Dual-Process Theories

Bertram Gawronski, Dillon M. Luke, & Laura A. Creighton

Dual-process theories propose that judgments and behavior can be understood as the product of two (sets of) qualitatively distinct processes, one being characterized by features of automatic processing and the other by features of controlled processing. This chapter provides an overview of dual-process theories in social psychology, integrating both historical and conceptual developments. Distinguishing between three broad classes of dual-process theories, the chapter reviews the most influential examples of (1) domain-specific dual-process theories, which focus on particular phenomena, (2) domain-independent dual-system theories, which identify general principles of information processing, and (3) formalized dual-process theories, which quantify the joint contributions of two distinct processes to behavioral responses. The chapter also discusses critical arguments against each type of dual-process theorizing, which are integrated in a general outlook on future directions.

Keywords: associative processing, attitudes, attribution, automaticity, control, dual-process theories, impression formation, moral judgment, persuasion, prejudice, propositional reasoning, rule-based processing; stereotyping

Introduction

Research in social psychology has been shaped and guided by dual-process theories for almost four decades (Chaiken & Trope, 1999; Sherman, Gawronski, & Trope, 2014). Whereas early dual-process theories focused primarily on domain-specific phenomena (e.g., attitude–behavior relations, attitude change, prejudice and stereotyping, impression formation, dispositional attribution, moral judgment), dual-process theorizing in the past two decades has shifted toward integrative models that aim to identify basic principles of information processing that generalize across content domains. These integrative models can be further divided into (1) domain-independent dual-system theories, which identify two distinct processing systems underlying judgments and behavior and (2) formalized dual-process theories, which quantify the contributions of two distinct processes to behavioral responses by means of mathematical modeling techniques. Although dual-process theories have become the target of increased criticism (e.g., Amodio, 2019; Arkes, 2016; Corneille & Stahl, 2019; De Houwer, 2019; De Neys, in press; Ferguson, Mann, & Wojnowicz, 2014; Keren & Schul, 2009; Kruglanski & Gigerenzer, 2011; Melnikoff & Bargh, 2018; Moors, 2014), it is difficult to imagine what the last four decades of social psychology would have been like without the theoretical guidance provided by dual-process theories.

The current chapter provides an overview of dual-process theories in social psychology with the goal of integrating both historical and conceptual developments since their emergence in the 1980s. Toward this end, we first identify basic features of dual-process theories, which serves as the basis for our review of specific theories. We then describe the most influential domain-specific theories that have set the foundation for the ubiquitous dual-process paradigm within social psychology. Expanding on this review, the following two sections discuss the tenets of domain-independent dual-system theories and formalized dual-process theories, which have gained considerable impact within and beyond social psychology during the past two decades. At the end of each section, we also discuss critical arguments against each type of dual-process theory, which we integrate in our outlook in the final section of this chapter.

What Are Dual-Process Theories?

A central feature of dual-process theories is that they propose two (sets of) qualitatively distinct processes to explain behavior (Gawronski, Sherman, & Trope, 2014; Moors, 2014). The term *process* in this definition can be further specified as the mental operations that transform environmental inputs into behavioral outputs (De Houwer, 2011; Gawronski & Bodenhausen, 2015a; Marr, 1982). Based on this conceptualization, relations between environmental inputs and behavioral outputs represent the phenomena in need of an explanation (*explanandum*) and the two (sets of) processes postulated by dual-process theories are supposed to explain these input-output relations (*explanans*). This conceptualization captures the terms *dual* and *process*. However, the term *dual-process theories* is typically used more narrowly to refer to theories that further assume that the postulated (sets of) processes differ in terms of automaticity features: one is assumed to operate in an automatic fashion, and the other is assumed to operate in a controlled fashion (Gawronski & Creighton, 2013; Moors, 2014).

The term *automatic* can be specified in terms of four questions about the conditions under which a given process operates (see Bargh, 1994; Moors, 2016). Does the process operate (1) when there is no goal to start the process (*unintentionality*), (2) when the amount of invested or available processing resources is small (*efficiency*), (3) when there is a goal to alter or stop the process (*uncontrollability*), and (4) when there is no conscious awareness (*unawareness*)? In the early days

of automaticity research, these features were assumed to occur in an all-or-none fashion (*dual-mode conceptualization*). However, it soon became clear that virtually no process meets all four criteria of automaticity (see Bargh, 1994; Fiedler & Hütter, 2014; Melnikoff & Bargh, 2018; Moors & De Houwer, 2006). Instead, most processes studied within social psychology involve combinations of the four features, requiring more nuance regarding the particular manner in which a given process is assumed to be automatic (*decompositional conceptualization*).

An emerging theme in dual-process theorizing is the importance of distinguishing between *operating principles* and *operating conditions* (Corneille & Stahl, 2019; Deutsch, 2015; Gawronski & Bodenhausen, 2009; Gawronski et al., 2014; Hütter & Rothermund, 2020; Moors, 2014; Sherman, Krieglmeyer, & Calanchini, 2014). Whereas the operating principles of a given process specify *how* the process translates environmental inputs into behavioral outputs, its operating conditions specify *when* the proposed translation is assumed to occur. This distinction is important for two reasons. First, whereas statements about operating principles are definitions that are true by scientific convention (i.e., they define what the process is), statements about operating conditions are empirically testable hypotheses about states of affairs (i.e., they state when the process should operate; see Deutsch, 2015; Gawronski et al., 2014). Second, theoretical statements about automaticity features specify operating conditions, not operating principles (Gawronski et al., 2014; Moors, 2016). Specifically, hypotheses about automaticity features state whether a given process should operate when there is no goal to start the process, when the amount of invested or available processing resources is small, when there is a goal to alter or stop the process, and when there is no conscious awareness. However, any such hypotheses do not specify how the process translates environmental inputs into behavioral outputs.

The emphasis on automaticity features in dual-process theories has sometimes led to misunderstandings that the umbrella terms *automatic* and *controlled* would characterize the intrinsic nature of the postulated processes (see Gawronski et al., 2014). However, a conceptually sound dual-process theory should specify both operating principles and operating conditions, and their respective specifications should be conceptually independent to avoid tautological hypotheses about the mapping between specific processes and their presumed automaticity features (Moors, 2014). For example, if a dual-process theory distinguishing between System-1 and System-2 processing states that System-1 processing is efficient and System-2 processing is inefficient without specifying how System-1 and System-2 processing translate environmental inputs into behavioral outputs (e.g., Dhar & Gorlin, 2013; Evans & Stanovich, 2013), the explanations provided by the theory would be circular. Resource-independent effects of environmental inputs on behavioral outputs would be explained in terms of System-1 processing and resource-dependent effects of environmental inputs on behavioral outputs would be explained in terms of System-2 processing. However, the only evidence for the involvement of two processes would be the differential resource-dependence that needs to be explained (see Arkes, 2016; Gawronski, 2013, Osman, 2013).

To the extent that dual-process theories include clear specifications of both operating principles and operating conditions that are (1) conceptually independent and (2) empirically identifiable, they can face two kinds of empirical challenges that make them "falsifiable" (Moors, 2014). First, a theory may be challenged by empirical data showing that the input-output relations attributed to one of the two processes occurs under conditions that conflict with the presumed operating conditions of that process. For example, a dual-process theory may explain correct solutions to logical problems in terms of a cognitively effortful process of rule-based processing, in contrast with a cognitively effortless process of associative processing (e.g., Evans, 2010). Yet, empirical data may suggest that performance in solving logical problems is unaffected by otherwise effective manipulations of cognitive resources (see De Neys & Pennycook, 2019). Such evidence would pose a challenge to the theory's hypotheses about operating conditions by suggesting that rule-based processing may be efficient rather than inefficient. However, it would not necessarily challenge the theory's ontological assumptions about the existence of the two processes.

Second, a theory may be challenged by empirical data showing that the input-output relations attributed to one of the two processes do not emerge at all. For example, a dual-process theory may explain effects of repeated stimulus co-occurrence in terms of an associative learning mechanism involving the formation of mental associations between the co-occurring stimuli, in contrast with a propositional learning mechanism involving the generation of mental propositions about the relation between co-occurring stimuli (e.g., Gawronski & Bodenhausen, 2006). Yet, empirical data may suggest that judgments and decisions are exclusively driven by specific relations between co-occurring stimuli (e.g., A starts vs. stops B) rather than their mere co-occurrence (see Corneille & Stahl, 2019). Such evidence would pose a more fundamental challenge to the theory, because it would question the theory's ontological assumptions about the existence of one of the two processes.

To further clarify the basic features of dual-process theories, it is also helpful to consider their potential theoretical alternatives. First, observed relations between environmental inputs and behavioral outputs may be explained by theories that postulate only one unitary process (*single-process theories*). According to such theories, the dichotomies of the four automaticity features (i.e., unintentional-intentional, efficient-inefficient, uncontrollable-controllable, unconscious-conscious) do not map onto two qualitatively distinct (sets of) processes. Instead, the operating conditions captured by the four features of automaticity are assumed to moderate the behavioral outputs produced by a single process in a parametric fashion. An illustrative example is the single-process hypothesis that cognitive resources merely influence the amount and complexity of considered information rather than the operation of two qualitatively distinct processes (see Arkes, 2016; De Houwer, Van Dessel, & Moran, 2020; Kruglanski & Thompson, 1999; Osman, 2013). Second, observed relations between environmental inputs and behavioral outputs may be explained by theories that postulate more than two processes (*multi-process theories*). According to such theories, human behavior is the product of multiple qualitatively distinct processes, each of which may be characterized by different features of automaticity (e.g., Amodio, 2019; Sherman et al., 2008). Finally, some theories explain relations between environmental inputs and behavioral outputs in terms of two distinct types of representation in memory (*dual-representation theories*). To the extent that these theories additionally postulate two qualitatively distinct processes operating on these representations, they would qualify as dual-process theories in terms of the above conceptualization (see Gawronski et al., 2014).[1] However, numerical hypotheses about processes and representations are independent, in that a theory may propose one process that operates on one type of representation (e.g., De Houwer et al., 2020), two processes that operate on one type of representation (e.g., Gawronski & Bodenhausen, 2006), one process that operates on two types of representation (e.g., Wilson, Lindsey, & Schooler, 2000), or two processes that operate on two types of representation (e.g., Smith & DeCoster, 2000). The current chapter focuses specifically on dual-process theories—that is, theories that postulate two qualitatively distinct (sets of) processes—independent of their postulated number of representations.

Domain-Specific Dual-Process Theories

Early dual-process theories tended to be domain-specific in the sense that they focused on particular phenomena in need of social-cognitive explanations. In the current section, we first review the core assumptions of the most influential theories of this kind and then discuss criticism that has been raised against domain-specific dual-process theories.

Attitude-Behavior Relations

One highly influential class of dual-process theories describes the mechanisms by which attitudes guide behavior. These theories have been inspired by recurring debates about whether and to what extent attitudes influence behavior (e.g., Wicker, 1969). By shifting the focus from the question of *do attitudes guide behavior?* to the question of *how do attitudes guide behavior?*, dual-process theorizing provided important insights into the conditions under which attitudes do and do not influence behavior.

MODE Model. The most prominent dual-process theory in this area is Fazio's (1990) *M*otivation and *O*pportunity as *DE*terminants (MODE) model, which specifies two distinct processes by which attitudes can guide behavior depending on the person's motivation and opportunity to engage in deliberate processing (for a review, see Fazio & Olson, 2014). A central component of the MODE model is the definition of *attitude* as the mental association between an object and a person's summary evaluation of that object (Fazio, 1995, 2007). To the extent that this association is sufficiently strong, the evaluation associated with the object may be activated automatically upon encountering that object (i.e., without intention to evaluate the object; see Fazio, Sanbonmatsu, Powell, & Kardes, 1986). Such automatically activated attitudes are assumed to influence an individual's spontaneous interpretation of the current situation, which in turn will guide the individual's behavior without them necessarily being aware of the attitude's influence (*spontaneous attitude–behavior process*). Alternatively, individuals may scrutinize specific attributes of the object and the current situation (*deliberate attitude–behavior process*). However, such deliberate analyses require that individuals have both the motivation and the opportunity (i.e., adequate time and cognitive resources) to engage in effortful information processing. Thus, to the extent that either the motivation or the opportunity to engage in effortful processing is low, automatically activated attitudes may guide behavior through their effect on the spontaneous construal of the current situation. However, if both the motivation and the opportunity to engage in effortful

[1] We use the term *dual-process theories* to refer to theories that postulate two distinct mental operations that translate inputs into outputs, and the term *dual-representation theories* to refer to theories that postulate two distinct formats in which information is stored in memory (see Gawronski et al., 2014).

processing are high, the impact of automatically activated attitudes on behavior will depend on particular aspects of the current situation, including specific attributes of the attitude object or salient norms (see Fazio, 1990).

By virtue of its assumptions about the processes underlying attitude-behavior relations, the MODE model also accounts for dissociations between explicit and implicit measures of attitudes (for a review, see Gawronski & Brannon, 2019). According to the MODE model, a central feature of implicit measures—such as evaluative priming (Fazio, Jackson, Dunton, & Williams, 1995) or the implicit association test (Greenwald, McGhee, & Schwartz, 1998)—is that they reduce participants' opportunity to engage in effortful processing. Hence, evaluative responses on implicit measures should reflect automatically activated attitudes. In contrast, verbally reported evaluations assessed by explicit measures are relatively easy to control. Thus, if either the motivation or the opportunity to engage in effortful processing is low, explicit measures should reveal the same automatically activated attitudes captured by implicit measures. If, however, both the motivation and the opportunity to engage in deliberate processing are high, explicit measures should reflect whatever evaluation is suggested by the outcome of deliberate inferences. Together with the MODE model's core assumptions about attitude-behavior relations, these hypotheses imply that implicit measures should be better predictors of spontaneous behavior, whereas explicit measures should be better predictors of deliberate behavior. These predictions have been confirmed in numerous studies on a wide range of topics (for a review, see Friese, Hofmann, & Schmitt, 2008).

Attitude Formation and Change

Expanding on the question of how attitudes guide behavior, another influential class of domain-specific dual-process theories is concerned with the formation and change of attitudes. Two prominent examples are the elaboration likelihood model (ELM; Petty & Cacioppo, 1986) and the heuristic systematic model (HSM; Chaiken, 1987), which identity the conditions under which different aspects of a persuasive message (e.g., strength of arguments, attractiveness of the source) influence the effectiveness of persuasive appeals. A more recent example is the associative-propositional evaluation (APE) model (Gawronski & Bodenhausen, 2006, 2011), which identifies factors leading to changes in evaluations captured by implicit measures, changes in evaluations captured by explicit measures, and changes in both.

Elaboration Likelihood Model. The central notion of Petty and Cacioppo's (1986) ELM is that attitude change occurs along an elaboration continuum whereby persuasion is determined by how motivated and able an individual is to engage in effortful information processing (for a review, see Teeny, Briñol, & Petty, 2016). The basic assumption is that the higher an individual's cognitive elaboration, the more likely they are to process all object-relevant information. At the high end of the elaboration continuum, people assess all of the available object-relevant information (e.g., strength of the presented arguments) and integrate this information with their stored knowledge in order to obtain a carefully considered evaluation (*central route*). Conversely, at the low end of the elaboration continuum, people engage in considerably less scrutiny of object-relevant information (*peripheral route*). When elaboration is low, attitude change can be effected from a cursory examination of the available information (e.g., by examining only a subset of the available information) or by the use of heuristics and other types of information processing shortcuts (e.g., *I agree with people I like*). Compared with attitudes that are changed through the central route, attitudes changed through the peripheral route are assumed to be relatively weak, susceptible to counterpersuasion, and less predictive of behavior.

A seminal finding predicted by the ELM is that high elaboration increases the impact of primary features of a persuasive message (e.g., argument quality), whereas low elaboration increases the impact of secondary aspects (e.g., source-related features). Consistent with these hypotheses, numerous studies found that, under conditions of low elaboration, participants were more persuaded by a counterattitudinal message when it was presented by a source with "high-value" characteristics (e.g., high attractiveness, high expertise) than when it was presented by a source with "low-value" characteristics (e.g., low attractiveness, low expertise), with argument strength having little impact. Conversely, under conditions of high elaboration, participants were more persuaded by a counterattitudinal message when the arguments were strong than when they were weak, with source characteristics having little impact (e.g., Petty, Cacioppo, & Schumann, 1983).

Heuristic-Systematic Model. Similar to Petty and Cacioppo's (1986) ELM, Chaiken's (1987) HSM describes two basic persuasion processes that may guide an individual's judgments of an attitude object (for a review, Chaiken & Ledgerwood, 2012). *Systematic processing* involves comprehensive consideration of object-relevant information (e.g., argument strength), which requires high levels of motivation and ability to engage in effortful processing. *Heuristic processing*, in contrast, relies on the activation, accessibility, and applicability of learned heuristics that require relatively few cognitive resources (e.g., *I agree with people I like*). According to

the HSM, the likelihood that an individual engages in systematic processing is guided by the *sufficiency principle*, which states that the motivation to engage in systematic processing increases with the extent to which an individual's desired level of confidence falls below their actual level of confidence. That is, individuals are more likely to engage in systematic processing when the difference between their desired and actual levels of confidence is large. Conversely, people are more likely to engage in heuristic processing when the difference between their desired and actual levels of confidence is small. Importantly, systematic processing may not necessarily lead to unbiased judgments, because systematic processing can be influenced by *defense motivation* and *impression motivation* instead of *accuracy motivation.* Whereas defense motivation refers to the desire to defend preexisting attitudes, impression motivation refers to the desire to hold attitudes that satisfy specific social goals.

Another central assumption of the HSM is that heuristic and systematic processing may co-occur and interact with each other to exert either independent or interdependent effects on evaluations. First, according to the model's *attenuation hypothesis*, systematic processing can completely override the effects of heuristic processing (e.g., Maheswaran & Chaiken, 1991). Such attenuation effects are likely to occur when systematic processing yields information that contradicts the validity of simple persuasion heuristics (e.g., strong arguments presented by an unattractive source, weak arguments presented by a source with high expertise). Second, the information generated by heuristic and systematic processing may jointly influence evaluations in an additive manner (e.g., Maheswaran, Mackie, & Chaiken, 1992). According to the model's *additivity hypothesis,* such effects are likely to occur when the two processing modes do not yield conflicting reactions (e.g., weak arguments presented by an unattractive source, strong arguments presented by a source with high expertise). Finally, when the message content is ambiguous, heuristic cues may bias the effects of systematic processing, as described by the model's *bias hypothesis*. For example, if the strength of a persuasive argument is ambiguous, the argument may be perceived as more convincing if it is presented by an expert than if it is presented by a layperson (e.g., Chaiken & Maheswaran, 1994).

Although the HSM and the ELM were developed independently at about the same time, the two theories have considerable overlap in terms of their core assumptions. For example, both models maintain that attitude change can occur through either (1) systematic/central processing that requires some degree of motivation and capacity or (2) heuristic/peripheral processing that is assumed to require little motivation or capacity. However, the two models differ in their assumptions about the relation between the proposed processes (Chen & Chaiken, 1999). Whereas the ELM assumes that more pronounced central processing is associated with less pronounced peripheral processing (and vice versa), the HSM assumes that systematic and heuristic processing can occur simultaneously, with either independent or interactive effects. Hence, the ELM holds that there is a trade-off between peripheral and central processing, such that the impact of one processing mode decreases as the impact of the other processing mode increases. In contrast, under the HLM's conceptualization, individuals can engage in systematic and heuristic processing simultaneously.

Associative-Propositional Evaluation Model. Although both the ELM and the HSM have been highly influential, a major limitation of the two theories is that they do not distinguish between attitude change on implicit and explicit measures. This feature makes them unable to explain discrepant patterns of change on the two kinds of measures (e.g., change on implicit but not explicit measures, change on explicit but not implicit measures). The associative-propositional evaluation (APE) model has been designed to fill this gap (Gawronski & Bodenhausen, 2006, 2011). The theoretical core of the APE model is the distinction between associative and propositional processes. The theory defines associative processes as the *activation* of associations in memory, driven by principles of similarity and contiguity. Propositional processes are defined as the *validation* of the information implied by activated associations, which is assumed to be guided by principles of cognitive consistency. To the extent that the set of information implied by activated associations is consistent, it will be used for judgments and behavioral decisions. If, however, the set of information implied by activated associations is inconsistent, aversive feelings of dissonance will induce a tendency to resolve the dissonance-provoking inconsistency before a judgment or behavioral decision is made (see Festinger, 1957). In such cases, inconsistency may be resolved either by rejecting one of the propositions within the set of inconsistent information or by searching for an additional proposition that resolves the inconsistency. To the extent that the inconsistency is resolved by rejecting one of the involved propositions, activated associations and endorsed propositional beliefs are assumed to diverge, because mere negation of a proposition (e.g., *it is not true that old people are bad drivers*) may not necessarily deactivate the mental associations underlying that proposition (i.e., the association between the concepts *old people* and *bad drivers*).

According to the APE model, evaluations captured by explicit measures reflect the proximal outcome of propositional processes, whereas evaluations captured

by implicit measures reflect the proximal outcome of associative processes. Yet, associative and propositional processes are assumed to interact with one another in that (1) associative activation processes provide the informational basis for propositional validation processes and (2) propositional validation processes can influence the activation of associations under specific conditions. According to the theory, associative processes should influence evaluations captured by explicit measures in a distal manner when activated associations are accepted as valid, but not when they are rejected as invalid. Conversely, propositional processes should influence evaluations captured by implicit measures in a distal manner when propositional inferences affirm the validity of new information, but not when they negate the validity of activated information. By virtue of these assumptions, the APE model provided novel predictions about the conditions under which a given factor should produce (1) changes on explicit but not implicit measures, (2) changes on implicit but not explicit measures, or (3) corresponding changes on explicit and implicit measures (e.g., Gawronski & LeBel, 2008; Gawronski & Strack, 2004; Whitfield & Jordan, 2009). According to the APE model, the specific pattern of change is determined by two questions: (1) Which of the two processes is directly influenced by a given factor, associative activation or propositional validation? (2) Do directly induced changes in one process lead to indirect changes in the other process?

Prejudice and Stereotyping

One of the most striking findings in research on prejudice and stereotyping is that public opinion polls in North America showed a steady decline in negative evaluations of racial minority groups after World War II, whereas racial conflicts showed only a moderate reduction (e.g., Greeley & Sheatsley, 1971; Taylor, Sheatsley, & Greeley, 1978). This discrepancy inspired social psychologists to postulate more subtle forms of racial prejudice, such as modern (McConahay, 1986), aversive (Gaertner & Dovidio, 1986), or symbolic (Sears, 1988) racism. The general notion underlying these constructs is that racial prejudice has simply changed its face, rather than been abandoned. A similar idea provided the inspiration for Devine's (1989) dissociation model, which was seminal in introducing the distinction between automatic and controlled processes to research on prejudice and stereotyping.

Dissociation Model. A central aspect of Devine's (1989) dissociation model is the distinction between the *knowledge* of a social stereotype and the *belief* in the accuracy of that stereotype. According to Devine, both low-prejudice and high-prejudice individuals tend to be familiar with the contents of prevailing cultural stereotypes. However, the two groups differ with respect to their personal beliefs about the accuracy of these stereotypes. To the extent that stereotypic knowledge is acquired during early childhood and highly overlearned through socialization, stereotypic knowledge is assumed to be activated automatically upon encountering members of stereotyped groups, and this occurs for both low-prejudice and high-prejudice individuals. In contrast, the rejection of stereotypic knowledge is assumed to be the result of egalitarian, nonprejudicial beliefs, which tend to be acquired later in the socialization process. Because these beliefs are less overlearned than earlier acquired stereotypic knowledge, suppressing the impact of automatically activated stereotypes in favor of egalitarian, nonprejudicial beliefs requires the operation of controlled processing. In other words, while the model assumes that automatic stereotype activation is equally strong and inescapable for both high-prejudice and low-prejudice individuals, the two groups differ at the level of controlled processing, such that low-prejudice but not high-prejudice individuals suppress the impact of automatically activated stereotypes with egalitarian, nonprejudicial beliefs. In terms of the four features of automaticity, these assumptions imply that the activation of social stereotypes occurs unintentionally, even though their impact on overt behavior can be controlled through effortful processes.

A notable aspect of Devine's (1989) dissociation model is that it implies a rather different view on the roles of personal beliefs and social contexts than the MODE model. Whereas the MODE model assumes that personal attitudes tend to be activated automatically and that the overt expression of these attitudes is sometimes suppressed when they conflict with social norms (Fazio et al., 1995), the dissociation model proposes that socially transmitted stereotypes are activated automatically and that the overt expression of these stereotypes is suppressed when they conflict with personal beliefs (Devine, 1989). In other words, whereas the MODE model locates an individual's "true self" at the level of automatic processes and extrinsic, social influences at the level of controlled processes, the dissociation model locates extrinsic, social influences at the level of automatic processes and the individual's "true self" at the level of controlled processes. Even though questions about what should be considered the "true self" are philosophical rather than empirical, these diverging views have important implications for the interpretation of automatic stereotypic biases. One example is the automatic tendency to misidentify harmless objects as weapons when they are held by a black person rather than a white person (for a review, see Payne & Correll, 2020). According to the MODE model, such unintentional errors reveal an individual's personal attitudes when the individual does not have the opportunity to adjust their automatic responses to egalitarian norms. In contrast, from the perspective of

Devine's (1989) model, unintentional errors in weapon identification reveal the ubiquitous influence of cultural stereotypes that may not necessarily reflect the individual's personal beliefs.

Impression Formation

Similar to Devine's (1989) dissociation model of prejudice and stereotyping, dual-process theories of impression formation emphasize the role of social category information in early processing stages. However, whereas Devine's model focuses particularly on the unintentional activation versus controlled suppression of stereotypes, dual-process theories of impression formation specify the conditions under which personal impressions of an individual are dominated by category-related or person-specific information.

Continuum Model. Fiske and Neuberg's (1990) continuum model of impression formation proposes that the processes by which people form opinions of other individuals operate along a continuum that reflects the degree to which perceivers utilize category-related versus person-specific information. The basic assumption of the model is that category information enjoys general priority because the processing of such information does not require substantial amounts of cognitive resources. Specifically, perceivers are assumed to categorize individuals based on salient category cues (e.g., gender, age, ethnicity), and this categorization is assumed to occur unintentionally upon encountering a target individual. Contingent on the relevance of the target for the perceiver's momentary goals, perceivers will direct their attention to individual attributes of the target, thereby moving toward the more thoughtful end of the processing continuum. If the target is judged to be irrelevant to the perceiver's momentary goals, the final impression of the target will be based exclusively on the initial categorization. If, however, the target is judged to be relevant to the perceiver's momentary goals, the perceiver will attempt to integrate person-specific attributes into a coherent impression of the target.

Overall, the continuum model assumes that perceivers attempt to maintain the impression implied by their initial categorization while processing individual attributes of the target. To the extent that the additional information is interpreted to be consistent with the initially identified category, the final impression of the target will be based on the initial categorization. If, however, the additional information is inconsistent with the initial categorization, perceivers will attempt to recategorize the target in an attempt to find a more suitable category than the initial one. For example, if person-specific attributes of a target individual seem inconsistent with the impression implied by his or her category membership, perceivers may use subtypes to assign the target to a more appropriate category than the initial, general category (Richards & Hewstone, 2001). If this recategorization process successfully integrates the available information about the target, the final impression will be based on this newly applied category. However, if the attempt to recategorize the target fails, perceivers are assumed to move on to a process of piecemeal integration, in which they engage in an attribute-by-attribute assessment of individual characteristics of the target. Yet, according to Fiske, Lin, and Neuberg (1999), such piecemeal integration occurs quite rarely, given that perceivers tend to construct ad hoc theories to account for contradictory information in the initial stages of the impression formation continuum (e.g., Kunda, Miller, & Claire, 1990; Leyens & Yzerbyt, 1992).

Dual-Process Model. Whereas Fiske and Neuberg's (1990) continuum model attributes a dominant role to top-down, category-based processing, Brewer's (1988) dual-process model argues that impression formation may take either a top-down or a bottom-up route (see also Brewer & Harasty-Feinstein, 1999). Both routes are assumed to start with an automatic identification of salient features of the stimulus person. This processing step can be described as the mere recognition of feature configurations (e.g., male, dark skin color, business suit). To the extent that the target is irrelevant to the perceiver, the processing sequence is assumed to remain at this level without further processing of category-related or person-related implications of the identified features. If, however, the target is relevant to the perceiver, further processing of the identified features can take either a top-down or a bottom-up route depending on the relative involvement of the perceiver.

Bottom-up processing is assumed to occur under conditions of high involvement, in which perceivers are assumed to adopt an interpersonal orientation. In this person-based processing mode, perceivers are assumed to draw inferences directly from an individual's identified features, which are integrated into a coherent impression of the target (*personalization*). Depending on the motivation and ability to engage in effortful processing, this person-based impression may be more or less complex. In other words, personalization is not an effortful process per se. Rather, the degree of cognitive elaboration is assumed to influence the complexity of the final impression, such that low elaboration will lead to relatively simple person-based impressions, whereas high elaboration will lead to relatively complex person-based impressions.

Top-down processing is assumed to occur under conditions of low involvement, in which perceivers are assumed to adopt an intergroup orientation. In this category-based processing mode, the target is initially categorized based on salient features (e.g., black

businessman). This categorization process, in turn, may activate stereotypic contents associated with the applied category, which serve as a filter for the integration of other target-specific information. Whereas target-specific information that is related to the stereotypic content of the category will be integrated into a coherent impression, target-specific information that is unrelated to the category stereotype will be ignored. To the extent that the category-related target information is consistent with the category stereotype, the process is assumed to stop at this point, leading to a target impression in line with the category stereotype (*stereotyping*). If, however, the category-related target information is inconsistent with the category stereotype, the inconsistency has to be resolved in order to achieve a coherent impression of the target. The result of the latter process is an individuated impression of the target, which is based on a systematic integration of target-specific information (*individuation*). However, this integration is still regarded as a category-based process, given that the initial categorization of the target serves as a filter for the processing of category-related versus category-unrelated target information. Thus, like person-based processing, category-based processing can be more or less effortful, such that stereotyping is the likely outcome of low elaboration, whereas individuation usually requires high elaboration.

Dispositional Attribution

Another important question in the context of impression formation is how perceivers make sense of other people's behavior. To describe the processes that underlie inferences from observed behavior, social psychologists in the 1960s proposed various theories of causal (e.g., Kelley, 1967) and dispositional (e.g., Jones & Davis, 1965) attribution. However, deviating from the predictions of these models, empirical research soon demonstrated that perceivers tend to give more weight to dispositional compared with situational factors (e.g., Jones & Harris, 1967; Ross, Amabile, & Steinmetz, 1977). This tendency to overestimate the role of dispositional compared with situational factors has become known as the *fundamental attribution error* (Ross, 1977). A particular instantiation of the fundamental attribution error is the *correspondence bias* (Gilbert & Malone, 1995), which is defined as the tendency to draw correspondent dispositional inferences from observed behavior even if the behavior is constrained by situational factors (for a discussion of conceptual differences between the fundamental attribution error and the correspondence bias, see Gawronski, 2004). In the 1970s, the discrepancy between theoretically derived predictions and empirical results led to the odd situation that the models that had originally been designed to explain perceivers' inferences acquired a normative status, such that empirically observed deviations were depicted as judgmental biases or errors instead of counterevidence against the proposed theories (see Trafimow, 2015). This situation did not change until the emergence of dual-process theories in the 1980s. These theories turned attention back to identifying the processes that underlie perceivers' inferences, with a particular focus on explaining when and why the correspondence bias occurs.

Three-Stage Model. One such dual-process theory is Gilbert's (1989) three-stage model of dispositional inference. According to this model, dispositional inferences involve three sequential processing steps that are claimed to require different amounts of cognitive resources: (1) *behavioral categorization* (i.e., what is the actor doing?), (2) *dispositional characterization* (i.e., what disposition does the behavior imply?), and (3) *situational correction* (i.e., what situational determinants might have caused the behavior?). Whereas behavioral categorization and dispositional characterization are assumed to occur unintentionally without requiring large amounts of cognitive resources, situational correction is assumed to be an intentional, relatively effortful process. Applied to the correspondence bias, these assumptions imply that the tendency to draw correspondent dispositional inferences from situationally constrained behavior should be lower when perceivers have both the motivation and the cognitive capacity to engage in the effortful process of situational correction. However, the tendency to commit the correspondence bias should be enhanced when either the motivation or the cognitive capacity to engage in effortful processing is low. These predictions have been confirmed in several studies that investigated effects of processing motivation (e.g., Vonk, 1999) and cognitive capacity (e.g., Gilbert, Pelham, & Krull, 1988) on dispositional inferences from situationally constrained behavior.

An important extension of Gilbert's (1989) three-stage model was proposed by Krull (1993), who merged Gilbert's (1989) assumptions about the effortfulness of situational correction with previous research on judgmental anchoring in dispositional inference (Quattrone, 1982). Deviating from Gilbert's (1989) assumption that social inferences follow a fixed sequence, Krull (1993) argued that the particular sequence of processes depends on the inferential goal of the perceiver. According to Krull, perceivers interested in inferring an actor's disposition will (1) categorize the behavior, (2) characterize a corresponding disposition, and then (3) correct these characterizations for situational constraints. If, however, perceivers are interested in the causal impact of situational factors, they will (1) categorize behavior, (2) characterize the situation, and then (3) correct these characterizations for dispositional factors. In other

words, the contents of both the characterization and the correction stage are assumed to depend on the inferential goal of the perceiver. Because correction processes are assumed to require more capacity compared with characterization processes, motivation and cognitive capacity to engage in effortful processing should have different effects on social inferences as a function of perceivers' inferential goals. Specifically, reduced cognitive elaboration should increase the tendency to commit the correspondence bias when perceivers have the goal of inferring dispositional characteristics of the actor. In contrast, reduced cognitive elaboration should have the opposite effect when perceivers are interested in characteristics of the situation (e.g., Krull, 1993; Krull & Erickson, 1995).

Two-Stage Model. Another influential theory that aims to identify the processes underlying dispositional attributions is Trope's (1986) two-stage model. According to this theory, trait judgments are the product of two sequential processes, which are described as *identification* and *inference*. At the identification stage, perceivers categorize momentarily available cues in trait-relevant terms. These cues may be related to the actor's behavior (*behavioral cues*), the situational context of the behavior (*situational cues*), or the actor's personal characteristics or group membership (*prior cues*). For example, a person's behavior might be categorized as anxious, the situational context as eliciting anxious behavior, and the actor as belonging to a stereotypically anxious group. To the extent that the relevant cues within each of the three dimensions are unambiguous, they fully constrain their corresponding categorizations. That is, behavioral cues fully determine the categorization of the behavior, situational cues fully determine the categorization of the situation, and prior cues fully determine the categorization of the actor. If, however, a particular cue is ambiguous, its categorization may be biased by contextual cues in an assimilative manner. For example, if an actor's behavior is ambiguously anxious, it may be perceived as more anxious when the situational context is known to be anxiety provoking (e.g., Snyder & Frankel, 1976) or when the actor belongs to a stereotypically anxious group (e.g., Kunda & Sherman-Williams, 1993). Similar effects may occur for the categorization of ambiguous situational cues (e.g., Trope & Cohen, 1989) and the categorization of ambiguous prior cues (e.g., Hugenberg & Bodenhausen, 2004), both of which may be biased by unambiguous behavioral cues. Because perceivers tend to consider their subjective categorizations "as perceptual givens rather than as context-derived" (Trope & Gaunt, 1999, p. 170), deliberate correction of such biased perceptions is rather unlikely, even when the validity of the biasing contextual information is discredited afterward (e.g., Trope & Alfieri, 1997). In combination with studies showing biasing effects of contextual cues under conditions of depleted cognitive resources (e.g., Trope & Alfieri, 1997; Trope & Gaunt, 2000), these findings suggest that the processes involved at the identification stage operate unintentionally, efficiently, and outside of conscious awareness (for a review, see Trope & Gaunt, 1999).

Once behavioral, situational, and prior cues have been categorized, the outputs of the identification stage serve as inputs for more or less deliberate dispositional inferences. At this stage, perceivers' categorizations of the behavior, the situation, and the actor are integrated into a unified judgment of the actor's disposition. However, in contrast to the generally assimilative nature of contextual influences at the identification stage, dispositional judgments at the inferential stage are influenced by the three kinds of information in different ways. Whereas behavioral and prior information influence trait judgments in a positive direction, situational information has a subtractive effect. For example, behavior that is categorized as anxious will facilitate correspondent inferences of dispositional anxiety. Similarly, categorization of the actor as a member of a stereotypically anxious group will also promote inferences of dispositional anxiety. Situational cues identified to elicit anxious behavior, however, should have a negative effect on inferences of dispositional anxiety. Such information should discount the informational value of the other two dimensions, and hence reduce correspondent inferences of dispositional anxiety (see Kelley, 1972). Importantly, whereas contextual influences at the identification stage are assumed to operate efficiently, unintentionally, and outside of conscious awareness, the integration of the different kinds of information at the inference stage is assumed to be a conscious, intentional process that requires varying amounts of cognitive resources depending on the salience of the three kinds of information (Trope & Gaunt, 2000).

Like Gilbert's (1989) three-stage model, Trope's (1986) two-stage model implies that the tendency to draw correspondent dispositional inferences from situationally constrained behavior should be more pronounced when cognitive elaboration is low than when it is high (but see Trope & Gaunt, 2000). However, within Trope's model, situational information can have two antagonistic effects on correspondent dispositional inferences: (1) a *direct negative effect* similar to the situational correction process in Gilbert's model and (2) an *indirect positive effect* that is mediated by biased categorizations of ambiguous behavior. For example, if an actor behaves anxiously, knowledge that the situational context is anxiety provoking should discount the informational value of anxious behavior, and hence *reduce* correspondent inferences of dispositional anxiety (i.e.,

direct negative effect). Yet, at the same time, the actor's behavior may be perceived as more anxious when the situational context is known to be anxiety provoking, and perceptions of stronger behavioral anxiety should *enhance* correspondent inferences of dispositional anxiety (i.e., indirect positive effect). Importantly, if the two effects occur at the same time, their effects on dispositional attributions may cancel each other out, leading perceivers to draw strong correspondent inferences from situationally constrained behavior even when they consider situational factors to have a strong impact on the target's behavior. In other words, they commit the correspondence bias, even though they do not commit the fundamental attribution error (see Gawronski, 2004).

Moral Judgment

A central question in moral psychology concerns the processes underlying moral judgments. Historically, research on this question has been dominated by rationalist theories, which place a strong emphasis on higher-order reasoning. This tradition is prominently reflected in Kohlberg's (1969, 1976) cognitive-developmental theory. According to this theory, moral judgments result from controlled processes that develop as a function of age according to a sequence of six stages, beginning with egocentric concerns and culminating in universal principles of justice. While heavily influential, the turn of the century brought a challenge to Kohlberg's theory in the form of dual-process theories suggesting that automatic processes play a central role in the generation of moral judgments.

Social Intuitionist Model. In contrast to rationalist approaches, Haidt's (2001) Social Intuitionist Model (SIM) holds that moral judgments are the product of moral intuitions, described as "the automatic output of an underlying, largely unconscious set of interlinked moral concepts" (p. 1040). According to this view, moral judgments arise from a highly efficient affective process that operates unintentionally and outside of conscious awareness. In this way, moral judgments are said to be similar to perception in the sense that they arise without awareness of the processes that produced them. Moral intuitions are claimed to be partly innate and partly shaped by culture, involving different dimensions such as care/harm and loyalty/betrayal (see Moral Foundations Theory; Graham et al., 2013).

While moral intuitions are assumed to provide the basis for moral judgments, moral reasoning is said to play a relatively minor role, primarily defending one's existing intuitions by means of post-hoc justifications (Haidt, 2001). Although these justifications typically reinforce one's own intuitive moral judgments (*post-hoc reasoning*), they can activate new intuitions in others and thereby influence their judgments (*reasoned persuasion*). Thus, reasoning is assumed to be most influential in interpersonal contexts like moral discussions in which individuals exchange reasons for their respective judgments. Persuasion in these contexts may be aided if individuals like one another or hindered if individuals dislike one another (*social persuasion*). While reasoning may primarily operate interpersonally, the SIM still allows for moral reasoning to influence one's own moral judgments by either (1) directly overriding existing intuitions (*reasoned judgment*) or (2) activating new moral intuitions (*private reflection*). However, these kinds of intrapersonal influences are assumed to be quite rare, in that they are limited to cases of conflicting moral intuitions and individuals who are highly motivated and trained to do so (e.g., moral philosophers).

Dual-Process Model. Integrating aspects of both intuitionist and rationalist approaches, Greene's (2008, 2014) dual-process model (DPM) aims to identify the processes underlying responses in moral dilemmas. The DPM was famously inspired by the trolley dilemma, a philosophical scenario in which a runaway trolley is set on a collision course with five railroad workers. In a variant called the *switch dilemma*, it is possible to pull a lever and divert the trolley to a different track, killing one worker instead of five (Foot, 1967). In a variant called the *footbridge dilemma*, it is possible to push a large man off a footbridge into the path of the trolley, killing the man but preventing the death of the five workers (Thomson, 1976). Judgments of these actions as acceptable have been described as *characteristically utilitarian* in the sense that they are consistent with maximizing overall welfare (i.e., saving the lives of five by sacrificing one; see Conway, Goldstein-Greenwood, Polacek, & Greene, 2018). In contrast, judgments of these actions as unacceptable have been described as *characteristically deontological* in the sense that they are consistent with notions of moral norms and duties (e.g., prohibition against killing others; see Conway et al., 2018). Past research (e.g., Greene, Sommerville, Nystrom, Darley, & Cohen, 2001) indicates that people are more likely to make utilitarian judgments in the switch dilemma (i.e., find it acceptable to pull the switch) and more likely to make deontological judgments in the footbridge dilemma (i.e., find it unacceptable to push the large man).

The DPM explains these findings by postulating two distinct processes underlying moral judgments. Whereas deontological judgments are assumed to arise from automatic emotional reactions to the idea of causing harm, utilitarian judgments are assumed to be the product of controlled cognitive analyses of costs and benefits for overall welfare. Although both variants of the trolley dilemma involve the same trade-off between the lives of one person versus five, the footbridge dilemma is different from the switch dilemma in that the focal action requires personal force to obtain the same outcome. Consequently, the footbridge dilemma

is said to elicit a stronger negative emotional reaction compared to the switch dilemma, thereby increasing the likelihood of deontological judgments.

In postulating that moral judgments can arise from two qualitatively distinct processes, the DPM integrates elements of both intuitionist and rationalist approaches. Adopting elements of intuitionist approaches, the DPM postulates that moral judgments can be rooted in automatic emotional processes. This claim concurs with the SIM in the contention that many moral judgments arise from automatic emotional processes, although the DPM makes the narrower claim that such processes primarily support judgments consistent with moral norms and rules. Adopting elements of rationalist approaches, the DPM postulates that moral judgments can be rooted in controlled cognitive processes. This claim conflicts with the SIM, which holds that such processes primarily serve to justify the outputs of automatic processes, only rarely influencing moral judgment itself. According to the DPM, controlled cognitive processes often override automatic emotional processes to produce judgments that maximize outcomes for the greater good.

Criticism of Domain-Specific Dual-Process Theories

Overall, the reviewed domain-specific dual-process theories have gained strong support in the form of empirically confirmed predictions, and many of them have been seminal in shaping the field of social psychology over the past decades. Nevertheless, these theories have also been the target of criticism. One critical argument is that many of them equate different information contents with distinct mental processes. For example, dual-process theories of persuasion have been criticized for conflating different types of information (e.g., characteristics of the source, message arguments) with different types of processes (e.g., peripheral vs. central; heuristic vs. systematic), even though the processes by which the two kinds of information are integrated into attitudinal judgments may be the same (Kruglanski & Thompson, 1999). Similar concerns have been raised against dual-process theories of impression formation (Chun & Kruglanski, 2006) and dual-process theories of dispositional attribution (Chun, Spiegel, & Kruglanski, 2002), which tend to conflate different kinds of information (e.g., category-related vs. individuating information; behavioral vs. situational information) with qualitatively distinct processes. According to this criticism, the different types of information in these theories may differ in terms of their overall complexity, and therefore require different amounts of processing resources. However, this does not necessarily mean that the different types of information influence behavioral responses via qualitatively distinct processes (see Kruglanski, Klein, Pierro, & Mannetti, 2014).

A related concern is that the existence of content-dependent principles of information processing suggested by domain-specific dual-process theories seems rather implausible. Instead, it seems more likely that the human mind is characterized by a set of basic principles that generalize across content domains (Kruglanski et al., 2014; Smith & DeCoster, 2000). From this perspective, the large number of domain-specific dual-process theories may be detrimental to scientific progress, because it hampers the identification of basic principles that may underlie all of the to-be-explained phenomena. These concerns have inspired the development of domain-independent dual-systems theories, which (1) resolve the conflation between process and content and (2) identify general principles of information processing that generalize across content domains.

Domain-Independent Dual-System Theories

Domain-independent dual-system theories aim to identify basic principles of social information processing that generalize across content domains. A shared feature of these theories is that they propose two mental systems that operate based on qualitatively distinct principles. In the following sections, we first review the core assumptions of the most prominent examples in the field of social psychology, and then discuss criticisms that have been raised against domain-independent dual-system theories.

Cognitive-Experiential Self Theory

Although not designed with this aim, the groundwork for domain-independent dual-system theories was put in place by Epstein's (1994) cognitive-experiential self-theory (CEST), which is based on his foundational work on the nature of the self-concept as a global theory of personality (Epstein, 1973). In broad terms, CEST proposes two interacting systems that are characterized by different processing principles. The first is described as the *experiential system*; the second is described as the *rational system* (for a review, see Epstein & Pacini, 1999).

The experiential system is assumed to operate in an automatic, effortless manner based on associative connections that are closely linked to affective principles of pleasure and pain (*what feels good or bad?*). Encoding of reality in the experiential system is claimed to occur in concrete images, metaphors, and narratives, involving holistic responses that are oriented toward immediate action. As such, responses driven by the experiential system are characterized by broad, schematic generalizations that tend to be incoherent, crudely integrated, and context specific. Changes in the experiential system are assumed to occur slowly, requiring repetitive or relatively intense experiences. These processing principles are assumed to be rooted in brain structures that developed early in evolution and

that have not been replaced by more recently evolved structures that build the foundation for the second, rational system.

The rational system is characterized by intentional, effortful processing that is based on logical relations between elements (*what is rational?*). Encoding of reality in the rational system is claimed to occur in abstract symbols, words, and numbers, involving analytic responses that are oriented toward delayed action. Thus, responses driven by the rational system are characterized by differentiated, highly integrated representations that tend to be abstract, logically coherent, and context independent. Changes in the rational system are assumed to occur more quickly compared with the experiential system, with changes depending on argument strength and availability of new evidence.

According to Epstein (1994), the two systems operate in parallel, such that each system can independently produce its own response tendency. In cases in which these response tendencies are incongruent, people tend to experience a "conflict between the head and the heart" (p. 710), such that the experiential system may produce an intuitive, affective response tendency that conflicts with a rational, logical response tendency produced by the rational system. At the same time, the two systems may interact with each other, such that preconscious processes in the experiential system may continuously influence conscious processing in the rational system. However, the proposed interaction between the two systems is assumed to be asymmetrical, because influences from the experiential system usually remain outside of conscious awareness. As a result, these influences often remain uncontrolled by the rational system, because people tend to be unaware that there is anything to control. Nevertheless, there can be individual and situational variations in the relative dominance of the two systems. For example, Epstein, Pacini, Denes-Raj, and Heier (1996) developed the Rational-Experiential Inventory (REI), which includes two individual difference measures that are specifically designed to identify stable individual differences in the dominance of intuitive-experiential and analytical-rational thinking styles. Other moderating factors include situational circumstances and emotional arousal. Whereas circumstances that require a formal analysis of the current situation are assumed to give priority to the rational system, emotional arousal is assumed to shift the balance toward the experiential system.

Associative vs. Rule-Based Memory Systems

Another milestone in domain-independent dual-system theorizing is Smith and DeCoster's (2000) conceptual integration of various domain-specific dual-process theories. Drawing on Sloman's (1996) distinction between associative and rule-based processing and McClelland and colleagues' work on fast-learning and slow-learning memory systems (McClelland, McNaughton, & O'Reilly, 1995), Smith and DeCoster (2000) argued that many phenomena identified and studied by domain-specific dual-process theories reflect the operation of two distinct memory systems that are guided by different processing principles: a slow-learning system that is characterized by associative processing and a fast-learning system that is characterized by rule-based processing. Associative processing is further specified as being structured by similarity and contiguity, drawing on simple associations between objects and events that are learned slowly over many experiences. Associative processing is assumed to occur automatically in a parallel fashion without awareness of the involved processing steps, even though their output may be accessible to conscious awareness. Rule-based processing, in contrast, is characterized as being structured by language and logic, drawing on symbolically represented rules that can be learned quickly with very few experiences. Attributing a dominant role to associative processing, rule-based processing is further assumed to occur optionally in a sequential fashion if both the motivation and the capacity to engage in effortful processing are present. Its processing steps are often accessible to conscious awareness, such that the applied rules of inference can be verbalized. Similar to Epstein's (1994) CEST, Smith and DeCoster (2000) propose an asymmetrical interaction between the two memory systems, such that rule-based processing may draw on inputs from both memory systems, whereas associative processing is exclusively based on the slow-learning system.

Using the distinction between associative and rule-based processing in the two memory systems, Smith and DeCoster (2000) integrated the phenomena identified and studied by various domain-specific dual-process theories within a single unifying framework. For example, peripheral/heuristic processing in dual-process theories of persuasion (Chaiken, 1987; Petty & Cacioppo, 1986) is characterized as the use of well-learned associations of salient cues (e.g., source attractiveness) with positive or negative evaluations. Central/systematic processing, in contrast, is described as the effortful search for relevant information that is evaluated using rule-based processes based on logical principles. Along the same lines, automatic attitude activation in Fazio's (1990) MODE model is described as automatic access to summary evaluations that are associated with an attitude object through repeated pairings. Deliberate analysis of an object's attributes, in contrast, is characterized as the search for and appraisal of relevant information based on logical rules of inference. Correspondent dispositional inferences in Gilbert's (1989) three-stage model are described as the

use of traits that are semantically associated with a person's observed behavior, whereas inferences about alternative causes (e.g., situational factors) are assumed to involve rule-based processes that are guided by principles of logical inference. Similar considerations apply to Devine's (1989) dissociation model of prejudice and stereotyping, in that automatic stereotype activation is assumed to be the result of highly overlearned associations between social groups and stereotypic information, whereas suppression of these automatically activated stereotypes involves effortful access to personal beliefs in order to override the impact of stereotypic information. Finally, automatic categorization in dual-process theories of impression formation (Brewer, 1988, Fiske & Neuberg, 1990) is described as the use of information that is associated with a person's salient category (e.g., gender, race, age), whereas individuation involves the processing and appraisal of multiple individual characteristics to form a personal impression.

System 1 vs. System 2

Working toward a theoretical integration of earlier research on heuristics and biases (for a review, see Gilovich, Griffin, & Kahneman, 2002), Kahneman (2003) presented a dual-system theory that shares many features with Smith and DeCoster's (2000) and Epstein's (1994) theories. To this end, Kahneman (2003) distinguished between two systems, generically described as *System 1* and *System 2* (see Stanovich & West, 2000), that are assumed to underlie intuition versus reasoning. Sharing characteristics of basic perceptual processes, intuitive processing in System 1 is described as fast, parallel, automatic, effortless, associative, slow learning, and emotional. In contrast, reasoning processes in System 2 are described as slow, serial, controlled, effortful, rule-governed, fast learning, and emotionally neutral. At the same time, information processing in the two systems is assumed to differ from basic perceptual processes, in that both intuition and reasoning can be evoked by verbal information, involving conceptual representations of the past, the present, and the future. These features differ from basic perceptual processes, which involve stimulus-bound percepts that are driven by current stimulation. Thus, whereas the outputs of System 1 may be described as intuitive *impressions*, the outputs of System 2 are *judgments* that can be based on intuitive impressions or deliberate reasoning. In that sense, an important function of System 2 is to monitor the activities and inputs of System 1. If no intuitive response is generated by System 1, judgments and behavior are exclusively computed by System 2. If, however, System 1 provides an intuitive response as input for System 2, System 2 may either (1) endorse this response, (2) adjust the response for other features that are recognized to be relevant, (3) correct the response for a recognized bias, or (4) block the response from overt expression if it is identified to violate a valid rule of inference.

Whereas the intuitive responses generated by System 1 are determined by the accessibility of mental contents (Higgins, 1996), processing in System 2 is guided by the application of logical rules of inference. In the case of heuristic judgments, highly accessible contents in System 1 will often pass the monitoring activities of System 2 through a process of attribute substitution. In general terms, attribute substitution occurs when an "individual assesses a specified target attribute of a judgment object by substituting a related heuristic attribute that comes more readily to mind" (Kahneman, 2003, p. 707). This process can be illustrated with the well-known Linda problem that has been used to demonstrate the conjunction fallacy (Tversky & Kahneman, 1983, p. 297): "*Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations*." According to Tversky and Kahneman (1983), people commit the conjunction fallacy when they judge Linda as more likely to be (1) a bank teller and active in the feminist movement than (2) a bank teller. Because the conjunction of two distinct events can never be more likely than one of the two events by itself, such a judgment violates basic principles of statistical probability. Drawing on Kahneman's (2003) conceptualization, this judgmental tendency can be explained by accessibility-driven attribute substitution, in that individuals substitute a relevant judgmental attribute (i.e., statistical probability) with an irrelevant, yet highly accessible attribute (i.e., feature similarity). Other examples of accessibility-driven attribute substitution include framing effects, in which people tend to show a preference for positively over negatively framed objects, events, and decisions (e.g., a sausage that is described as 80% lean vs. 20% fat), even if the two descriptions are semantically equivalent.

Reflection-Reflexion Model

The reflection–reflexion model proposed by Lieberman (2003; see also Lieberman, Gaunt, Gilbert, & Trope, 2002) combines the basic notion of dual-system theorizing with insights in social-cognitive neuroscience. Deviating from conceptualizations that describe automatic processes as more efficient variants of insufficiently practiced controlled processes (e.g., Bargh, 1997), the reflection–reflexion model argues that automatic and controlled processes operate on the basis of qualitatively distinct representations with distinct neural underpinnings. According to Lieberman (2003), the automatic–controlled distinction is insufficient, if not misleading, because it simply describes the operating conditions of a given process

(i.e., when does the process operate?) without specifying the underlying computational properties (i.e., what is the process doing?). Because the reflection–reflexion model assumes distinct representational underpinnings, it also allows for complex interactions between automatic and controlled processes. Such interactions are difficult to reconcile with the view that automatic and controlled processes draw on the same underlying mental representations, the only difference being that automatic processes operate more efficiently as a result of practice.

The first system, called the *X-system* with reference to the term *reflexive*, is proposed to involve the amygdala, basal ganglia, and lateral temporal cortex. Reflexive processes in the X-system link affect and meaning to currently represented stimuli by means of simple stimulus-stimulus associations (semantic meaning) or stimulus-outcome associations (affective meaning). These associations build the foundation for a person's implicit theories and generalized expectations about the world, which are assumed to develop slowly over time and over extended periods of learning. The neurons in the X-system are highly interdependent, in that they are mutually influenced by the neurons they are influencing. As a result, activation of associations in the X-system operates in a parallel fashion on the basis of similarity and pattern matching, such that observed relations of the format "if *p*, then *q*" will create a reflexive tendency to draw the logically incorrect inference "if *q*, then *p*."

The second system, called the *C-system* with reference to the term *reflective*, is proposed to involve the anterior cingulate cortex, prefrontal cortex, and medial temporal lobe. The operation of the C-system is assumed to be conditional on the failure of the X-system to achieve a momentary goal, such that reflective processes are initiated only if (1) the implicit theories or generalized expectancies in the X-system are violated, or (2) there are no implicit theories or generalized expectancies in the X-system that are applicable to guide behavior in a novel situation. As such, the primary function of the C-system is to handle context-specific "exceptions to the rule" for which the generalizations in the X-system are not prepared. Reflective inferences in the C-system are assumed to operate in a sequential manner on the basis of causal and logical relations, which allows the C-system to block logically incorrect inferences of *p* in the presence of *q* from observed relations of the format "if *p*, then *q*."

Reflective-Impulsive Model

Another highly influential dual-system theory is Strack and Deutsch's (2004) reflective-impulsive model (RIM). The RIM argues that human behavior is guided by two simultaneously operating systems of information processing, which are described as the *reflective system* (RS) and the *impulsive system* (IS). Even though the two systems are assumed to operate in parallel, the IS enjoys priority over the RS because the RS operates only under conditions of sufficient cognitive capacity, whereas information processing in the IS is assumed to be resource-independent. Similar to other dual-system theories, the RIM states that the IS operates based on simple associative links between elements that are formed and activated according to the principles of similarity and contiguity. Information processing in the RS, in contrast, is assumed to involve propositionally represented relations between elements, which are tagged with truth values (i.e., true vs. false). These characteristics make the RS capable of various operations that cannot be performed by the IS, the most important being the processing of negations and representations of the future. Thus, even though accessible associations in the IS provide the basis for propositional representations in the RS, their functionally distinct operating principles can have different behavioral implications if processing in the RS involves the negation of activated associations in the IS (see Deutsch, Gawronski, & Strack, 2006) or conflicts between immediate desires and long-term goals (see Metcalfe & Mischel, 1999).

Another central assumption of the RIM concerns the translation of mental representations into behavior. The RIM assumes that the RS and the IS influence behavior through a common pathway that includes the activation of behavioral schemata of varying abstractness (Norman & Shallice, 1986). These behavioral schemata can be activated directly through the spread of activation from momentarily accessible associations in the IS, which may elicit an impulsive tendency to either approach or avoid a given object. Alternatively, behavioral schemata may be activated indirectly through behavioral intentions generated in the RS, which are guided by (1) the subjective hedonic quality of future states that may result from a given behavior (i.e., value) and (2) the subjective probability with which the behavior may produce the focal outcome (i.e., expectancy).

Going beyond dual-process theories that focus primarily on cognitive and affective processes, the RIM attributes an important role to motivational processes, which may operate in the IS in two ways. First, the RIM integrates basic principles of homoeostatic dysregulation, such that deprivation of basic needs (e.g., hunger) is assumed to activate behavioral schemata that are linked to successful satiation of these needs through a history of past experiences. Second, the RIM assumes that motivational orientations to approach or avoid an object may be elicited by mere processing of positive or negative information, mere perception of approach or avoidance movements, the experience of positive or negative affect, or the execution of approach or avoidance motor actions.

Conversely, motivational orientations of approach or avoidance are assumed to facilitate the processing of information, the experience of affective states, and the execution of behavior that are compatible with the current motivational state.

Criticism of Domain-Independent Dual-System Theories

Although the reviewed dual-system theories may seem very similar, each of them deserves credit for unique contributions to the field of social psychology. Epstein's (1994) CEST has made a significant contribution by laying the groundwork for domain-independent dual-system theorizing; Smith and DeCoster's (2000) memory-systems model has made a significant contribution by providing a conceptual integration of various domain-specific dual-process theories; Kahneman's (2003) distinction between System-1 and System-2 processing has been highly influential in shaping research on judgment and decision-making; Lieberman's (2003) reflection-reflexion model deserves credit for integrating dual-system theorizing with insights in social-cognitive neuroscience; and Strack and Deutsch's (2004) RIM has been highly influential in various applied areas dealing with self-regulatory conflicts (for an overview, see Deutsch, Gawronski, & Hofmann, 2017). Nevertheless, domain-independent dual-system theories have also been the target of abundant criticism.

One concern is that the mapping of multiple dualities hypothesized by dual-system theories is conceptually and empirically implausible (Keren & Schul, 2009; Melnikoff & Bargh, 2018; Moors, 2014). For example, describing one system as affective, automatic, holistic, and associative and the other one as cognitive, controlled, analytic, and logical (e.g., Epstein & Pacini, 1999) implies that cognitive processes cannot be automatic, holistic, or associative—an assumption that seems questionable on both conceptual and empirical grounds. Indeed, the mapping of multiple dualities can be criticized for ignoring an important lesson from early research on automaticity, indicating that there is virtually no process that is characterized by all four features of automaticity (i.e., unintentional, efficient, uncontrollable, unconscious; see Bargh, 1994; Fiedler & Hütter, 2014; Melnikoff & Bargh, 2018). Dual-system theories not only seem to ignore this insight; they exacerbate the problem by adding more dualities to the lists of characteristics that supposedly describe the two processing systems.

Another concern is that domain-independent dual-system theories seem unfalsifiable (Keren & Schul, 2009). However, a fair evaluation of this criticism requires a more nuanced analysis of what makes a theory unfalsifiable. According to Popper (1934), the degree to which a theory is falsifiable increases with the number of possible events that should *not* happen according to the theory (see Gawronski & Bodenhausen, 2015b). To the extent that any of these prohibited events occur, the theory would be in conflict with an empirical observation, which implies that it is falsifiable. From this perspective, a theory would be unfalsifiable if it does not prohibit any events. Is it true that domain-independent dual-system theories are unfalsifiable in the sense that they do not prohibit any events? The answer to this question is *no*. In fact, this answer is rooted in the same mapping of dualities that has been criticized as conceptually and empirically implausible. Imagine a dual-system theory that uses eight dualities to characterize the two hypothesized systems (e.g., associative, automatic, slow-learning, holistic vs. rule-based, controlled, fast-learning, analytical). Technically, a set of eight dualities implies a total of $2 \times 2 \times 2 \times 2 = 16$ potential combinations of system properties. Yet, by mapping one of the respective duality features onto one system and the respective others onto the other system, the theory suggests that only two of the 16 combinations exist, thereby prohibiting the other 14 (see Keren & Schul, 2009; Melnikoff & Bargh, 2018). From this perspective, dual-system theories are actually highly falsifiable, because they permit only a very small subset of potential feature combinations. Thus, although the proposed mapping of multiple dualities can be criticized for rendering dual-system theories conceptually and empirically implausible, it simultaneously refutes the criticism that dual-system theories are unfalsifiable in terms of Popper's (1934) conceptualization.[2]

However, there is another way in which a theory can be unfalsifiable, which has been described as *irrefutable* to distinguish it from Popper's (1934) original conceptualization (see Gawronski & Bodenhausen, 2015b; Quine & Ullian, 1978). According to this alternative conceptualization, a theory is irrefutable if it can be reconciled with any empirical outcome without requiring significant adjustments to the currently held set of theoretical core assumptions (see also Platt, 1964). The most common reason for this to occur is when there is no straightforward way to link the constructs proposed by a given theory to observable empirical events. In line with this concern, dual-system theories have been

[2] In response to the criticism that the proposed mappings of multiple dualities are conceptually and empirically implausible, some dual-system theorists argued that the hypothesized mappings are meant to describe mere correlations between features rather than perfect overlap (e.g., Evans & Stanovich, 2013). Although such an interpretation addresses the implausibility concern, it reduces the falsifiability of dual-system theories in terms of Popper's (1934) criterion (see Keren, 2013).

criticized for not providing sound criteria for the identification of whether the process underlying observed relations between environmental inputs and behavioral outputs is associative or rule-based (Kruglanski & Gigerenzer, 2011; Moors, 2014). For example, some dual-system theorists suggest that a process qualifies as associative if people are unaware of the involved processing steps, and as rule-based if people are aware of the involved processing steps (e.g., Lieberman, 2009; Smith & DeCoster, 2000). Other theorists suggest that a process qualifies as associative if it does not require a large amount of processing resources, and as rule-based if it does require a large amount of processing resources (e.g., Dhar & Gorlin, 2013; Evans & Stanovich, 2013). However, such conceptualizations can be criticized for using operating conditions (i.e., conscious vs. unconscious; efficient vs. inefficient) to identify operating principles (i.e., associative vs. rule-based) without linking the latter to specific input-output relations (Gawronski et al., 2014; Moors, 2014). Thus, if a given input-output relation is explained in terms of associative processes and the emergence of this input-output relation turns out to depend on conscious awareness or cognitive resources, the input-output relation would simply be recategorized as the product of rule-based processes (or vice versa) without requiring any significant adjustment to the currently held set of theoretical core assumptions. Hence, although dual-system theories are falsifiable in terms of Popper's (1934) original criterion, they can be criticized for being irrefutable if they do not include clear specifications of the input-output relations that can be expected to result from associative versus rule-based processes (see Heycke & Gawronski, 2020, for a discussion of how associative and propositional learning can be identified independent of automaticity features).

A final concern is that extant dual-system theories seem redundant in the sense that it seems impossible to derive conflicting predictions from these theories (Keren & Schul, 2009). To the extent that this concern is valid, it would be justified to question the value of having multiple distinct dual-system theories. Although domain-specific dual-process theories may be criticized for their narrow focus and for conflating process and content, they are less susceptible to this criticism because their predictions are unique and complementary rather than empirically redundant.

Formalized Dual-Process Theories

Simultaneous to the emergence of domain-independent dual-system theories, social cognition researchers became concerned that many behavioral phenomena may not be process-pure, but instead reflect the joint contributions of automatic and controlled processes. This insight posed a challenge for a dominant approach in dual-process research, where one process is often equated with responses on one type of task (e.g., implicit measures) and the other process with responses on another task (e.g., explicit measures). To address these concerns, researchers have adopted mathematical modeling procedures from cognitive psychology to quantify the joint contributions of distinct processes to behavioral responses within a single task. In context of dual-process theories, the most prominent example is Jacoby's (1991) process dissociation (PD) model, which quantifies the independent contributions of two distinct processes to behavioral outcomes.

Process Dissociation Model

The basic idea of PD is that automatic and controlled processes sometimes work in concert to produce a behavioral response, while at other times automatic and controlled processes work in opposition to each other (for a review, see Payne & Cameron, 2014). For example, many implicit measures—such as the implicit association test (Greenwald et al., 1998) or evaluative priming (Fazio et al., 1995)—involve one class of trials that is described as *compatible* and another class of trials that is described as *incompatible*. The basic idea is that both automatic and controlled processes will lead to the correct response on compatible trials, thereby facilitating fast and accurate responses. On incompatible trials, however, only controlled processes will lead to the correct response, whereas automatic processes will lead to the incorrect response, thereby inhibiting fast and accurate responses. For example, in an implicit association test designed to measure automatic racial bias (Greenwald et al., 1998), controlled identification of a black face will produce an accurate response regardless of whether black faces are mapped onto the same key as negative words (i.e., prejudice-compatible block) or positive words (i.e., prejudice-incompatible block). In contrast, automatic racial bias will produce an accurate response when black faces are mapped onto the same key as negative words (i.e., prejudice-compatible block), but an incorrect response when black faces are mapped onto the same key as positive words (i.e., prejudice-incompatible block).

These influences can be depicted graphically in a processing tree that describes how automatic and controlled processes may jointly determine correct versus incorrect responses on a given task (see Figure 1). If the controlled process succeeds (depicted as C in Figure 1), participants will show the correct response on both compatible and incompatible trials. If, however, the controlled process fails (depicted as $1 - C$ in Figure 1), the behavioral outcome depends on the operation of the automatic process. If the automatic process succeeds (depicted as A in Figure 1), participants will show the correct response on

compatible trials but the incorrect response on incompatible trials. Yet, if the automatic process fails (depicted as *1 – A* in Figure 1), the model assumes a bias in the opposite direction, such that participants will show the incorrect response on compatible trials but the correct response on incompatible trials.

A major advantage of the PD model over traditional dual-process theories is that it provides mathematical equations that can be used to quantify the relative contributions of distinct processes to performance on tasks in which automatic and controlled processes can work in concert or in opposition. The basic idea is to derive two equations from the proposed structure in the processing tree, one for the probability of correct (vs. incorrect) responses on compatible trials and one for the probability of correct (vs. incorrect) responses on incompatible trials. In Jacoby's (1991) PD model, these equations include two parameters as unknowns: *C,* which is supposed to capture the impact of the controlled process; and *A,* which is supposed to capture the impact of the automatic process. Using the empirically observed probabilities of correct responses on compatible and incompatible trials in a given data set, the particular values of these unknowns can be calculated through simple algebra.

For example, using the graphical depiction of Jacoby's (1991) model in Figure 1, the probability of a correct response on compatible trials should be equal to all processing paths from left to right that lead to a correct response in the "compatible" column. The two paths that produce such a response are *Controlled Process Succeeds*, which can be depicted as *C*, and *Controlled Process Fails* in conjunction with *Automatic Process Succeeds*, which can be depicted as *(1 – C) × A*. Thus, in statistical terms, the probability of a correct response on compatible trials can be described as:

$$p(\text{correct} \mid \text{compatible}) = C + (1 - C) \times A$$

The same logic can be applied to the probability of a correct response on incompatible trials.[3] The two paths that produce such a response are *Controlled Process Succeeds*, which is again depicted as *C*, and *Controlled Process Fails* in conjunction with *Automatic Process Fails*, which is depicted as *(1 – C) × (1 – A)*. Based on these processing paths, the probability of a correct response on incompatible trials can be described as:

$$p(\text{correct} \mid \text{incompatible}) = C + (1 - C) \times (1 - A)$$

Through the use of linear algebra, these equations can be solved for *C* and *A*, which allows researchers to quantify the relative impact of automatic and controlled processes. Without going into the details of the mathematical conversion, the controlled process can be quantified algebraically as:

$$C = p(\text{correct} \mid \text{compatible}) - p(\text{incorrect} \mid \text{incompatible})$$

Using the value computed for *C*, the automatic process can then be calculated as:

$$A = p(\text{incorrect} \mid \text{incompatible}) / (1 - C)$$

Estimates for C and A can be calculated for each participant in a given sample, allowing the use of these estimates as dependent variables in experimental designs or as individual-difference variables in correlational designs.

Applications of Process Dissociation

Because the PD model is content-agnostic, it is possible to apply it to a wide range of phenomena where two processes work in concert or opposition (for reviews, see Hütter & Klauer, 2016; Payne & Cameron, 2014). A prominent example we have already mentioned is research that has used PD to quantify the contributions of automatic and controlled processes to responses on implicit measures. A closely related example is the use of PD to disentangle the contributions of automatic and controlled processes to racial bias in weapon identification (for a review, see Payne & Correll, 2020). Payne (2001) developed a sequential priming task in which participants are briefly presented with either a black or a white face prime, followed by a target picture showing either a gun or a harmless object. The target picture is quickly replaced by a black-and-white pattern mask, and participants' task is to indicate whether the target picture showed a gun or a harmless object. The common result is that harmless objects are more frequently misidentified as guns when the face prime was black than when it was white, whereas guns are more frequently misidentified as harmless objects when the face prime was white than when it was black (for a review, see Payne & Correll, 2020). Using the PD model, Payne (2001) calculated separate estimates reflecting participants' ability to correctly identify guns and harmless objects (reflected in the model's *C* parameter) and racial bias in guessing the nature of the target stimulus when participants were unable to identify the stimulus (reflected in the model's *A* parameter).

Other illustrative examples include the use of PD to disentangle the role of distinct processes in judgment and decision-making. For example, expanding on Tversky and Kahneman's (1974) seminal work on heuristics and biases, Ferreira, Garcia-Marques, Sherman, and Sherman (2006) used PD to quantify the contributions of heuristic and rule-based reasoning to judgmental biases arising from the representativeness heuristic, including the conjunction fallacy (Tversky &

[3] Note that $p(\text{correct} \mid \text{compatible}) = 1 - p(\text{incorrect} \mid \text{compatible})$. Correspondingly, $p(\text{correct} \mid \text{incompatible}) = 1 - p(\text{incorrect} \mid \text{incompatible})$.

Kahneman, 1983) and base-rate neglect (see Kahneman & Tversky, 1972). Expanding on research on the role of conscious and unconscious thought in decision-making (for a review, see Dijksterhuis, Strick, Bos, & Nordgren, 2014), Damian and Sherman (2013) used PD to disentangle the contributions of intuitive and rule-based processing to the two modes of thinking. Finally, expanding on Greene's (2008, 2014) dual-process model of moral judgment, Conway and Gawronski (2013) used PD to disentangle deontological and utilitarian inclinations in moral dilemma judgments.

Criticism of Process Dissociation

Compared with domain-specific dual-process theories and domain-independent dual-system theories, formalized dual-process theories such as the PD model have the unique advantage that they provide a mathematical tool to quantify the contributions of distinct processes to behavioral responses (for a more extensive discussion of advantages, see Klauer, 2015). At the same time, there have been controversies surrounding the conceptual meaning of the two parameters of the PD model (see Gawronski & Creighton, 2013). A common assumption in research using PD is that the two model parameters provide direct access to automatic and controlled processes. This assumption is rooted in depictions of the two parameters as *automatic* (using the acronym *A*) and *controlled* (using the acronym *C*). Such depictions are problematic for at least two reasons. First, as outlined in the initial sections of this chapter, different features of automaticity do not necessarily covary (Bargh, 1994; Fiedler & Hütter, 2014; Moors & De Houwer, 2006; Melnikoff & Bargh, 2018), which makes generic descriptions of the PD parameters as automatic or controlled misleading as long as it is not specified in which particular sense the captured process is assumed to be automatic or controlled (i.e., unintentional, efficient, uncontrollable, unconscious). Second, depictions of the two parameters as automatic and controlled conflate operating principles (i.e., what is the process doing?) with operating conditions (i.e., when does the process operate?). Although PD provides a valuable tool to quantify the impact of two qualitatively distinct processes, the conditions under which these processes operate have to be confirmed empirically by means of conceptually appropriate manipulations. There is nothing in the PD equations for the two parameters that would guarantee that the process captured by *A* is automatic and the process captured by *C* is controlled (for a more detailed discussion, see Gawronski & Creighton, 2013).

Another concern is that the two parameters of the PD model are often insufficient to capture the processes underlying behavioral responses. This concern is reflected in the development of formalized multi-process alternatives to the PD model that include more than two parameters. For example, to capture the processes underlying responses on implicit measures, Conrey, Sherman, Gawronski, Hugenberg, and Groom (2005) developed a multi-process model that quantifies the contributions of four distinct processes (for a review, see Sherman et al., 2014): the likelihood that an automatic association is activated (described as *Association Activation* or *AC*); the likelihood that the correct response to the stimulus can be determined (described as *Discriminability* or *D*); the likelihood that an automatic association is successfully overcome in favor of the correct response (described as *Overcoming Bias* or *OB*); and the likelihood that a general response bias (e.g., right-hand bias) drives the response (described as *Guessing* or *G*). To study the processes underlying responses to moral dilemmas (see Greene, 2008, 2014), Gawronski, Armstrong, Conway, Friesdorf, and Hütter (2017) developed a multi-process model that quantifies the contribution of three distinct factors: (1) sensitivity to consequences, (2) sensitivity to moral norms, and (3) general preference for inaction versus action. Finally, to capture the processes underlying moral judgments of right and wrong (see Haidt, 2001), Cameron, Payne, Sinnott-Armstrong, Scheffer, and Inzlicht (2017) developed a multi-process model that quantifies the contribution of three distinct processes: (1) intentional moral judgment, (2) unintentional moral judgment, and (3) general response bias in moral judgment. Although these formalized models are domain-specific in the sense that they have been designed for particular content areas, they share the concern that the two parameters of the PD model are insufficient to capture the complexity of processes underlying responses in the respective content areas.[4]

Outlook

Although dual-process theories have inspired an enormous amount of research (for overviews, see Chaiken & Trope, 1999; Sherman et al., 2014), they have also become the target of increased criticism. In the preceding sections, we have already reviewed critical arguments against specific kinds of dual-process theories. In the final sections of this chapter, we discuss some ongoing debates that emerged in response to broader issues in the literature, offering

[4] A more fundamental critique is that the parameters of all formal models capture patterns of stimulus-response relations at the behavioral level rather than explanatory mental constructs (De Houwer & Moors, 2015; Gawronski & Bodenhausen, 2015a). From this perspective, the number of parameters in a formal model has no implications for the number of underlying mental processes (for discussions, see Gawronski, Conway, Hütter, Luke, Armstrong, & Friesdorf, 2020; Heycke & Gawronski, 2020).

considerations regarding the current state of theorizing and directions for future developments.

Formation, Representation, and Behavioral Expression

An emerging theme in current debates about dual-process theories is the need for more nuanced distinctions between (1) the processes involved in the formation of mental representations, (2) the structure of the resulting representations, and (3) the processes by which these representations influence judgments and behavior (Corneille & Stahl, 2019; De Houwer et al.; 2020; Gawronski, Brannon, & Bodenhausen, 2017; Kurdi & Dunham, 2020; Mandelbaum, 2016; see also Ferguson et al., 2014).[5] Although it can be difficult to disambiguate the nature of a given effect in terms of the three components (see De Houwer & Moors, 2015), conclusions pertaining to one component may not necessarily generalize to the other two.

Regarding the presumed role of two distinct processes in the formation of mental representations, a recent review by Corneille and Stahl (2019) raises significant questions about the existence of a learning process involving the automatic formation of associative links in memory (see also Mitchell, De Houwer, & Lovibond, 2009). Based on their review, the authors concluded that the available evidence is better explained by single-process propositional theories assuming that all learning effects are driven by the controlled generation and validation of mental propositions about specific stimulus relations (e.g., De Houwer et al., 2020). These concerns have inspired novel lines of research using formal modeling approaches to provide more nuanced insights into the presumed role of associative and propositional processes in the formation of mental representations (Heycke & Gawronski, 2020; Kukken, Hütter, & Holland, 2020). Ironically, some of the evidence obtained in this research poses an explanatory challenge for both dual-process and single-process theories (e.g., Heycke & Gawronski, 2020), raising important new questions about the processes underlying the formation of mental representations.

Expanding on Corneille and Stahl's (2019) skeptical conclusion about the existence of two functionally distinct learning mechanisms, further questions arise regarding the presumed structure of mental representations. Although some dual-process theories propose two qualitatively distinct representations in memory (e.g., Lieberman, 2003; Smith & DeCoster, 2000), most dual-process theories hypothesize a single associative store as the basis for the operations of two qualitatively distinct processes (e.g., Fazio, 1990; Gawronski & Bodenhausen, 2006; Strack & Deutsch, 2004). Yet, if there is no evidence for an associative learning process that shapes mental representations over and above propositional learning (Corneille & Stahl, 2019), questions could be raised about the representational assumptions of either type of theory (see De Houwer et al.; 2020; Mandelbaum, 2016). On the one hand, dual-process theories assuming two distinct memory systems (e.g., associative vs. rule-based) could be criticized for proposing a distinct associative system that does not seem necessary to explain the available evidence. On the other hand, dual-process theories assuming a single associative store face the question of whether they are able to capture the complexity of mental representations arising from propositional learning. Whereas some dual-process theorists argued that associative networks are capable of capturing these complexities (e.g., Gawronski, Brannon, & Bodenhausen, 2017), critics remain rather skeptical about this possibility (e.g., De Houwer et al., 2020).

Regarding the behavioral expression of mental representations, a debated issue is whether it is necessary to postulate two qualitatively distinct processes to account for the available evidence. Questioning a core assumption of many dual-process theories, proponents of single-process theories have argued that the evidence accumulated by dual-process researchers can be fully explained by a single process of rule-based inference (Kruglanski & Gigerenzer, 2011) or propositional reasoning (De Houwer et al., 2020) operating during the behavioral expression of mental representations. Yet, in evaluating this criticism, it is worth noting that the proposed single-process explanations include assumptions that, upon closer scrutiny, show a high resemblance to the explanations provided by dual-process theories. For example, in their critique of the dual-process distinction between associative and propositional processes, Kruglanski and Gigerenzer (2011) postulated that all judgments are the product of a single epistemic process of applying inferential rules to judgment-relevant information, which is assumed to be shaped by (1) the accessibility of stored information and (2) the perceived relevance of accessible information. Somewhat ironically, accessibility and relevance are claimed to influence judgments via a "two-step process" that strongly resembles the hypothesized functions of (1) associative activation in determining the accessibility of stored information and (2) propositional validation in

[5] Some domain-specific dual-process theories are explicit about whether they aim to identify processes underlying the formation of mental representations (e.g., Chaiken, 1987; Petty & Cacioppo, 1986) or processes underlying the behavioral expression of mental representations (e.g., Fazio, 1990), while others explicitly distinguish between the two stages of information processing (e.g., Gawronski & Bodenhausen, 2006, 2014). However, the distinction remains unclear in various other theories and it has become increasingly blurry in domain-independent dual-system theories.

determining the perceived relevance of activated information (see Gawronski, Brannon, & Bodenhausen, 2017). Thus, although single-process hypotheses regarding the behavioral expression of mental representations may not include any reference to qualitatively distinct processes, their assumptions are very similar to the assumptions of dual-process theories, the only difference being the use of different terms (e.g., dual-process vs. two-step process).

Different from the criticism presented by proponents of single-process theories, other researchers suggested that more than two processes are needed to explain the complexity of human behavior. We have already reviewed some of these arguments in the context of formalized dual-process theories, which have been criticized for not capturing the full set of processes underlying responses on implicit measures (e.g., Conrey et al., 2005), responses to moral dilemmas (e.g., Gawronski, Armstrong, et al., 2017), and moral judgments of right and wrong (e.g., Cameron et al., 2017). A similar concern has been raised by Amodio (2019), who proposed an interactive memory systems model of social cognition that includes five distinct memory systems with unique learning mechanisms. The five memory systems are further assumed to differ in their involvement in the production of different kinds of behavior, depending on the specific affordances of the relevant behaviors (e.g., prospective planning, impression judgments, action decisions). Drawing on the distinction between formation, representation, and behavioral expression, Amodio's (2019) interactive memory systems model proposes (1) five distinct learning mechanisms, (2) five distinct memory systems, and (3) a total of 12 pathways by which the proposed memory systems are involved in the production of judgments and behavior. Needless to say, this proliferation of processes and representations stands in stark contrast to the argument that the findings generated by dual-process research can be fully explained by single-process theories (e.g., Arkes, 2016; Corneille & Stahl, 2019; De Houwer et al., 2020; Kruglanski & Gigerenzer, 2011; Osman, 2013; for a discussion, see Van Dessel, Gawronski, & De Houwer, 2019).

How Many Processes Are There?

Given these conflicting views, a major question is: How many processes are there? In addressing this question, it is important to consider that existence claims—such as claims about the existence of one, two, or multiple processes—are ontological statements. In the philosophy of science, ontological statements fall into the realm of metaphysics, which means that they cannot be tested empirically (Popper, 1934). In other words, we cannot test empirically if there are one, two, or multiple processes. Yet, researchers can make decisions about the usefulness of ontologies by empirically testing assumptions *about* the proposed entities. To the extent that the predictions derived from a given theory are confirmed, the entities proposed by the theory are typically assumed to exist. However, if the predictions derived from a given theory are continuously disconfirmed, it seems likely that researchers will at some point reject the theory, and, by extension, its ontological claims. Note, however, that in these cases it is not the existence claims themselves that are confirmed or disconfirmed, but the assumptions that are made about the proposed entities.

Another important issue in this context is that theories can often be reconciled with prediction-incongruent findings by means of post-hoc assumptions (Gawronski & Bodenhausen, 2015b; Lakatos, 1970). Indeed, some researchers have argued that the flexibility of post-hoc assumptions makes it impossible to empirically disconfirm any of the involved theories, regardless of whether they propose one, two, or multiple processes (e.g., De Houwer et al., 2020). Although this may be true in principle (see Quine, 1953), there are some important issues to consider. First, as we noted earlier in this chapter, some theories can be reconciled with any empirical outcome without requiring significant adjustments to their theoretical core, and such theories can be criticized for being irrefutable (Gawronski & Bodenhausen, 2015b; Quine & Ullian, 1978). Strong theories are refutable in the sense that a reconciliation with prediction-incongruent outcomes would require non-trivial modifications of their theoretical core assumptions instead of minor changes in the theoretical "periphery" (Platt, 1964). Second, post-hoc assumptions to reconcile a theory with prediction-incongruent findings can be evaluated in terms of whether they generate novel empirical predictions. To the extent that this is the case and the newly derived predictions are empirical confirmed, the theoretical trajectory can be described as a *progressive problem shift* (Lakatos, 1970) and there would be no reason to question the validity of the theory. However, if post-hoc assumptions to reconcile a given theory with prediction-incongruent findings do not lead to novel predictions or their novel predictions are empirically disconfirmed, the theoretical trajectory can be described as a *degenerative problem shift* (Lakatos, 1970) that raises significant questions about the value of the theory. Thus, if a theory has a longer history of degenerative problem shifts, researchers may dismiss the theory and, by extension, its ontological claims. From this perspective, future developments in the debate about dual-process theories may be evaluated in terms of (1) the strength of theorizing in terms of the refutability criterion and (2) the ability to respond to prediction-incongruent findings with progressive rather than degenerative problem shifts. These evaluation

criteria apply equally to single-process, dual-process, and multi-process theories.

Conclusion

Although dual-process theories have become the target of increased criticism, their fundamental impact on research in social psychology is undeniable. Yet, faced with the challenges from single-process and multi-process alternatives, the future of dual-process theorizing will likely depend on whether (1) the conceptual criticism put forward by these alternatives can be effectively addressed and (2) empirical challenges can be integrated in a way that leads to novel predictions that can be empirically confirmed. Although proponents of dual-process theories can point to an impressive history of groundbreaking insights, dual-process theorizing can only become stronger if these issues are taken seriously.

References

Amodio, D. M. (2019). Social Cognition 2.0: An interactive memory systems account. *Trends in Cognitive Sciences*, 23, 21-33.

Arkes, H. R. (2016). A levels of processing interpretation of dual-system theories of judgment and decision making. *Theory & Psychology, 26,* 459-475.

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (pp. 1-40). Hillsdale, NJ: Erlbaum.

Bargh, J. A. (1997). The automaticity of everyday life. In R. S. Wyer, Jr. (Ed.), *Advances in social cognition* (Vol. 10, pp. 1-61). Mahwah, NJ: Erlbaum.

Brewer, M. B. (1988). A dual-process model of impression formation. In T. K. Srull & R. S. Wyer (Eds.), *Advances in social cognition* (Vol. 1, pp. 1-36). Hillsdale, NJ: Erlbaum.

Brewer, M. B. & Harasty-Feinstein, A. S. (1999). Dual-processes in the representation of persons and social categories. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 255-270). New York: Guilford Press.

Cameron, C. D., Payne, B. K., Sinnott-Armstrong, W., Scheffer, J. A., & Inzlicht, M. (2017). Implicit moral evaluations: A multinomial modeling approach. *Cognition, 158,* 224-241.

Chaiken, S. (1987). The heuristic model of persuasion. In M. P. Zanna, J. M. Olson, & C. P. Herman (Eds.), *Social influence: The Ontario Symposium* (Vol. 5, pp. 3-39). Hillsdale, NJ: Erlbaum.

Chaiken, S., & Ledgerwood, A. (2012). A theory of heuristic and systematic information processing. In P. A. M. van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (pp. 246-266). Thousand Oaks, CA: Sage.

Chaiken, S., & Maheswaran, D. (1994). Heuristic processing can bias systematic processing: effects of source credibility, argument ambiguity, and task importance on attitude judgment. *Journal of Personality and Social Psychology, 66,* 460-473.

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

Chen, S., & Chaiken, S. (1999). The Heuristic-Systematic Model in its broader context. In S. Chaiken, & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 73-96). New York: Guilford Press.

Chun, W. Y., & Kruglanski, A. W. (2006). The role of task demands and processing resources in the use of base-rate and individuating information. *Journal of Personality and Social Psychology, 91,* 205-217.

Chun, W. Y., Spiegel, S., & Kruglanski, A. W. (2002). Assimilative behavior identification can also be resource dependent: A unimodel perspective on personal-attribution phases. *Journal of Personality and Social Psychology, 83,* 542-555.

Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. (2005). Separating multiple processes in implicit social cognition: The Quad-Model of implicit task performance. *Journal of Personality and Social Psychology, 89,* 469-487.

Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision-making: A process dissociation approach. *Journal of Personality and Social Psychology, 104,* 216-235.

Conway, P., Goldstein-Greenwood, J., Polacek, D., & Greene, J. D. (2018). Sacrificial utilitarian judgments do reflect concern for the greater good: Clarification via process dissociation and the judgments of philosophers. *Cognition*, *179*, 241-265.

Corneille, O., & Stahl, C. (2019). Associative attitude learning: A closer look at evidence and how it relates to attitude model. *Personality and Social Psychology Review, 23,* 161-189.

Damian, R. I., & Sherman, J. W. (2013). A process-dissociation examination of the cognitive processes underlying unconscious thought. *Journal of Experimental Social Psychology*, *49*, 228-237.

De Houwer, J. (2011). Why the cognitive approach in psychology would profit from a functional approach and vice versa. *Perspectives on Psychological Science, 6,* 202-209.

De Houwer, J. (2019). Moving beyond System 1 and System 2. Experimental Psychology, 66, 257-265.

De Houwer, J., & Moors, A. (2015). Levels of analysis in social psychology. In B. Gawronski & G. Bodenhausen (Eds.), *Theory and Explanation in Social Psychology* (pp. 24-40). New York: Guilford Press.

De Houwer, J., Van Dessel, P., & Moran, T. (2020). Attitudes beyond associations: On the role of propositional representations in stimulus evaluation. *Advances in Experimental Social Psychology, 61,* 127-183.

De Neys, W. (in press). On dual- and single-process models of thinking. *Perspectives on Psychological Science*.

De Neys, W., & Pennycook, G. (2019). Logic fast and slow: Advances in dual-process theorizing. *Current Directions in Psychological Science, 28,* 503-509.

Deutsch, R. (2015). Duality models in social psychology. In B. Gawronski, & G. V. Bodenhausen (Eds.), *Theory and explanation in social psychology* (pp. 132-156). New York: Guilford Press.

Deutsch, R., Gawronski, B., & Hofmann, W. (Eds). (2017). *Reflective and impulsive determinants of human behavior*. New York: Psychology Press.

Deutsch, R., Gawronski, B., & Strack, F. (2006). At the boundaries of automaticity: Negation as reflective operation. *Journal of Personality and Social Psychology, 91,* 385-405.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56,* 5-18.

Dhar, R., & Gorlin, M. (2013). A dual-system framework to understand preference construction processes in choice. *Journal of Consumer Psychology, 23,* 528-542.

Dijksterhuis, A., Strick, M., Bos, M., & Nordgren, L. (2014). Prolonged thought: Proposing Type 3 processing. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 355-368). New York: Guilford Press.

Epstein, S. (1973). The self-concept revisited. Or a theory of a theory. *American Psychologist, 28,* 404-416.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist, 49,* 709-724.

Epstein, S., & Pacini, R. (1999). Some basic issues regarding dual-process theories from the perspective of Cognitive-Experiential Self-Theory. In S. Chaiken, & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 462-482). New York: Guilford Press.

Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive-experiential and analytical-rational thinking styles. *Journal of Personality and Social Psychology, 71,* 390-405.

Evans, J. St. B. T. (2010). *Thinking twice: Two minds in one brain*. Oxford: Oxford University Press.

Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher-order cognition: Advancing the debate. *Perspectives of Psychological Science, 8,* 223-241.

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. *Advances in Experimental Social Psychology, 23,* 75-109.

Fazio, R. H. (1995). Attitudes as object-evaluation associations: Determinants, consequences, and correlates of attitude accessibility. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength* (pp. 247-282). Mahwah, NJ: Erlbaum.

Fazio, R. H. (2007). Attitudes as object-evaluation associations of varying strength. *Social Cognition*, 25, 603-637.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology, 69,* 1013-1027.

Fazio. R. H., & Olson, M. A. (2014). The MODE model; Attitude-behavior processes as a function of motivation and opportunity. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 155-171). New York: Guilford Press.

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

Ferguson, M. J., Mann, T. C., & Wojnowicz, M. T. (2014). Rethinking duality: Criticisms and ways forward. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 578-594). New York: Guilford Press.

Ferreira, M. B., Garcia-Marques, L., Sherman, S. J., & Sherman, J. W. (2006). A dual-process approach to judgment under uncertainty. *Journal of Personality and Social Psychology, 91,* 797-813.

Festinger, L. (1957). *A theory of cognitive dissonance.* Evanston, IL: Row Peterson.

Fiedler, K., & Hütter, M. (2014). The limits of automaticity. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 497-513). New York: Guilford Press.

Fiske, S. T., Lin, M., & Neuberg, S. L. (1999). The continuum model: Ten years later. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 231-254). New York: Guilford Press.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology, 23,* 1-74.

Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review*, *5*, 5-15.

Friese, M., Hofmann, W., & Schmitt, M. (2008). When and why do implicit measures predict behavior? Empirical evidence for the moderating role of opportunity, motivation, and process reliance. *European Review of Social Psychology, 19,* 285-338.

Gaertner, S. L., & Dovidio, J. F. (1986). The aversive form of racism. In J. F. Dovidio, & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 61-89). Orlando, FL: Academic Press.

Gawronski, B. (2004). Theory-based bias correction in dispositional inference: The fundamental attribution error is dead, long live the correspondence bias. *European Review of Social Psychology, 15,* 183-217.

Gawronski, B. (2013). What should we expect from a dual-process theory of preference construction in choice? *Journal of Consumer Psychology, 23,* 556-560.

Gawronski, B., Armstrong, J., Conway, P., Friesdorf, R., & Hütter, M. (2017). Consequences, norms, and generalized inaction in moral dilemmas: The CNI model of moral decision-making. *Journal of Personality and Social Psychology, 113,* 343-376.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132,* 692-731.

Gawronski, B., & Bodenhausen, G. V. (2009). Operating principles versus operating conditions in the distinction between associative and propositional processes. *Behavioral and Brain Sciences, 32,* 207-208.

Gawronski, B., & Bodenhausen, G. V. (2011). The associative-propositional evaluation model: Theory, evidence, and open questions. *Advances in Experimental Social Psychology, 44,* 59-127.

Gawronski, B., & Bodenhausen, G. V. (2014). The associative-propositional evaluation model: Operating principles and operating conditions of evaluation. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 188-203). New York: Guilford Press.

Gawronski, B., & Bodenhausen, G. V. (2015a). Social-cognitive theories. In B. Gawronski, & G. V. Bodenhausen (Eds.), *Theory and explanation in social psychology* (pp. 65-83). New York: Guilford Press.

Gawronski, B., & Bodenhausen, G. V. (2015b). Theory evaluation. In B. Gawronski, & G. V. Bodenhausen (Eds.), *Theory and explanation in social psychology* (pp. 3-23). New York: Guilford Press.

Gawronski, B., & Brannon, S. M. (2019). Attitudes and the implicit-explicit dualism. In D. Albarracín & B. T. Johnson (Eds.), *The handbook of attitudes. Volume 1: Basic principles* (2nd edition, pp. 158-196). New York: Routledge.

Gawronski, B., Brannon, S. M., & Bodenhausen, G. V. (2017). The associative-propositional duality in the representation, formation, and expression of attitudes. In R. Deutsch, B. Gawronski, & W. Hofmann (Eds.), *Reflective and impulsive determinants of human behavior* (pp. 103-118). New York: Psychology Press.

Gawronski, B., Conway, P., Hütter, M., Luke, D. M, Armstrong, J., & Friesdorf, R. (2020). On the validity of the CNI model of moral decision-making: Reply to Baron and Goodwin (2020). *Judgment and Decision Making, 15,* 1054-1072.

Gawronski, B., & Creighton, L. A. (2013). Dual-process theories. In D. E. Carlston (Ed.), *The Oxford handbook of social cognition* (pp. 282-312). New York: Oxford University Press.

Gawronski, B., & LeBel, E. P. (2008). Understanding patterns of attitude change: When implicit measures show change, but explicit measures do not. *Journal of Experimental Social Psychology, 44,* 1355-1361.

Gawronski, B., Sherman, J. W., & Trope, Y. (2014). Two of what? A conceptual analysis of dual-process theories. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 3-19). New York: Guilford Press.

Gawronski, B., & Strack, F. (2004). On the propositional nature of cognitive consistency: Dissonance changes explicit, but not implicit attitudes. *Journal of Experimental Social Psychology, 40,* 535-542.

Gilbert, D. T. (1989). Thinking lightly about others: Automatic components of the social inference process. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 189-211). New York: Guilford Press.

Gilbert, D. T. & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin, 117,* 21-38.

Gilbert, D. T., Pelham, B. W., & Krull, D. S. (1988). On cognitive busyness: When person perceivers meet persons perceived. *Journal of Personality and Social Psychology, 54,* 733-740.

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of*

intuitive judgment. Cambridge, UK: Cambridge University Press.

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. *Advances in Experimental Social Psychology, 47,* 55-130.

Greeley, A. M., & Sheatsley, P. B. (1971). Attitudes toward racial integration. *Scientific American, 225,* 13-19.

Greene, J. D. (2008). The secret joke of Kant's soul. In W. Sinnott-Armstrong (Ed.), *Moral psychology. The neuroscience of morality* (Vol. 3, pp. 35-79). Cambridge, MA: MIT Press.

Greene, J. (2014). Beyond point-and-shoot morality: Why cognitive (neuro)science matters for ethics. *Ethics*, *124*, 695-726.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*, 2105-2108.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. K. L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74,* 1464-1480.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*, 814-834.

Heycke, T., & Gawronski, B. (2020). Co-occurrence and relational information in evaluative learning: A multinomial modeling approach. *Journal of Experimental Psychology: General, 149,* 104-124.

Higgins, E. T. (1996). Knowledge activation: Accessibility, applicability, and salience. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 133-168). New York: Guilford Press.

Hütter, M., & Klauer, K. C. (2016). Applying processing trees in social psychology. *European Review of Social Psychology, 27,* 116-159.

Hütter, M., & Rothermund, K. (2020). Automatic processes in evaluative learning. *Cognition and Emotion, 34,* 1-20.

Hugenberg, K., & Bodenhausen, G. V. (2004). Ambiguity in social categorization: The role of prejudice and facial affect in racial categorization. *Psychological Science, 15,* 342-345.

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

Jones, E. E. & Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. *Advances in Experimental Social Psychology*, *2,* 219- 266.

Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology, 3,* 1-24.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58,* 697-720.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3,* 430-454.

Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska symposium on motivation* (pp. 192-240). Lincoln: University of Nebraska Press.

Kelley, H. H. (1972). Causal schemata and the attribution process. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: perceiving the causes of behavior* (pp. 151-174). Morristown, NJ: General Learning Press.

Keren, G. (2013). A tale of two systems: A scientific advance or a theoretical stone soup? *Perspectives on Psychological Science, 8,* 257-262.

Keren, G., & Schul, Y. (2009). Two is not always better than one: A critical evaluation of two-systems theories. *Perspectives on Psychological Science, 4,* 533-550.

Klauer, K. C. (2015). Mathematical modeling. In B. Gawronski, & G. V. Bodenhausen (Eds.), *Theory and explanation in social psychology* (pp. 371-389). New York: Guilford Press.

Kohlberg, L (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D.A. Goslin (Ed.), *Handbook of socialization theory and research* (pp. 347-480). Chicago, IL: Rand McNally.

Kohlberg, L. (1976). Moral stages and moralization: The cognitive-developmental approach. In T. Lickona (Ed.), *Moral Development and behavior: Theory, research, and social issues* (pp. 31-53). New York: Hot, Rinehart, and Winston.

Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberative judgments are based on common principles. *Psychological Review, 118,* 97-109.

Kruglanski, A. W., Klein, K. M., Pierro, A., & Mannetti, L. (2014). The unimodal unfolding. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 514-529). New York: Guilford Press.

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

Krull, D. S. (1993). Does the grist change the mill? The effect of the perceiver's inferential goal on the process of social inference. *Personality and Social Psychology Bulletin, 19,* 340-348.

Krull, D. S. & Erickson, D. J. (1995). Judging situations: On the effortful process of taking dispositional information into account. *Social Cognition, 13,* 417-438.

Kukken, N., Hütter, M., & Holland, R. (2020). Are there two independent evaluative conditioning effects in relational paradigms? Dissociating effects of CS-US pairings and their meaning. *Cognition and Emotion, 34,* 170-187.

Kunda, Z., Miller, D. T., & Claire, T. (1990). Combining social concepts: The role of causal reasoning. *Cognitive Science, 14,* 551-577.

Kunda, Z., & Sherman-Williams, B. (1993). Stereotypes and the construal of individuating information. *Personality and Social Psychology Bulletin, 19,* 90-99.

Kurdi, B., & Dunham, Y. (2020). Propositional accounts of implicit evaluation: Taking stock and looking ahead. *Social Cognition, 38,* s42-s67.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos, & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91-195). Cambridge: Cambridge University Press.

Leyens, J.-P., & Yzerbyt, V. Y. (1992). The ingroup overexclusion effect: The impact of valence and confirmation on information search. *European Journal of Social Psychology, 22,* 549-569.

Lieberman, M. D. (2003). Reflective and reflexive judgment processes: A social cognitive neuroscience approach. In J. P. Forgas, K. R. Williams, & W. von Hippel (Eds.), *Social judgments: Implicit and explicit processes* (pp. 44-67). New York: Cambridge University Press.

Lieberman, M. D. (2009). What zombies can't do: A social cognitive neuroscience approach to the irreducibility of reflective consciousness. In J. St. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 293-316). Oxford: Oxford University Press.

Lieberman, M. D., Gaunt, R., Gilbert, D. T., & Trope, Y. (2002). Reflection and reflexion: A social cognitive neuroscience approach to attributional inference. *Advances in Experimental Social Psychology, 34,* 199-249.

Maheswaran, D., & Chaiken, S. (1991). Promoting systematic processing in low-motivation settings: Effects of incongruent information on processing and judgment. *Journal of Personality and Social Psychology, 61,* 13-25.

Maheswaran, D., Mackie, D. M., & Chaiken, S. (1992). Brand name as an heuristic cue: The effects of task importance and expectancy confirmation on consumer judgments. *Journal of Consumer Psychology, 1,* 317-336.

Mandelbaum, E. (2016). Attitude, inference, association: On the propositional structure of implicit bias. *Noûs, 50,* 629-658.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York: Freeman.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102,* 419-457.

McConahay, J. B. (1986). Modern racism, ambivalence, and the modern racism scale. In J. D. Dovidio, & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 91-125). Orlando, FL: Academic Press.

Melnikoff, D. E., & Bargh, J. A. (2018). The mythical number two. *Trends in Cognitive Sciences, 22,* 280-293.

Metcalfe, J., & Mischel, W. (1999). A hot/cool-system analysis of delay of gratification. Dynamics of willpower. *Psychological Review, 106,* 3-19.

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences*, *32*, 183-198.

Moors, A. (2014). Examining the mapping problem in dual-process models. In J. W., Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 20-34). New York: Guilford Press.

Moors, A. (2016). Automaticity: Componential, causal, and mechanistic explanations. *Annual Review of Psychology, 67,* 263-287.

Moors, A., & De Houwer, J. (2006). Automaticity: A conceptual and theoretical analysis. *Psychological Bulletin, 132*, 297-326.

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research* (pp. 1-18). New York: Plenum.

Osman, M. (2013). A case study: Dual-process theories of higher cognition. *Perspectives on Psychological Science, 8,* 248-252.

Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, *81*, 181-192.

Payne, B. K., & Cameron, C. D. (2014). Dual-process theory from a process dissociation perspective. In J. W., Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 107-120). New York: Guilford Press.

Payne, B. K., & Correll, J. (2020). Race, weapons, and the perception of threat. *Advances in Experimental Social Psychology, 62,* 1-50.

Petty, R. E., & Cacioppo, J. T. (1986). The Elaboration Likelihood Model of persuasion. *Advances in Experimental Social Psychology, 19,* 123-205.

Petty, R. E., Cacioppo, J. T., & Schumann, D. W. (1983). Central and peripheral routes to advertising effectiveness: The moderating role of involvement. *Journal of Consumer Research, 10*, 135-146.

Platt, J. R. (1964). Strong inference. *Science, 146,* 347-353.

Popper, K. R. (1934). *Logic of scientific discovery*. New York: Basic Books.

Quattrone, G. A. (1982). Overattribution and unit formation: When behavior engulfs the person. *Journal of Personality and Social Psychology, 42,* 593-607.

Quine, W. V. O. (1953). Two dogmas of empiricism. In W. V. O. Quine, *From a logical point of view* (pp. 20-46). Cambridge, MA: Harvard University Press.

Quine, W. V. O., & Ullian, J. S. (1978). *The web of belief* (2nd ed.). New York: McGraw-Hill.

Richards, Z., & Hewstone, M. (2001). Subtyping and subgrouping: Processes for the prevention and promotion of stereotype change. *Personality and Social Psychology Review, 5,* 52–73.

Ross, L. D. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. *Advances in Experimental Social Psychology, 10,* 173-220.

Ross, L. D., Amabile, T. M., & Steinmetz, J. L. (1977). Social roles, social control, and biases in social-perception processes. *Journal of Personality and Social Psychology, 35,* 485-494.

Sears, D. O. (1988). Symbolic racism. In P. A. Katz, & D. A. Taylor (Eds.), *Eliminating racism: Profiles in controversy* (pp. 53-84). New York: Plenum.

Sherman, J. W., Gawronski, B., Gonsalkorale, K., Hugenberg, K., Allen, T. J., & Groom, C. J. (2008). The self-regulation of automatic associations and behavioral impulses. *Psychological Review, 115,* 314-335.

Sherman, J. W., Gawronski, B., & Trope, Y. (Eds.). (2014). *Dual-process theories of the social mind*. New York: Guilford Press.

Sherman, J. W., Krieglmeyer, R., & Calanchini, J. (2014). Process models require process measures. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.). *Dual process theories of the social mind* (pp. 121-138). New York: Guilford Press.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119,* 3-22.

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review, 4,* 108-131.

Snyder, M. L. & Frankel, A. (1976). Observer bias: A stringent test of behavior engulfing the field. *Journal of Personality and Social Psychology, 34,* 857-864.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences, 23,* 645-726.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review, 8,* 220-247.

Taylor, D. G., Sheatsley, P. B., & Greeley, A. M. (1978). Attitudes toward racial integration. *Scientific American, 238,* 42-49.

Teeny, J., Briñol, P., & Petty, R. E. (2016). The elaboration likelihood model: Understanding consumer attitude change. In C. V. Jansson-Boyd & M. J. Zawisza (Eds.). *Routledge international handbook of consumer psychology* (pp. 390-410). Abingdon, UK: Routledge.

Thomson, J. J. (1976). Killing, letting die, and the trolley problem. *The Monist, 59,* 204-217.

Trafimow, D. (2015). Rational actor theories. In B. Gawronski, & G. V. Bodenhausen (Eds.), *Theory and explanation in social psychology* (pp. 245-265). New York: Guilford Press.

Trope, Y. (1986). Identification and inferential processes in dispositional attribution. *Psychological Review, 93*, 239-257.

Trope, Y. & Alfieri, T. (1997). Effortfulness and flexibility of dispositional judgment processes. *Journal of Personality and Social Psychology, 73,* 662-674.

Trope, Y. & Cohen, O. (1989). Perceptual and inferential determinants of behavior-correspondent attributions. *Journal of Experimental Social Psychology, 25,* 142-158.

Trope, Y. & Gaunt, R. (1999). A dual-process model of overconfident attributional inferences. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 161-178). New York: Guilford Press.

Trope, Y. & Gaunt, R. (2000). Processing alternative explanations of behavior: Correction or integration? *Journal of Personality and Social Psychology, 79,* 344–354.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124-1131.

Tversky, A., & Kahneman, D. (1983). Extensional vs. intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90,* 293-315.

Van Dessel, P., Gawronski, B., & De Houwer, J. (2019). Does explaining social behavior require multiple memory systems? *Trends in Cognitive Sciences, 23,* 368-369.

Vonk, R. (1999). Effects of outcome dependency on correspondence bias. *Personality and Social Psychology Bulletin, 25,* 382-389.

Whitfield, M., & Jordan, C. H. (2009). Mutual influences of explicit and implicit attitudes. *Journal of Experimental Social Psychology, 45,* 748-759.

Wicker, A. W. (1969). Attitudes versus actions: The relationship of verbal and overt behavioral responses to attitude objects. *Journal of Social Issues, 25,* 41-78.

Wilson, T. D., Lindsey, S. & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review, 107,* 101-126.

Author Note

The current chapter is a revised and updated version of an earlier chapter with the same title by Gawronski and Creighton (2013) in the first edition of this handbook. Preparation of the current chapter was supported by National Science Foundation Grant BCS-1649900 to Bertram Gawronski. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

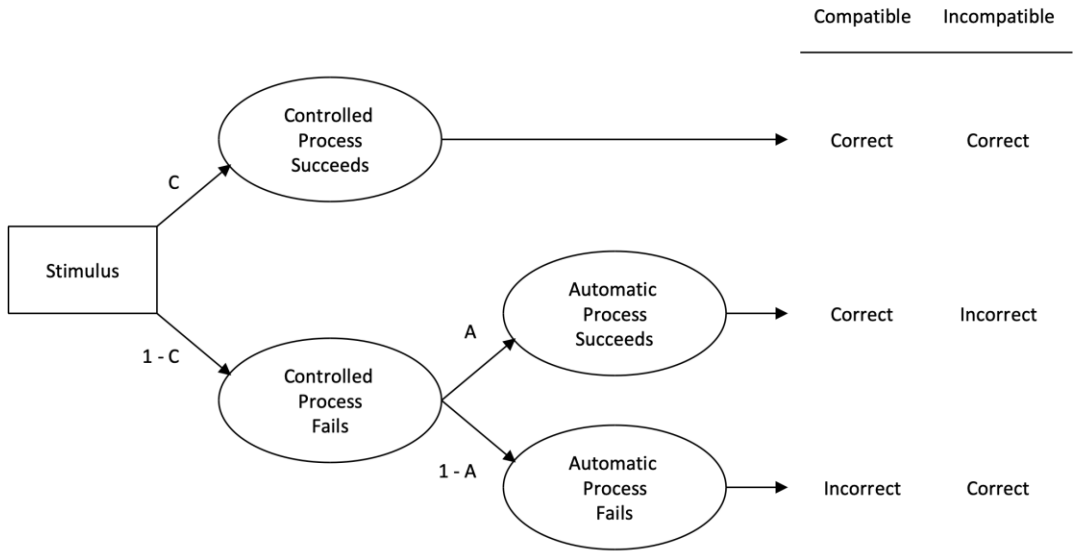

Figure 1. Process dissociation model of automatic and controlled processing. Figure adapted from Conrey, Sherman, Gawronski, Hugenberg, and Groom (2005). Reprinted with permission.