with parallel constraint satisfaction. In *Proceedings of the 31st annual conference of the cognitive science society* (pp. 2469–2474). Cognitive Science Society.

Miller, S. A., DeKay, M. L., Stone, E. R., & Sorenson, C. M. (2013). Assessing the sensitivity of information distortion to four potential influences in studies of risky choice. *Judgment and Decision Making*, *8*(6), 662–677. https://doi.org/10.1017/S1930297500004708

Mischkowski, D., Glöckner, A., & Lewisch, P. (2021). Information search, coherence effects, and their interplay in legal decision making. *Journal of Economic Psychology*, *87*, 102445. https://doi.org/10.1016/j.joep.2021.102445

Mondak, J. J., & Halperin, K. D. (2008). A Framework for the Study of Personality and Political Behaviour. *British Journal of Political Science*, *38*(2), 335–362. https://doi.org/10.1017/S0007123408000173

Moulines, C. U. (1996). Chapter 1: Structuralism: The Basic Ideas. In W. Balzer & C. U. Moulines (Eds.), *Structuralist Theory of Science* (pp. 1–14). DE GRUYTER. https://doi.org/10.1515/9783110879421.1

Munro, G. D., & Ditto, P. H. (1997). Biased Assimilation, Attitude Polarization, and Affect in Reactions to Stereotype-Relevant Scientific Information. *Personality and Social Psychology Bulletin*, *23*(6), 636–653. https://doi.org/10.1177/0146167297236007

Muraven, M., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin*, *126*(2), 247–259.

Muthukrishna, M., & Henrich, J. (2019). A problem in theory. *Nature Human Behaviour*, *3*(3), 221–229. https://doi.org/10.1038/s41562-018-0522-1

Mynatt, C. R., Doherty, M. E., & Dragan, W. (1993). Information relevance, working memory, and the consideration of alternatives. *The Quarterly Journal of Experimental Psychology Section A*, *46*(4), 759–778. https://doi.org/10.1080/14640749308401038.

Navarro, D. J., & Perfors, A. F. (2011). Enlightenment grows from fundamentals. In *Behavioral and Brain Sciences* (Vol. 34, Issue 4). https://doi.org/10.1017/S0140525X11000367

Nelson, J. D. (2005). Finding Useful Questions: On Bayesian Diagnosticity, Probability, Impact, and Information Gain. *Psychological Review*, *112*(4), 979–999. https://doi.org/10.1037/0033-295X.112.4.979

Nelson, J. D., McKenzie, C. R. M., Cottrell, G. W., & Sejnowski, T. J. (2010). Experience Matters. *Psychological Science*, *21*(7), 960–969.

https://doi.org/10.1177/0956797610372637

Newell, A. (1990). *Unified theories of cognition*. Harvard University Press.

Newell, B. R. (2005). Re-Visions of rationality? *Trends in Cognitive Sciences*, *9*, 11–15.

Newell, B. R., & Bröder, A. (2008). Cognitive processes, models and metaphors in decision research. *Judgment and Decision Making*, *3*, 195–204.

Newell, B. R., Rakow, T., Weston, N. J., & Shanks, D. R. (2004). Search strategies in decision making: The success of "success." *Journal of Behavioral Decision Making*, *17*(2), 117–137. https://doi.org/10.1002/bdm.465

Newell, B. R., & Shanks, D. R. (2003). Take the best or look at the rest? Factors influencing "one-reason" decision making. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *29*, 53–66. http://dx.doi.org/10.1037//0278-7393.29.1.53

Nickerson, R. S. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology*, *2*(2), 175–220. https://doi.org/10.1037/1089-2680.2.2.175

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review*, *108*(2), 291–310. https://doi.org/10.1037/0033-295X.108.2.291

Nurek, M., & Kostopoulou, O. (2016). What you find depends on how you measure it: reactivity of response scales measuring predecisional information distortion in medical diagnosis. *PloS One*, *11*(9).

Nurek, M., Kostopoulou, O., & Hagmayer, Y. (2014). Predecisional information distortion in physicians' diagnostic judgments: Strengthening a leading hypothesis or weakening its competitor? *Judgment and Decision Making*, *9*(6), 572–585. https://doi.org/10.1017/S1930297500006434

Nyhan, B., & Reifler, J. (2010). When Corrections Fail: The Persistence of Political Misperceptions. *Political Behavior*, *32*(2), 303–330. https://doi.org/10.1007/s11109-010-9112-2

Nyhan, B., Reifler, J., & Ubel, P. A. (2013). The Hazards of Correcting Myths About Health Care Reform. *Medical Care*, *51*(2), 127–132. https://doi.org/10.1097/MLR.0b013e318279486b

Oberauer, K., & Lewandowsky, S. (2019). Addressing the theory crisis in psychology. *Psychonomic Bulletin & Review*, *26*(5), 1596–1618. https://doi.org/10.3758/s13423-019-01645-2

Oeberst, A., & Imhoff, R. (2023). Toward Parsimony in Bias Research: A Proposed Common Framework of Belief-Consistent Information Processing for a Set of Biases. *Perspectives on Psychological Science*, *18*(6), 1464–1487. https://doi.org/10.1177/17456916221148147

Oeberst, A., Mischkowski, D., & Imhoff, R. (2025). Belief-Consistent Information Processing or Coherence-Based Reasoning: Integrating Two Parsimonious Frameworks for Biases. *European Journal of Social Psychology*, *55*(7), 1131–1135. https://doi.org/10.1002/ejsp.70006

Orquin, J. L., Lahm, E. S., & Stojić, H. (2021). The visual environment and attention in decision making. *Psychological Bulletin*, *147*(6), 597–617. https://doi.org/10.1037/bul0000328

Osgood, C. E., & Tannenbaum, P. H. (1955). The principle of congruity in the prediction of attitude change. *Psychological Review*, *62*(1), 42–55. https://doi.org/10.1037/h0048153

Palminteri, S., & Lebreton, M. (2022). The computational roots of positivity and confirmation biases in reinforcement learning. *Trends in Cognitive Sciences*, *26*(7), 607–621. https://doi.org/10.1016/j.tics.2022.04.005

Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior and Human Performance*, *16*(2), 366–387. https://doi.org/10.1016/0030-5073(76)90022-2

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *14*, 534–552.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge University Press.

Plott, C. R., & Levi, M. D. (1978). A model of agenda influence on committee decision. *American Economic Review*, *68*, 146–160.

Polman, E., & Russo, J. E. (2012). Commitment to a developing preference and predecisional distortion of information. *Organizational Behavior and Human Decision Processes*, *119*(1), 78–88. https://doi.org/10.1016/j.obhdp.2012.05.004

Popkin, S. L. (1991). *The reasoning voter: Communication and persuasion in presidential campaigns*. University of Chicago Press.

Popper, K. R. (n.d.). *Logik der Forschung (11th ed.)*. Mohr Siebeck.

Popper, K. R. (1970). Normal science and its dangers. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 51–58). Cambridge University Press.

Pyszczynski, T., & Greenberg, J. (1987). Toward an integration of cognitive and motivational perspectives on social inference: A biased hypothesis-testing model. In *Advances in experimental social psychology* (20th ed., pp. 297–340). Academic Press.

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org

Rahn, W. M. (1993). The Role of Partisan Stereotypes in Information Processing about Political Candidates. *American Journal of Political Science*, *37*(2), 472–496. https://doi.org/10.2307/2111381

Rakow, T., Newell, B. R., Fayers, K., & Hersby, M. (2005). Evaluating three criteria for establishing cue-search hierarchies in inferential judgement. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *31*, 1088–1104. http://dx.doi.org/10.1037/0278-7393.31.5%0A.1088

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108. https://doi.org/10.1037/0033-295X.85.2.59

Ratcliff, R., & Smith, P. L. (2004). A Comparison of Sequential Sampling Models for Two-Choice Reaction Time. *Psychological Review*, *111*(2), 333–367. https://doi.org/10.1037/0033-295X.111.2.333

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion Decision Model: Current Issues and History. *Trends in Cognitive Sciences*, *20*(4), 260–281. https://doi.org/10.1016/j.tics.2016.01.007

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review*, *9*(3), 438–481. https://doi.org/10.3758/BF03196302

Read, S. J., Vanman, E. J., & Miller, L. C. (1997). Connectionism, Parallel Constraint Satisfaction Processes, and Gestalt Principles: (Re)Introducing Cognitive Dynamics to Social Psychology. *Personality and Social Psychology Review*, *1*(1), 26–53. https://doi.org/10.1207/s15327957pspr0101_3

Rights, J. D., & Sterba, S. K. (2019). Quantifying explained variance in multilevel models: An integrative framework for defining R-squared measures. *Psychological Methods*, *24*(3), 309–338. https://doi.org/10.1037/met0000184

Rigoli, F. (2021). A computational perspective on faith: religious reasoning and Bayesian decision. *Religion, Brain & Behavior*, *11*(2), 147–164. https://doi.org/10.1080/2153599X.2020.1812704

Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionst model of decision making. *Psychological Review*, *108*(2), 370–392. https://doi.org/10.1037/0033-295X.108.2.370

Rubin, E. (1921). *Visuell wahrgenommene figuren: Studien in psychologischer analyse*. Gyldendalske boghandel.

Russo, J. E. (1977). The value of unit price information. *Journal of Marketing Research*, *14*, 193–201.

Russo, J. E. (2015). The predecisional distortion of information. *Neuroeconomics, Judgment, and Decision Making.*, 91–110. http://libproxy.nps.edu/login?url=http://search.proquest.com/docview/1605987206?accountid=12702%5Cnhttp://sfxhosted.exlibrisgroup.com/nps?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:book&genre=bookitem&sid=ProQ:PsycINFO&atitle=The+predecisional+d

Russo, J. E. (2018). Bayesian Revision vs. Information Distortion. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01550

Russo, J. E., Carlson, K. A., & Meloy, M. G. (2006). Choosing an Inferior Alternative. *Psychological Science*, *17*(10), 899–904. https://doi.org/10.1111/j.1467-9280.2006.01800.x

Russo, J. E., Carlson, K. A., Meloy, M. G., & Yong, K. (2008). The goal of consistency as a cause of information distortion. *Journal of Experimental Psychology: General*, *137*(3), 456.

Russo, J. E., Medvec, V. H., & Meloy, M. G. (1996). The distortion of information during decisions. *Organizational Behavior and Human Decision Processes*, *66*(1), 102–110. https://doi.org/10.1006/obhd.1996.0041

Russo, J. E., Meloy, M. G., & Medvec, V. H. (1998). Predecisional Distortion of Product Information. *Journal of Marketing Research*, *35*(4), 438–452. https://doi.org/10.1177/002224379803500403

Russo, J. E., Meloy, M. G., & Wilks, T. J. (2000). Predecisional distortion of information by auditors and salespersons. *Management Science*, *46*(1), 13–27. https://doi.org/10.1287/mnsc.46.1.13.15127

Russo, J. E., & Yong, K. (2011). The distortion of information to support an emerging evaluation of risk. *Journal of Econometrics*, *162*(1), 132–139. https://doi.org/10.1016/j.jeconom.2010.07.004

Sanbonmatsu, D. M., Posavac, S. S., Kardes, F. R., & Mantel, S. P. (1998). Selective hypothesis testing. *Psychonomic Bulletin and Review*, *5*(2), 197–220. https://doi.org/10.3758/BF03212944

Sassenberg, K., Landkammer, F., & Jacoby, J. (2014). The influence of regulatory focus and group vs. individual goals on the evaluation bias in the context of group decision making. *Journal of Experimental Social Psychology*, *54*, 153–164. https://doi.org/10.1016/j.jesp.2014.05.009

Sawicki, V., Wegener, D. T., Clark, J. K., Fabrigar, L. R., Smith, S. M., & Bengal, S. T. (2011). Seeking Confirmation in Times of Doubt. *Social Psychological and Personality Science*, *2*(5), 540–546. https://doi.org/10.1177/1948550611400212

Scharf, S. E., Jekel, M., & Glöckner, A. (2022). Awareness of Option Attractiveness Increases the Attraction Search Effect: Modeling the Awareness Effect in an Extended iCodes Model. *Decision*, *9*(1). https://doi.org/10.1037/dec0000162

Scharf, S. E., Wiegelmann, M., & Bröder, A. (2019). Information search in everyday decisions: The generalizability of the attraction search effect. *Judgment and Decision Making*, *14*(4), 488–512. https://doi.org/10.1017/S193029750000615X

Schmittat, S. M., & Englich, B. (2016). If you judge, investigate! Responsibility reduces confirmatory information processing in legal experts. *Psychology, Public Policy, and Law*, *22*(4), 386–400.

Schulte-Mecklenbeck, M., Johnson, J. G., Böckenholt, U., Goldstein, D. G., Russo, J. E., Sullivan, N. J., & Willemsen, M. C. (2017). Process-Tracing Methods in Decision Making: On Growing Up in the 70s. *Current Directions in Psychological Science*, *26*(5), 442–450. https://doi.org/10.1177/0963721417708229

Simon, D. (2004). A third view of the black box: Cognitive coherence in legal decision making. *University of Chicago Law Review*, *71*(2), 511–586.

Simon, D., & Holyoak, K. J. (2002). Structural Dynamics of Cognition: From Consistency Theories to Constraint Satisfaction. *Personal and Social Psychology Review*, *6*(6), 283–294. https://doi.org/10.2139/ssrn.318722

Simon, D., Krawczyk, D. C., & Holyoak, K. J. (2004). Construction of Preferences by Constraint Satisfaction. *Psychological Science*, *15*(5), 331–336. https://doi.org/10.1111/j.0956-7976.2004.00678.x

Simon, D., Pham, L. B., Le, Q. A., & Holyoak, K. J. (2001). The emergence of coherence over the course of decision making. *Journal of Experimental Psychology: Learning, Memory,* *and Cognition*, *27*(5), 1250–1260.

Simon, D., & Read, S. J. (2025). Toward a General Framework of Biased Reasoning: Coherence-Based Reasoning. *Perspectives on Psychological Science*, *20*(3), 421–459. https://doi.org/10.1177/17456916231204579

Simon, D., Snow, C. J., & Read, S. J. (2004). The redux of cognitive consistency theories: Evidence judgments by constraint satisfaction. *Journal of Personality and Social Psychology*, *86*(6), 814–837. https://doi.org/10.1037/0022-3514.86.6.814

Simon, D., & Spiller, S. A. (2016). The Elasticity of Preferences. *Psychological Science*, *27*(12), 1588–1599. https://doi.org/10.1177/0956797616666501

Simon, H. A. (1954). Bandwagon and Underdog Effects and the Possibility of Election Predictions. *Public Opinion Quarterly*, *18*(3), 245. https://doi.org/10.1086/266513

Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on Generality (COG): A Proposed Addition to All Empirical Papers. *Perspectives on Psychological Science*, *12*(6), 1123–1128. https://doi.org/10.1177/1745691617708630

Söllner, A., & Bröder, A. (2016). Toolbox or adjustable spanner? A critical comparison of two metaphors for adaptive decision making. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *42*, 215–237. http://dx.doi.org/10.1037/xlm0000162

Söllner, A., Bröder, A., Glöckner, A., & Betsch, T. (2014). Single-process versus multiple-strategy models of decision making: Evidence from am information intrusion paradigm. *Acta Psychologica*, *146*, 84–96.

Söllner, A., Bröder, A., & Hilbig, B. E. (2013). Deliberation versus automaticity in decision making: Which presentation format features facilitate automatic decision making? *Judgment and Decision Making*, *8*(3), 278–298.

https://doi.org/10.1017/S1930297500005982

Spektor, M. S., & Wulff, D. U. (2024). Predecisional information search adaptively reduces three types of uncertainty. *Proceedings of the National Academy of Sciences*, *121*(47). https://doi.org/10.1073/pnas.2311714121

Svenson, O. (1979). Process descriptions of decision making. *Organizational Behavior and Human Performance*, *23*(1), 86–112. https://doi.org/10.1016/0030-5073(79)90048-5

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, *50*(3), 755–769. https://doi.org/10.1111/j.1540-5907.2006.00214.x

Tetlock, P. E., & Kim, J. I. (1987). Accountability and Judgment Processes in a Personality Prediction Task. *Journal of Personality and Social Psychology*, *52*(4), 700–709. https://doi.org/10.1037/0022-3514.52.4.700

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences*, *12*(3), 435–467. https://doi.org/10.1017/S0140525X00057046

Thagard, P. (1998). Coherence as constraint satisfaction. *Cognitive Science*, *22*(1), 1–24. https://doi.org/10.1016/S0364-0213(99)80033-0

Thagard, P. (2000). Probabilistic networks and explanatory coherence. *Cognitive Science Quarterly*, *1*(1), 91–114.

Thagard, P. (2003). Why wasn't O.J. convicted? Emotional coherence in legal inference. *Cognition and Emotion*, *17*(3), 361–383. https://doi.org/10.1080/02699930244000002

Thorngate, W. (1980). Efficient decision heuristics. *Behavioral Science*, *25*(3), 219–225. https://doi.org/10.1002/bs.3830250306

Tormala, Z. L., & Petty, R. E. (2002). What doesn't kill me makes me stronger: The effects of resisting persuasion on attitude certainty. *Journal of Personality and Social Psychology*, *83*(6), 1298–1313. https://doi.org/10.1037/0022-3514.83.6.1298

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

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323. https://doi.org/10.1007/BF00122574

Tversky, A., & Shafir, E. (1992). Choice under Conflict: The Dynamics of Deferred Decision. *Psychological Science*, *3*(6), 358–361. https://doi.org/10.1111/j.1467-9280.1992.tb00047.x

Unkelbach, C., & Greifeneder, R. (2013). A general model of fluency effects in judgment and decision making. In *The Experience of Thinking: How the Fluency of Mental Processes Influences Cognition and Behavior*. https://doi.org/10.4324/9780203078938

Van Bavel, J. J., & Pereira, A. (2018). The Partisan Brain: An Identity-Based Model of Political Belief. *Trends in Cognitive Sciences*, *22*(3), 213–224. https://doi.org/10.1016/j.tics.2018.01.004

van Dongen, N., van Bork, R., Finnemann, A., Haslbeck, J. M. B., van der Maas, H. L. J., Robinaugh, D. J., de Ron, J., Sprenger, J., & Borsboom, D. (2025). Productive explanation: A framework for evaluating explanations in psychological science. *Psychological Review*, *132*(2), 311–329. https://doi.org/10.1037/rev0000479

Van Lange, P. A. M. (2013). What We Should Expect From Theories in Social Psychology. *Personality and Social Psychology Review*, *17*(1), 40–55. https://doi.org/10.1177/1088868312453088

Van Lange, P. A. M., Higgins, E. T., & Kruglanski, A. W. (Eds.). (2022). *Social psychology: Handbook of basic principles* (3rd ed.). Guilford Press.

van Rooij, I., & Blokpoel, M. (2020). Formalizing Verbal Theories. *Social Psychology*, *51*(5), 285–298. https://doi.org/10.1027/1864-9335/a000428

Vandello, J. A., Goldschmied, N. P., & Richards, D. A. R. (2007). The Appeal of the Underdog. *Personality and Social Psychology Bulletin*, *33*(12), 1603–1616. https://doi.org/10.1177/0146167207307488

Wason, P. C. (1968). Reasoning about a Rule. *Quarterly Journal of Experimental Psychology*, *20*(3), 273–281. https://doi.org/10.1080/14640746808400161

Wertheimer, M. (1938). Laws of organization in perceptual forms. In W. D. Ellis (Ed.), *A source book of Gestalt psychology* (pp. 71–94). Hartcourt, Brace and Co. (Original work published 1923).

Westermann, R. (2000). Festinger's Theory of Cognitive Dissonance: A Structuralist Theory-Net. In W. Balzer, J. D. Sneed, & C. U. Moulines (Eds.), *Structuralist Knowledge Representation:Paradigmatic examples* (pp. 189–217). BRILL. https://doi.org/10.1163/9789004457805_011

Westmeyer, H. (2004). Zum Aufstieg und Niedergang der Wissenschaftstheorie in der Psychologie [On the rise and decline of philosophy of science in psychology: A personal construction]. In E. H. Witte (Ed.), *Methodologische, methodische und historische Entwicklungen in der Sozialpsychologie* (pp. 13–34). Pabst Science Publishers.

Wickham, H. (2016a). Data Analysis. In *ggplot2* (Use R!, pp. 189–201). Springer. https://doi.org/10.1007/978-3-319-24277-4_9

Wickham, H. (2016b). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wolf, A., Ounjai, K., Takahashi, M., Kobayashi, S., Matsuda, T., & Lauwereyns, J. (2019). Evaluative Processing of Food Images: Longer Viewing for Indecisive Preference Formation. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00608

Young, M. J., Tiedens, L. Z., Jung, H., & Tsai, M.-H. (2011). Mad enough to see the other side: Anger and the search for disconfirming information. *Cognition and Emotion*, *25*(1), 10–21. https://doi.org/10.1080/02699930903534105

Yu, R., & Chen, L. (2015). The need to control for regression to the mean in social psychology studies. *Frontiers in Psychology*, *5*(1574), 1–8. https://doi.org/10.3389/fpsyg.2014.01574

Zilker, V., & Pachur, T. (2023). Attribute attention and option attention in risky choice. *Cognition*, *236*. https://doi.org/10.1016/j.cognition.2023.105441

Zuckermann, M., Knee, C. R., Hodgins, H. S., & Miyake, K. (1995). Hypothesis confirmation: The joint effect of positive test strategy and acquiescence response set. *Journal of Personality and Social Psychology*, *68*, 52–60.