

Causal Conceptions in Social Explanation and Moral Evaluation: A Historical Tour

Mark D. Alicke¹, David R. Mandel², Denis J. Hilton³,
Tobias Gerstenberg⁴, and David A. Lagnado⁵

¹Department of Psychology, Ohio University; ²Socio-Cognitive Systems Section, Defence Research and Development Canada (DRDC) Toronto, Ontario, Canada, and Department of Psychology, York University, Ontario, Canada; ³Department of Psychology, University of Toulouse, France; ⁴Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology; and ⁵Cognitive, Perceptual, and Brain Sciences Department, University College London, United Kingdom

Perspectives on Psychological Science
2015, Vol. 10(6) 790–812
© Her Majesty the Queen in Right of
Canada, as represented by Defence
Research and Development
Canada 2015
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1745691615601888
pps.sagepub.com



Abstract

Understanding the causes of human behavior is essential for advancing one's interests and for coordinating social relations. The scientific study of how people arrive at such understandings or explanations has unfolded in four distinguishable epochs in psychology, each characterized by a different metaphor that researchers have used to represent how people think as they attribute causality and blame to other individuals. The first epoch was guided by an "intuitive scientist" metaphor, which emphasized whether observers perceived behavior to be caused by the unique tendencies of the actor or by common reactions to the requirements of the situation. This metaphor was displaced in the second epoch by an "intuitive lawyer" depiction that focused on the need to hold people responsible for their misdeeds. The third epoch was dominated by theories of counterfactual thinking, which conveyed a "person as reconstructor" approach that emphasized the antecedents and consequences of imagining alternatives to events, especially harmful ones. With the current upsurge in moral psychology, the fourth epoch emphasizes the moral-evaluative aspect of causal judgment, reflected in a "person as moralist" metaphor. By tracing the progression from the person–environment distinction in early attribution theories to present concerns with moral judgment, our goal is to clarify how causal constructs have been used, how they relate to one another, and what unique attributional problems each addresses.

Keywords

causal judgment, attribution, moral judgment, social perception

Whether people can initiate their own actions is the central question in perennial debates about free will, but there is no dispute as to whether people are perceived to cause their own actions. Not only are people viewed as prime movers of their behavior, their actions are presumed to be impelled by various intentions and motives. Causation is the link that cements these prior mental states to the outcomes that ensue and, along with intention, is the primary criterion for evaluating social behavior. The process by which people derive causal inferences about human action has occupied more than 50 years of psychological theory and research.

The study of how causal inferences are made in the service of evaluating social behavior has unfolded in four overlapping, but distinguishable, research epochs in psychology, each emphasizing different aspects of causal

judgment (see Fig. 1 for an overview). Each epoch was guided by a metaphor that represented researchers' beliefs about the layperson's primary orientation in assessing behavioral causation. The first epoch consists of the classic attribution period of the 1960s and 1970s. The guiding metaphor during this period was that people acted as "intuitive scientists" in deciding whether an observed behavior reflected something unique about the person who performed it or was instead a typical reaction to features of the situation in which it occurred (Heider, 1958; Kelley, 1967).

Corresponding Author:

Mark D. Alicke, Department of Psychology, Ohio University, 229
Porter Hall, Athens, OH 45701
E-mail: alicke@ohio.edu

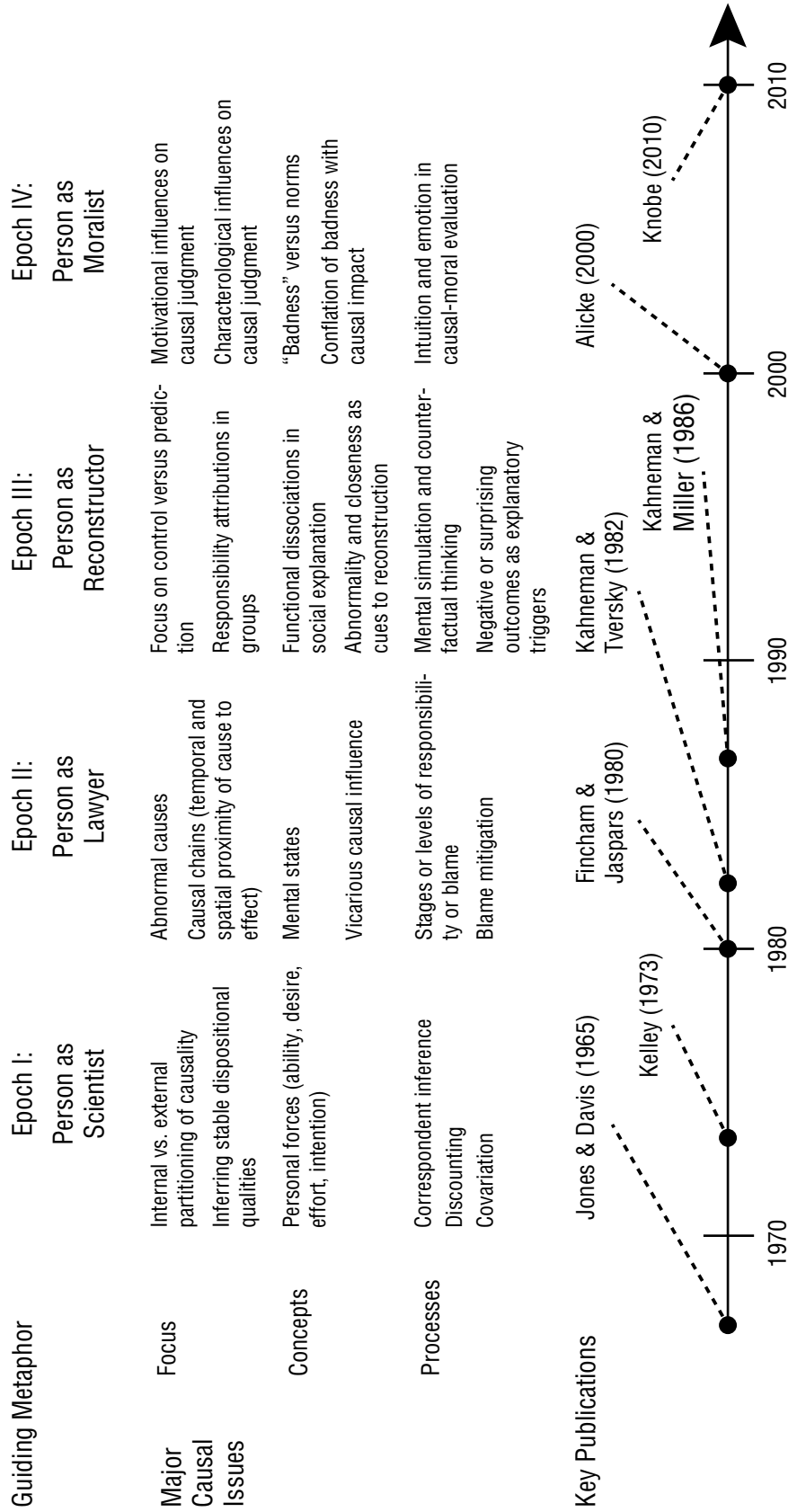


Fig. 1. Timeline of causal judgment contributions.

The second epoch was ushered in when the intuitive scientist metaphor of causal judgment gave way to an “intuitive-lawyer” metaphor (Fincham & Jaspars, 1980). This approach derived from the legal philosophers Hart and Honoré (1959) and emphasized the aspects of causal influence that were relevant for ascribing responsibility. The intuitive lawyer perspective in Epoch 2 likened lay attribution to the legal decision-making process of assessing whether the actor’s causal role in a harmful event warranted civil or criminal sanctions—the main difference being that lay attributors concentrated on everyday social misdeeds rather than on criminal offenses.

The third “person-as-reconstructor” epoch in causal judgment research was inaugurated with Kahneman and Tversky’s (1982) article on the simulation heuristic, followed a few years later by Kahneman and Miller’s (1986) norm theory. Although the study of counterfactual reasoning soon ramified into a sophisticated research area involving perceptions of causation, control, preventability, and regret (for overviews, see Mandel, Hilton, & Catellani, 2005; Roese & Olson, 1995), it began with the observation that surprising or counternormative events, especially unfortunate ones, impel observers to consider how things might have turned out differently, which, in turn, influences their cause and blame attributions. In this regard, the third epoch represents another step on the path from viewing people as objective causal analysts to individuals whose attributions serve multiple goals and motives.

The fourth epoch in the causal judgment literature completes the cycle from the person as scientist to the “person as moralist” (Knobe, 2010; Tetlock, 2002) in that it pertains to how people appraise and form moral judgments about behavior. In the person-as-moralist perspective, blame and praise are placed at the center of causal investigations of behavior and its consequences. The fourth epoch represents a radical departure from the causal conceptions of previous epochs by positing a bidirectional relationship between causation and blame: Not only does the perception that an actor caused a harmful outcome lead to increased blame (Mandel, 2010) but ascriptions of blame influence causal attributions (Alicke, 2000; Alicke, Rose, & Bloom, 2011).

Across the four epochs that we identify, the content and goal of causal analysis shift increasingly from identifying the conditions that enable observers to make dispositional inferences during the person-as-scientist epoch to understanding the role of causal judgment in evaluations of blame and morality in the person-as-moralist epoch.

The Present Article

With the ascendance of moral psychology among social, cognitive, and evolutionary researchers, our goal in this

article is to trace the history of the causal construct in each of the four research epochs as it applies to assessments of blame and moral character. Although causation is the thread that conjoins these research epochs, it has been treated as a separate topic within each, and, as a result, a complete picture has never been drawn of how different facets of causation are linked to blame and moral judgment. By showing how each of the aspects of causation that was emphasized during the four epochs is relevant to understanding the process by which people assess blame and moral character and express approbation or disapprobation for behavior and its consequences, we hope to show not only how older causal ideas are related to new ones but also how current perspectives in moral psychology can benefit if researchers reconsider and incorporate earlier causal conceptions.

Epoch 1: Causation in Classic Attribution Approaches

In the classic attributional approach that dominated the study of person perception from the late 1960s through much of the 1970s, researchers modeled lay causal judgment on traditional scientific notions of causation. Consistent with regularity theories (Hume, 1748/2007; Mill, 1843), people were assumed to derive causal beliefs from patterns of covariation among events. The same formal tools (e.g., analysis of variance [ANOVA]) that scientists used to infer strengths of fertilizers were approximated by laypeople to infer strengths of character (cf. Gigerenzer, 1991). What came to be known as “attribution theory” originated in Fritz Heider’s seminal writings in which he introduced the distinction between personal and situational contributions to behavior and its outcomes. Although Heider’s approach to person perception was the impetus for later attributional developments, the covariation approach was developed most fully in Harold Kelley’s ANOVA model of causal assignment.

Heider’s naïve analysis of action

One of the central and enduring issues in the psychology of perception concerns how people maintain invariant perceptions of environmental stimuli that are constantly in flux. In the seminal publication *The Psychology of Interpersonal Relations*, Heider (1958) transported this issue to the study of human action—that is, to the question of how perceivers infer stable dispositional qualities in others (e.g., their introversion or extraversion) on the basis of observations of fluctuating behaviors (e.g., loud and expressive in some situations, quiet and reserved in others). As Malle (2004, 2011) noted, Heider’s analysis of personal causality had two distinct strands that were not always carefully distinguished in later theorizing: the distinction between personal and environmental forces and

between personal and impersonal causality. The former is most relevant to the classic attributional goal of identifying stable properties of the person or the environment, whereas the latter—which embodies Heider's distinction between goal-directed, purposive actions and those that are inadvertent or accidental—speaks more directly to central issues in blame and moral evaluation.

In the early history of attribution theory, however, the most influential aspect of Heider's (1958) book for subsequent attributional developments was the person-environment dichotomy that Kelley (1967, 1971) subsequently refined. The observer's task in Heider's analysis was to disentangle the personal and environmental forces that conspired to create the event and its effects. Given that the event occurred, the observer works backward to ascertain what the person and the environment each contributed to it. Because Heider's view of causation was informed by basic perceptual principles, his goal was to identify the conditions that produce attributions of invariant dispositional properties against the background of facilitating and inhibiting environmental forces. His emphasis, therefore, was on the personal forces (i.e., ability, desire, effort) that instigate an event unless thwarted by environmental impediments. According to this view, abilities and desires impel goal-directed behavior, and environmental conditions facilitate or inhibit these objectives.

Heider was probably the first theorist in the *person perception* tradition to include mentalistic components in his conception of personal causation. In fact, the conflation of causal and mentalistic concepts in the perception of events was adumbrated in Heider and Simmel's (1944) demonstration that people spontaneously construe the movement of geometric figures in terms of emotions and motives. For example, three geometrical shapes moving in and around a square were described by almost all participants with animistic phrases, such as objects chasing each other, being angry with one another, and so forth. These findings point naturally toward a view of causation based on human action concepts, such as intent and motive.

A crucial aspect of Heider's (1958) personal force conception is that intention is treated as an integral component of personal causation. Heider assumed that personal causation will be discounted if an effect is brought about inadvertently, accidentally, or under compulsion. In Heider's view, unintended actions do not reveal stable dispositional characteristics and do not, therefore, qualify as personal causes. This view contrasts with most philosophical and psychological perspectives in which causation and intention are treated as separate components of an action sequence. If a wedding guest, Ralph, slips on the floor and knocks over the cake, he is certainly the cause of the cake's derangement and the

wedding party's dismay, even if he had no desire to cause the mishap. In Heider's approach, however, the main task in attributing dispositional qualities to actors is less about identifying the necessary or sufficient conditions for producing an effect than of estimating whether environmental forces were sufficient to overcome personal ones, thereby negating dispositional inferences. Because accidents lack intent, there is scant reason to claim that Ralph is a vindictive, cake-smashing sort of person, although other attributions may apply (e.g., that Ralph is clumsy).

Correspondent inference and discounting

The empirical studies in which researchers laid the groundwork for classic attributional theorizing demonstrated what eventually came to be known as the discounting principle. These studies (e.g., Jones, Davis, & Gergen, 1961; Thibaut & Riecken, 1955) showed that situational pressures (e.g., responding to a high-status person's request) led observers to discount the role of internal dispositions (e.g., generosity, introversion) in producing the behavior in question (e.g., agreeing to donate blood, behaving quietly in an interview for a position as a submariner).

However, later findings with other experimental paradigms (e.g., the attitude-attribution paradigm of Jones & Harris, 1967) revealed a failure to discount situational factors sufficiently, leading to the identification of a *correspondence bias* (Jones, 1979; Jones & Davis, 1965; also known as the *fundamental attribution error*; Gilbert & Malone, 1995; Ross, Amabile, & Steinmetz, 1977), whereby people fail to account sufficiently for situational pressures when evaluating behavioral causes. This claim dovetailed with the emerging view in the judgment and decision-making literature of the 1970s and 1980s that people fail to respect normative inferential models (Kahneman, Slovic, & Tversky, 1982; Nisbett & Ross, 1980). Although insufficient discounting of situational pressures undoubtedly occurs (e.g., Bierbrauer, 1979), we note that this tendency is far from inevitable (for relevant reviews, see Gawronski, 2004; Hastie, in press; Hilton, 2007; McClure, 1998). Moreover, Morris and Larrick's (1995) probabilistic analysis of discounting (going beyond the simple ANOVA model) showed that people's failure to discount is sometimes rationally justified (see also Walker, Smith, & Vul, 2015). Finally, insufficient discounting of situational factors could also be motivated by social functionalist objectives other than accurate assessment, such as being an "intuitive prosecutor" who would rather err on the side of ascribing too much blame to wrongdoers than to err on the side of ascribing too little (Hamilton, 1980; Tetlock, 2002).

Furthermore, cross-cultural research has shown a general tendency for East Asians to favor situational attributions more than Americans do (J. G. Miller, 1984; Morris, Nisbett, & Peng, 1995), although it is not clear whether this is due to their having different beliefs about the general relationship of situational pressures to behavior (Kashima, Siegal, Tanaka, & Kashima, 1992) or to their reasoning differently from the same information. It is important to note that in most of the studies conducted in the four epochs we review here, researchers have relied on participant samples from “WEIRD” societies—namely, those that are Western, Educated, Industrialized, Rich, and Democratic (Henrich, Heine, & Norenzayan, 2010).

As we elaborate in Epoch 2, discounting is an important aspect of blame and responsibility judgments. Although in classic attributional approaches researchers carried over Heider's (1958) person–situation scheme, they actually studied a very limited range of situational pressures, such as acceding to the request of a high-status person, or in the case of correspondence bias, enacting role-consistent behavior. In the context of deciding whether or to what extent a person is to blame for harmful actions, however, a much larger range of constraints is potentially influential, such as the actor's background, cognitive capacities, and emotional vulnerabilities as well as situational provocations, temptations, and compulsions. One lesson to be gleaned from attributional research in Epoch 1 is that people (at least in Western cultures) are generally conservative in mitigating blame on these bases; that is, they often give short shrift to situational pressures in ascribing blame.

Covariation and the ANOVA model

Heider's (1958) seminal attribution book was sympathetically reviewed by Kelley (1960), who adopted Heider's suggestion that the layperson was like a scientist who used Mill's (1843) canons of inductive logic to determine the locus of causality—namely, that a cause is “that condition which is present when the effect is present and which is absent when the effect is absent” (Heider, 1958 p. 15). Kelley's covariation model brought the intuitive scientist metaphor into full bloom. Unfortunately, Kelley's analogy with statistical techniques—specifically, the ANOVA—to convey a theory of commonsense inference proved to be a double-edged sword, as the layperson's apparent deviation from the canons of scientific inference led to the blanket assumption that people were poor, rather than simply naïve, scientists (e.g., Kahneman et al., 1982; Nisbett & Ross, 1980). We briefly illustrate this model and show how deviations from it were sometimes overstated.

Kelley's model is centered on the roles of consistency, distinctiveness, and consensus information in determining

personal or situational attributions. According to Kelley's covariational formula, assessing these variables requires information about the actor's behavior toward the same object or entity over repeated observations (consistency), toward other entities of the same class (distinctiveness), and of other people's interactions with the same entity (consensus). Consider an example in which Alice wonders whether she was at fault for hurting her roommate Betty's feelings. Alice might review evidence about her propensity to hurt others and any available information about Betty's susceptibility to being hurt. If Alice notes that many other people hurt Betty's feelings (high consensus), whereas she (Alice) rarely hurts other people's feelings (high distinctiveness), then this would point to a situational attribution (e.g., Betty is oversensitive) given that the effect (hurt feelings) covaries with Betty's presence (i.e., the hurt feelings occur only when she is the partner in a social interaction but not otherwise). Studies showed that observers could indeed make the distinctions specified in Kelley's model (e.g., Cheng & Novick, 1990; Försterling, 1989; Hilton, Smith, & Kim, 1995; Van Overwalle, 1997). These studies also suggested that the widespread assumption that people ignore or underuse consensus information (a claim based on a particular response methodology; e.g., McArthur, 1972; Orvis, Cunningham, & Kelley, 1975) was exaggerated. This was an important counterweight to the depiction of the layperson as a poor scientist (see Hilton, 2007, 2012, for other examples).

The cogency of Kelley's ANOVA model accounts for its being the only one of the original attribution theories that continues to receive significant attention in social psychology textbooks and to have a widespread influence outside of social psychology (Quayle & Naidoo, 2012). Nevertheless, the ANOVA model has serious limitations, as Kelley (1973) also recognized. First, the crucial distinction that Heider introduced between personal and impersonal causality is ignored, and instead, causes are partitioned into the internal–external scheme. The internal–external model works well for establishing whether the actor's behavior reflects something relatively unique about his or her personality or ability, but it has no provisions for evaluating the actor's intentions and motives, which are the fundamental criteria for assessing blame and moral character. Second, research inspired by the ANOVA model tends to depict an actor's behavior toward a passive stimulus or entity, which deemphasizes the interpersonal behaviors that give rise to moral evaluations, typically involving two active agents who may interact in complex ways (Gray, Young, & Waytz, 2012). Third, independence assumptions are made in the ANOVA model in which more complex causal chains and networks are ignored, reducing their applicability to real-world causal problems (see Kelley, 1983).

Summary of Epoch 1

It is fair to say that Heider's (1958) theory was better equipped to address problems in blame and moral judgment than the wave of attribution models that immediately followed. Heider's notion of personal force combined both the behavioral and mental (i.e., intent) elements that are essential to moral evaluation, and in fact, he applied his naïve analysis of action directly to the problem of ascribing responsibility. Although inspired by Heider's writings, the next generation of attribution theories emphasized the problem of inferring dispositional invariances against the background of facilitative and inhibitory environmental forces. Kelley's sophisticated formalization of the internal–external problem as a covariation analysis is particularly useful for predicting how the actor will behave in future situations and for evaluating one's own behavior and outcomes in the same situation. Although Kelley's ANOVA model did not emphasize blame ascription or moral evaluation, identifying stable properties of the actor or situation is clearly relevant for these judgments. If the environment is the prepotent cause of a harmful event, then the actor's culpability for harmful outcomes is mitigated or negated. More generally, information about personal norms (distinctiveness and consistency) is vital for inferring the intentions or motives that underlie the actor's behavior, whereas general norms (consensus) are perhaps the most important consideration in evaluating the moral appropriateness of the actor's behavior.

Nevertheless, classic attributional analyses from the person-as-scientist epoch are not particularly well-suited to handle events that include, for example, multiple causes, actors in different roles with different responsibilities, causal preemption, and causal overdetermination. To understand such events, we need a broader repertoire of concepts than those found in the ANOVA model, for which we turn to the intuitive-lawyer perspective reflected in the second research epoch.

Epoch 2: The Legal Model

The advent of the “person-as-lawyer” metaphor represented a shift from the person–environment problem of classic attribution theory to a broader interest in attributing responsibility for harmful or potentially harmful events. Influential articles by Hamilton (1976, 1978, 1980), Lloyd-Bostock (1983), and especially Fincham and Jaspars (1980) transported issues in criminal and tort law to the attributional context. Fincham and Jaspars, drawing heavily on the legal philosopher Herbert Hart (e.g., Hart, 1968; Hart & Honoré, 1959), showed how traditional attributional approaches were inadequate to address complex issues in responsibility, including intervening causes;

omissions to act; unintentional, negligent, or reckless harm-doing; vicarious responsibility; and failed attempts.

An influential contribution of Hart and Honoré (1959) to the study of responsibility and blame was to note that lay observers and legal decision makers typically seek causal explanations for individual events rather than universal covering laws accounting for relations among event types. If there is a fire in an operating room, for example, laws describing the relationship between oxygen and combustion miss the point; what observers want to know is what caused this particular fire. Causal explanations in blame and responsibility ascription, therefore, typically invoke abnormal conditions, such as the presence of a lighted cigarette in the operating room (e.g., Hilton & Slugoski, 1986). Furthermore, voluntary and deliberate human actions are especially likely to be cited as causes because they are perceived as “self-willed” and are, in this regard, ultimate causes beyond which no further influences need be traced (e.g., Hilton, McClure, & Sutton, 2010). When multiple causes compete for priority, therefore, abnormal conditions and voluntary, deliberate human actions are likely to be seen as the most influential causes of the event's outcomes.

This difference in causal orientation between identifying stable properties of a person or the environment and deciding whether a human act caused a harmful outcome was a primary impetus for Fincham and Jaspars's (1980) metaphor shift from scientist to lawyer. If the attribution theories of Epoch 1 gave short shrift to blame and moral judgment, the intuitive-lawyer approach placed these issues front and center.

The person-as-lawyer metaphor that governed Epoch 2 highlighted the lay attributor's use of blame or responsibility criteria—such as intent, foresight, foreseeability, and (of course) causation—and broadened the scope of events from the type that Epoch 1 emphasized to include those with multiple causes and multiple agents, outcomes that occur or are thwarted in circuitous ways, and agents who are only remotely connected to the eventual harm (e.g., a bartender being held responsible for damages caused by a drunken patron).

Early studies during this period provided empirical tests of “stage” theories of responsibility attribution, including Shultz, Schleifer, and Altman's (1981; also see Shultz & Schleifer, 1983) simple entailment model, which posited that responsibility entailed causation on the basis of Piaget's (1932) developmental stages (Sedlak, 1979). During this period, Heider's conception of responsibility was formalized into inferential stages of association, causality, foreseeability, intentionality, and supererogation (Fincham & Jaspars, 1979; for an earlier test of Heider's model, see Shaw & Sulzer, 1964). Research directed at the primary components of blame (i.e., intent, causation, foresight and foreseeability, mitigating circumstances)

generally confirmed rational expectations, including the findings (a) that people are blamed most for intentional harm, followed by unintended but negligent harm, and least by accidental harm (Karlovac & Darley, 1988; Shultz & Wright, 1985; Shultz, Wright, & Schleifer, 1986); (b) that people who cause harm directly are blamed more than those whose influence is more remote (Johnson & Drobny, 1987); and (c) that people are blamed more for harms they foresaw or should have foreseen than for detrimental outcomes that were unforeseen or unforeseeable (Brewer, 1977; Schlenker, Britt, Pennington, Murphy, & Doherty, 1994).

In the same way that the scientist metaphor of Epoch 1 represents an idealized scientist who adheres to something like Mill's (1843) canons of logic, the lawyer metaphor of Epoch 2 uses formal legal principles to delineate the ordinary person's blame criteria. Furthermore, just as people have been shown to be imperfect scientists, their jurisprudential credentials are not without flaws, as evidenced in phenomena such as victim-blaming (Lerner, 1980; Lerner & Miller, 1978) and the tendency to conflate an actor's "badness" with estimations of his or her intentions and causal involvement (discussed further in Epoch 4).

Next, we consider the aptness of the lawyer metaphor in light of the particular motivations and orientations that differentiate legal and lay judgments. In particular, we examine differences between legal and lay assessments with regard to the types of offenses that they address as well as differences in the evidence that is marshaled to ascribe legal responsibility versus everyday blame. We then look at the mental and behavioral criteria for blame and responsibility and consider some of the practical difficulties they entail for establishing causal responsibility. Finally, we briefly compare the meaning of discounting in Epoch 2 with that in Epoch 1, specifically in the guise of Epoch 2's assessment of excuses and justifications that potentially mitigate or negate responsibility.

Type of offense and scope of evidence

Because the analyses of attribution processes in Epoch 1 terminated with a dispositional or situational inference, the relevance of dispositions—typical emotional reactions, attitudes and beliefs, abilities, and behavioral tendencies—for everyday blame and responsibility judgments was rarely acknowledged. In addition to dispositional inferences, ordinary person perception makes use of nonverbal cues (DePaulo & Friedman, 1998), assumptions about the actor's thoughts and goals (Krauss & Fussell, 1991), and beliefs about what the actor is likely to do in the future (Pronin, Olivola, & Kennedy, 2008). In stark contrast to legal determinations, ordinary blame judgments can be based solely on inferences of this kind,

such as when a partner in a relationship blames the other for what he assumes are negative thoughts that she has about him, for not having enough positive thoughts about him, or for not thinking about him at all. The legal apparatus, of course, is engaged only by external conduct. Although inferences about mental states such as intentions and motives are germane to guilt, liability, and sanction, they do not suffice to bring charges or to award damages unless attended by harmful or potentially harmful actions.

Regardless of whether an injury is accessible to legal redress, the dispositional attributions that were emphasized in Epoch 1 are ordinarily crucial for deciding how to respond to an insult or injury (Bayles, 1982; Goodwin, Piazza, & Rozin, 2014; Nadler & McDonnell, 2012; Pizarro & Tannenbaum, 2011). Furthermore, it is precisely because of the presumed influence of dispositional inferences, and the assumption that jurors will use such information to convict, that information about a defendant's past history or bad character is strictly regulated in criminal trials (Saks & Spellman, 2015, Chapter 5). This practice is supported by data demonstrating a relationship between conviction rates and whether a jury learned of a defendant's previous crimes (Eisenberg & Hans, 2009) and by studies showing that conviction rates are greater when the previous crimes are related to the present one (Wissler & Saks, 1985). More directly, experimental evidence suggests that blame and perceived causation are increased by general information about the defendant's bad moral character and that these relationships are mediated by emotional reactions to the character information (Nadler, 2012; Uhlmann, Pizarro, & Diermeier, 2015). Finally, studies have shown that blame and causal judgments are influenced by global dispositional information (such as likableness) in addition to dispositional information that is related to a specific offense (e.g., an aggressive personality matched to an aggressive crime; Alicke & Yurak, 1995; Alicke & Zell, 2009).

In addition to the legal system's restriction on character evidence, legal jurisdiction applies almost exclusively to offenses to person, property, and reputation: You can, it has been said, sue your neighbor for breaking your jaw but not for breaking your heart. In fact, most emotional (e.g., break-up of a relationship, loss of self-esteem) and material (e.g., loss of a job, failure to be accepted into a school) injuries that people cause each other lie completely outside of legal jurisdiction. Legal enquiries into causation, therefore, are relatively unambiguous in the sense that there is a material outcome to be explained, such as physical injury or property loss. Social outcomes, such as emotional distress or loss of self-esteem, can be more nebulous: It would be difficult, for example, to specify the appropriate analysis for ascertaining whether a parent caused her child's low self-esteem; whether one

spouse was the cause of the other's loss of ambition; or whether one friend's constant criticism of another's singing, dancing, or writing caused her to forego a career in these fields. Furthermore, even if such an analysis was viable, establishing the validity of the relevant facts would pose an often insurmountable challenge.

Mental force

Heider's (1958) "levels-of-responsibility" formulation contains ambiguities that reduce its effectiveness as a model for blame attribution. Although it is certainly true that causing harm is worse than merely being associated with it, the later stages in Heider's formulation represent different criteria for blame rather than stepwise progressions of increasing blameworthiness. For example, whether an actor foresees harm is independent of whether he or she causes it. Intention is also a separate criterion: People can bring about intended outcomes in unforeseen ways, or they can achieve outcomes that were foreseen but unintended. Finally, whereas Heider anticipated later developments in the field by recognizing that observers conflate intention with causation, thereby viewing intentional actions as more causal, he provided very little analysis of the concept of intention or, for that matter, of any of the mental criteria for blame.

The American Law Institute's (1962) Model Penal Code provides a useful way to conceptualize the varying degrees of mental force that underlie different offenses.

The American Law Institute distinguishes among four classes of mental force in decreasing order of criminal seriousness: to act purposely (with practical certainty of the outcome and desire for it), knowingly (with practical certainty of the outcome but lacking desire for it), recklessly (with some probabilistic estimate of the outcome and lacking desire), or negligently (with neither foresight nor desire for the outcome but in circumstances in which the outcome would have been foreseen by most reasonable people). Many studies have demonstrated that people who act purposely are blamed more severely than those whose actions are accidental or characterized by a lesser state of mental force by both adults (e.g., Darley & Shultz, 1990; Gray & Wegner, 2008; Lagnado & Channon, 2008) and children (Darley, Klosson, & Zanna, 1978). More recent research has confirmed the Model Penal Code's levels of wrongdoing (as evidenced in participants' rankings of the degree of punishment that each level of wrongdoing warrants), with the exception that participants generally failed to distinguish between the levels of acting knowingly and recklessly (Shen, Hoffman, Jones, Greene, & Marois, 2011).

As we discuss further in Epoch 4, causal estimations are often conflated with the degree of badness that these different mental states reflect. With all things being equal,

therefore, actors who behave purposively should be seen as more causal than those who behave knowingly, who, in turn, should be seen as more causal than reckless actors, and so forth. Studies on reduced degrees of mental force, such as negligence and recklessness, have been less numerous than those involving perceptions of intent. In one germane study, Lagnado and Channon (2008) explored attributions for negligent versus reckless actions in an experimental design in which they varied both the subjective and objective foreseeability that an action would lead to an unintended bad outcome. In one scenario, for example, Lucy assembles a flat-pack chair and either has a high or low degree of belief that the chair will hold-up in circumstances in which the objective probability of the chair holding-up is either high or low. Her sister sits on the chair, which collapses, causing her injury. Both subjective and objective foreseeability increased blame attributions: People rated the actor as more blameworthy when she expected the adverse outcome (i.e., she acted recklessly) and when the outcome was objectively more likely (i.e., she acted negligently; Lagnado & Channon, 2008).

Behavioral force and causal chains

The difference between an act and its outcomes is rarely distinguished in classic attributional analyses. For example, observers might learn that Ralph tripped over Joan's feet while dancing, subsequent to which they presumably review Ralph and Joan's dancing histories to decide which one is clumsy. In blame and responsibility analyses, by contrast, behavior and its outcomes are usually unpacked, allowing for the fact that people can be blamed for both their actions and for the consequences of those actions. A stark legal example is changing an assault charge to murder when the victim dies 6 months after the assault. A nonlegal example is tracing an adult's obnoxious behavior to the childrearing practices of his or her parents.

Several interesting attributional problems arise when a person's culpable actions represent one element in a complex chain of events leading to a harmful outcome (see Wolff, Barbey, & Hausknecht, 2010). In the following section, we consider the actor's perceived personal force when he or she conspires with another to commit an offense or when he or she simply encourages another to do so; we also consider the various excuses and justifications that mitigate culpability.

Remote and vicarious causal influence

People can be held responsible for the actions of others even when they occupy remote positions in the causal chain by which others committed their offenses. According

to vicarious liability laws, for example, the owner of a bar could face criminal sanctions if a bartender he or she employed sold alcoholic beverages to an underage client. Imposing criminal penalties on a blameless individual (assuming the owner took every reasonable precaution to prevent minors from acquiring liquor in his or her establishment) is controversial in the law (Prosser, Wade, & Schwartz, 1982). As will become clear in Epoch 4, an important question in this regard is whether lay observers' endorsement of such sanctions is based on their perceptions of the owner's causal role in the outcome or simply is due to a difference in values, such as if observers harbored unfavorable views of alcohol or bar owners.

Fascinating examples of second-hand or divided responsibility can be found in the legal categories of *complicity* and *conspiracy*. Complicity occurs when, for example, Marjorie actively or knowingly encourages or enables Hannah to commit a crime, whereas conspiracies involve two or more parties working in concert to implement a criminal plan. In the case of complicity, observers are likely to ascribe the most personal force to Marjorie if she forcefully exhorts Hannah or if she stands to gain from Hannah's actions (Young & Phillips, 2011). In contrast to complicity, a striking feature of criminal conspiracy is that it is one of the few areas in the law in which planning an offense is a crime independent of whether the conspirators actuate their plans. When the conspirators do act, the observer must identify the causal roles of each party. In such circumstances, in which there are few guidelines for apportioning causal influence, perceptions of the badness of each conspirator's motives and actions tend to be conflated with personal force assessments. This may help to explain why, when nonreckless "good" and reckless "bad" actors both causally contribute to an accident, blame assigned to the good actor is discounted in light of the bad actor's role but not vice versa (Mandel, 2010; see also Kominsky, Phillips, Gerstenberg, Lagnado, & Knobe, 2015).

Disruptions in personal force: Mitigation of blame

Even when actors are the predominant cause of harmful outcomes, blame and personal force attributions can be mitigated by excuses and justifications (Markman & Tetlock, 2000; Shaver, 1985; Weiner, Amirkhan, Folkes, & Verette, 1987; for a review, see Snyder & Higgins, 1988). In the law, excuses acknowledge wrongdoing but claim that the actor's personal control was diminished, whereas justifications assert that the behavior enacted was the best option available among poor alternatives—a "choice of evils" (Greenawalt, 1986).

Akin to the internal-external partition of attribution theory, excuses involve either diminution in capacity or

situational pressures (Alicke, 2000; Weiner et al., 1987). Capacity-based excuses include cognitive limitations, physical defects, or emotional disorders, whereas situational excuses involve threats, provocations, and compulsions. The extent to which excused transgressions are forgiven depends on several factors, such as the perceived match between excuse and transgression (P. H. Kim, Dirks, Cooper, & Ferrin, 2006), the sincerity with which the excuse was offered (Zechmeister, Garcia, Romero, & Vas, 2004), how severe the outcome was (Bennett & Earwaker, 1994), as well as the time at which the excuse was offered (Frantz & Bennigson, 2005). Even if an excuse was successful in distancing the actor from what happened, the excuse-giver runs a risk of experiencing disadvantages in the long run, such as being seen as a generally unreliable person (Schlenker, Pontari, & Christopher, 2001).

Blame is mitigated when observers conclude that most people would have behaved in the same manner as the actor under similar situational pressures (Fincham & Jaspars, 1983). In Epoch 2, therefore, discounting involves mitigating the actor's causal force on the outcome in a way that reduces blame, whereas in Epoch 1, discounting usually means mitigating a dispositional inference in light of situational factors. In Epoch 1, for example, the fact that everyone would be angered by the insult that Juan hurls at Martin mitigates the attribution of an angry personality to Martin. In Epoch 2, the fact that few people would haul off and hit Juan on the basis of this insult leads to a failure to mitigate Martin's blameworthiness for hitting Juan.

Legal guidelines for excuses and justifications include cognitive criteria, such as whether a defendant knew what he or she was doing or was able to distinguish right from wrong, and emotional stipulations, such as whether the defendant was able to control his or her emotions (Greenawalt, 1986). Everyday excuses and justifications permit a much broader array of mitigating circumstances than the law permits. For example, some observers might endorse a "tit-for-tat" rule whereby ill treatment by one party justifies retaliation by the other. Justifications of this sort are rarely effective in the law.

Ordinary excuses are also broader because of the obvious fact that they apply to harmful and offensive social behaviors and gaffes that have no legal currency, such as mistaking one person for another, forgetting an acquaintance's name, or not keeping a promise (P. H. Kim et al., 2006). Common excuses for harmful or offensive social behaviors include claims such as "It was an accident," "I didn't know it was wrong," "I just wasn't thinking," "I couldn't help it," "I was forced to," and "I didn't know this would happen" (Markman & Tetlock, 2000). Each of these everyday claims has a more circumscribed legal counterpart. For example, the legal analogue to the "I couldn't

help it” excuse corresponds to claims of diminished capacity, provocation, and some formulations of the insanity defense (i.e., the “control” test). In everyday social life, such claims may be proffered for mental or physical abuse, deleterious personal habits (e.g., alcohol and drug abuse, sex addiction), emotional problems, and greedy and selfish behavior. Similarly, the “I was forced to” excuse corresponds in the law to compulsion, duress, and superior orders (Young & Phillips, 2011). In ordinary social life, such excuses are common refrains of younger siblings and individuals in low-status positions.

Summary of Epoch 2

The shift from the scientist metaphor in Epoch 1 to the lawyer metaphor in Epoch 2 represents a change in orientation from causal analysis as a way of establishing stable properties of the actor or the environment to causal ascription as an essential step on the path to assigning blame. Viewing the layperson’s attributional task through the lens of jurisprudence generally provides a much richer array of causal concepts than the scientist analogy invoked. In addition to highlighting the importance of identifying particular events as causes—those that explain another particular event rather than a class of events—the legal perspective raises issues such as the nearness in space and time of the cause in the causal chain that produces an outcome (i.e., causal proximity), causal overdetermination (i.e., multiple sufficient causes of an outcome), and shared agent causation in cases of conspiracy and complicity. Although character evidence is much more restricted in the law than in ordinary social judgment, dispositional inferences—which are the terminus of attributional investigation in Epoch 1—become an important datum for assessing intention, motive, and (ultimately) blame in Epoch 2. Finally, whereas discounting in Epoch 1 generally refers to discounting dispositional causes of behavior in light of constraining situational forces, the nature of constraints—both on the capacity to enact certain behaviors and on the ability to surmount situational obstacles—becomes a much broader issue in Epoch 2. Constraints can occur at many points along the causal chain that produces, or that fails to produce, a harmful outcome, and the nature of the constraint is a vital factor in assessing the actor’s blameworthiness for the outcome that occurred or that was averted.

Epoch 3: The Counterfactual Model

As noted previously, advances in the heuristics and biases research program that Daniel Kahneman and Amos Tversky initiated began to influence attribution research in the late 1970s (for overviews, see Gilovich, Griffin, &

Kahneman, 2002; Kahneman et al., 1982; Nisbett & Ross, 1980). This program provided a formidable challenge to the person-as-scientist metaphor and, more generally, to rational models of human judgment. One effect of that challenge was to chip away interest in classical attribution models that were premised on rational-assessor assumptions, such as Kelley’s ANOVA model.

A seminal work in that regard was Kahneman and Tversky’s (1982) brief chapter on the simulation heuristic. They proposed that as much as people use the mental ease with which instances are retrieved from memory as a predominant cue or heuristic for judging frequency, people also use the ease with which mental constructions come to mind as heuristics for judging, among other things, causality and blame. People, they hypothesized, are prone to making “downhill climbs” that restore an unusual episode to the normal default. People seldom altered or “undid” aspects of episodes in the direction of making them less normal, which would represent an energy-intensive “uphill climb.” For example, a driver would have an easier time mentally “undoing” a car accident after he took an unusual route home (by imagining having taken his usual route) than the other way around. When thinking about how events could have unfolded differently, people are therefore more likely to construct simulations that restore episodes to their normal, default value by mentally undoing abnormal antecedents of the episode’s focal outcome. These notions were expanded in Kahneman and Miller’s (1986) norm theory, which spurred the development of a new research area in social psychology on counterfactual thinking and its functional consequences (for overviews, see Epstein & Roese, 2008; Mandel et al., 2005; Roese, 1997).

An important idea in norm theory is that psychological reactions to events are automatically constructed after an event has concluded. These postevent computations were posited to affect attributional inquiry by shaping what needs to be explained. What people most need to explain, according to norm theory, are those outcomes that evoke surprise by violating people’s expectancies, especially when those violations take the form of aversive, unwelcome results (Taylor, 1991). In this sense, norm theory was built on earlier research in which researchers had likewise concluded that unexpected outcomes were more likely than expected ones to trigger explanatory searches for the cause or causes of the outcome (Hastie, 1984; Pyszczynski & Greenberg, 1981; Wong & Weiner, 1981). However, unlike the earlier studies making this point, Kahneman and colleagues forged this idea with generative proposals regarding how people can use mental simulation and counterfactual thinking to arrive at causal explanations. The effect of this new paradigm on attribution research was profound because it shifted research attention away from the “disposition-versus-situation” question to a more

general focus on explaining unexpected or unwanted outcomes (for examples of studies in which researchers attempted to distinguish the effects of expectancy violation and negative outcomes on causal search, see also Bohner, Bless, Schwarz, & Strack, 1988; Kanazawa, 1992). The constructive effect of this paradigm shift was to stimulate interest in new attributional questions and, in some sense, to rejuvenate this research area.

Psychological closeness and abnormality

Norm theory spurred research efforts directed at discovering the rules that govern surprise and at the consequences of considering alternatives to the surprising events or outcomes. Kahneman and Varey (1990) elaborated that what people find most surprising is not necessarily what they thought had the lowest subjective probability of occurring at the start of an episode. For example, one might be surprised by the underdog that loses in a sports match if that player had the momentum or propensity to win through much of the game. Surprise or expectancy violation, therefore, has much to do with one's perception of the dynamical properties of the episode itself, which may change observers' outcome expectancies over short intervals.

In particular, the concept of psychological closeness played an important role in the development of norm theory. The perceived closeness of a surprising outcome to the expected one was presumed to be an important determinant of the ease with which one would mentally undo an outcome. Thus, for example, Kate—a commuter who misses a bus by 10 s after having to take a minute to locate her keys—is more likely to ruminate about how she might have made it in time than another commuter, Tonya, who did exactly the same things beforehand but missed the bus by 5 min. Kate, who almost caught the bus, is more likely to psychologically “kick” herself and assign self-blame for the outcome. She is likely to move from thinking about what might have been to feeling that she ought to have caught the bus (D. T. Miller & Turnbull, 1990). Tonya, who missed the bus by several minutes, however, is unlikely to feel as if she almost caught the bus and, accordingly, is less likely than Kate to blame herself.

The abnormality of events leading up to a surprising or negative outcome was also posited to influence blame assignment. For instance, as noted earlier, the driver who had a car accident on his way home from work by an unusual route that he whimsically decided to take would be expected to blame himself more than a comparable driver who simply took his usual route (Kahneman & Tversky, 1982). It is easier to mentally undo the abnormal action preceding the accident than it is to undo the usual

action. The implication of this is that blame ascriptions will be influenced by factors that had no indicative value in foresight and thus should not, by normative accounts, factor into the assessment. Of course, the effect of counterfactual thinking on blame assignment is not a causal necessity. If outcome mutability (i.e., the ease with which one can imagine the outcome of a given case having turned out differently) is low or nonexistent or an actor has taken reasonable precautionary measures to avoid a negative outcome, antecedent mutability (i.e., the ease with which one can imagine an antecedent of the outcome of the case having been different) may not increase blame assignment (Alicke, Davis, Buckingham, & Zell, 2008). For instance, imagine a soldier who had carefully followed all military procedures for ensuring his or her safety but who was nevertheless killed in the line of duty. It might be difficult to mentally undo the soldier's death because the profession of arms requires putting oneself in harm's way when necessary. Moreover, the studies demonstrating such abnormality effects had methodological limitations. For instance, the length of description given to abnormal events is typically longer than that given to normal events in vignette studies, and when such confounds are controlled, abnormality effects have been eliminated (Trabasso & Bartalone, 2003).

The counterfactual cues to causality controversy

The findings of the 1980s and early 1990s suggested that blame and victim compensation judgments were affected by antecedent mutability (e.g., MacRae, 1992; D. T. Miller & McFarland, 1986). This led some researchers to posit that the effect of counterfactual, “if only” thinking—namely, thinking about how things might have turned out differently—on such judgments was mediated by causal inferences. This idea was presaged in Kahneman and Tversky's (1982) chapter, in which they proposed that people mentally simulate counterfactuals to probe the causal structure of episodes in need of explanation. For instance, one might do this either by negating an antecedent (e.g., undoing having taken an unusual route home) and observing whether it would undo the outcome in the simulation (the car accident that happened on the way home) or by examining whether a change in antecedent conditions (e.g., increasing security presence) would correspondingly change the propensity toward a particular outcome or even make the outcome seem inevitable (reducing the likelihood of terrorist attacks at public events such as the Boston Marathon). In fact, a number of investigators subsequently proposed that counterfactual conditionals of the form “If only *X* hadn't occurred, then *Y* wouldn't have either” were necessarily causal. In this formulation, *X* usually stood for an

abnormal antecedent condition, and *Y* usually stood for a negative or surprising outcome. For instance, Wells and Gavanski (1989) proposed, “an event will be judged as causal of an outcome to the extent that mutations to that event would undo the outcome” (p. 161). Likewise, Roese and Olson (1995) argued, “counterfactuals, by virtue of the falsity of their antecedents [which contrast with the actual antecedents], represent one class of conditional propositions that are always causal” (p. 11). Yet, it is easy to show that counterfactual conditionals can represent noncausal statements—for example, “if my name were John rather than David it would only have four letters” (for further discussion, see J. Kim, 1973; Mandel, 2011).

It is also important to distinguish the deliberate construction of counterfactuals from counterfactuals that automatically spring to mind. For example, imagine a woman who lost her husband in one of the planes that was crashed into the Twin Towers on 9/11. After the event, she might have the following spontaneous ruminative thoughts: “If only my husband had taken the later plane,” or “If only I had forced him to stay for our son’s birthday,” thus mentally undoing factors that, in reality, she could have herself undone to prevent the traumatic event. These events all have a causal connection to the husband’s death (which if undone would have prevented it in a mental simulation) and thus would be considered causal of the outcome according to the definitions given earlier (Roese & Olson, 1995; Wells & Gavanski, 1989). However, intuition indicates that it is unlikely that people would see either of these actions as the cause of his death. In the case of 9/11, intuition indicates that most people (including the victims’ families) would see the actions of Al-Qaeda in hijacking the aircraft as the cause of the passengers’ deaths; however, the traumatized wife, when ruminating about the events that led to the loss of her husband, will dwell on other factors (e.g., “If only I had persuaded him to stay at home for our son’s birthday”).

By the mid-1990s, researchers began finding that the antecedents people identified in their “if only” statements also did not necessarily coincide with what they would choose to ascribe as the cause of the focal (usually negative) outcome (e.g., Davis, Lehman, Wortman, Silver, & Thompson, 1995; McEleney & Byrne, 2006; N’gbala & Branscombe, 1995). In these studies and others (e.g., Giroto, Legrenzi, & Rizzo, 1991), it was evident that people were more likely to focus on personally controllable antecedents in their counterfactual “if only” thoughts than in their causal attributions. Such research led to a proposed functional dissociation between causal and counterfactual ascriptions (Mandel & Lehman, 1996). Specifically, it was proposed that causal ascriptions focus mainly on identifying antecedents of an outcome that would yield predictive information in future cases,

whereas counterfactual ascriptions focus mainly on identifying how an actor could have controlled events so that the actual (negative) outcome might have been prevented. This divergence is arguably adaptive because both prediction and control are important functional requirements, yet they do not necessarily converge on the same antecedents (Mandel & Lehman, 1996).

This kind of divergence between causes and counterfactuals was observed in studies in which researchers examined real-life cases of trauma and loss. For instance, although a majority of parents in one study who had lost a child to sudden infant death syndrome reported that they did not feel they had caused their child’s death, most also reported thinking about how their child might still be alive if only they had done something differently, such as checking on the child more frequently (Davis et al., 1995). Moreover, the frequency of such counterfactuals predicted the degree to which they felt responsible and guilty for their child’s death. In a subsequent study of respondents with spinal-cord injuries, Davis, Lehman, Silver, Wortman, and Ellard (1996) found that, compared with independent raters, involved respondents assigned the same level of causal importance to their actions but assigned much more self-blame. Moreover, when Davis et al. examined only those responses that implicated the respondents’ personal actions, they found that counterfactuals predicted self-blame controlling for causal assessments. However, causal assessments did not predict self-blame controlling for counterfactuals. Using an experimental approach, Mandel and Dhami (2005) asked sentenced prisoners to think either factually (e.g., “what you actually did”) or counterfactually (e.g., “what you might have done differently”) about the role they played in events leading to their capture, conviction, and sentencing. Compared with the factually focused prisoners, counterfactually focused prisoners blamed themselves more and felt guiltier, and the relation between the thought focus manipulation and guilt was fully mediated by self-blame (Mandel & Dhami, 2005). In a related study (Dhami, Mandel, & Souza, 2005), prisoners who reported positing “upward” counterfactuals about how their trial or sentencing might have turned out better also felt their trial and sentence were less fair and reported feeling angrier than prisoners who had not reported such thoughts. Although these studies showed that counterfactual undoing thoughts had a wide range of attributional and affective consequences, many of them did not seem to be mediated by judgment of what “the” cause or main causes of the focal outcome happened to be.

Refinements to the functional account

The emerging functional account in Epoch 3 accorded well with covariation theories of causal inference (e.g.,

Cheng & Novick, 1990, 1992) because it posited that causal ascriptions were mainly focused on yielding predictive knowledge. The crediting causality model (Spellman, 1997; Spellman, Kincannon, & Stose, 2005) bridged the gap among these literatures. Whereas most covariation theorists had focused on how causal reasoners induce causal relations among types of events over a large set of cases, Spellman and colleagues adapted that approach to the problem of causal selection in the specific case, proposing that observers work backward from the known outcome, first asking themselves what the prior probability of that outcome was at the start of the case, and then updating that probability conditional on each key event in the temporal sequence culminating in the outcome. For example, consider a game in which two team players sequentially toss fair coins for a prize that they will win if and only if their coins land on the same side. If Player 1 gets Heads and Player 2 gets Tails, either player's outcome could have been changed to make the joint outcome a winning combination. The crediting causality model predicts that Player 2 will be seen as more causal because the probability of winning at the start is $\frac{1}{2}$, and it remains $\frac{1}{2}$ after Player 1 tosses the coin. However, after Player 2 goes, the probability of winning goes to zero. Thus, Player 2 makes all the difference in changes to the probability of the outcome in this case. Although Spellman (1997), who used this problem, did not ask participants about their causal inferences, she found that the majority of participants tended to think that Player 1 will blame Player 2 more than the other way around (see also D. T. Miller & Gunasegaram, 1990, who developed the coin-toss problem).

However, the crediting causality model cannot account for observers' causal ascriptions in causal preemption cases. For instance, in one experiment (Mandel, 2003, Experiment 1), participants read a story in which a man puts a lethal dose of poison into the drink of a rival organized crime boss, Mr. Wallace, only for the latter to be assassinated by another rival who runs Wallace's car off the road before the poison has had time to act. Although the crediting causality model predicts that the first assassin would be deemed the cause (because he raises the probability of Mr. Wallace's death to 1), most observers select the second assassin as the cause but undo Mr. Wallace's death by undoing both assassination attempts or by undoing Mr. Wallace's (presumably controllable) involvement in a life of organized crime (Jimenez-Leal, 2013; Mandel, 2003; see also Walsh & Sloman, 2011).

Using judgment dissociation theory, Mandel (2003, 2005) accounted for such findings by proposing that observers are influenced by multiple functional goals when formulating explanations of such events. Their counterfactual explanations focus on what would have been sufficient in the circumstances to prevent the

outcome as it actually came about as well as other functionally equivalent outcomes that might otherwise have materialized. Thus, undoing both assassins' actions and Mr. Wallace's life in crime successfully would have prevented not only his actual death by car crash but also his inevitable death by poisoning should the crash have failed to kill him. Observers' causal explanations, however, tend to focus on how the specific outcome came about. Thus, although the first assassin's poison raised the probability of Mr. Wallace's death to certainty, leaving no room for the second assassin to further increase the probability of that outcome, most observers nevertheless view the second assassination attempt as causal and correspondingly blame the second assassin.

Indeed, examples such as this flag a deeper problem for any account of causation that relies purely on dependency relations (probabilistic or counterfactual) and suggest that people are most focused on the actual causal processes leading to the effects in question. Simple as this concept of actual cause seems, it has so far eluded satisfactory formal analysis and is the subject of considerable debate in the philosophy of causation (see the collection in Collins, Hall, & Paul, 2004). Nevertheless, a broad range of psychological research shows that processes, mechanisms, and narratives play a key role in causal reasoning (see Sloman & Lagnado, 2015, for a review), and simple, counterfactual, or probabilistic accounts struggle to capture these riches.

Counterfactuals and responsibility attributions to individuals in groups

The relationship among causal attribution, counterfactual judgments, and controllability discussed in the previous section is also applicable to situations in which multiple individuals jointly cause a given effect. As social animals, humans regularly coordinate their efforts to effectuate outcomes that they would be incapable of producing individually. Humans engage in team sports, vote in elections, participate in demonstrations, and pay their taxes to contribute to public goods such as education or national security. When multiple individual contributions combine to determine a joint outcome, the extent to which a given individual controls the outcome is usually diminished (Bandura, 1977; Kerr, 1989). Indeed, such situations invite an addition to the excuses offered in Epoch 2. Despite acknowledging the problem of global warming, an individual might justify his or her transatlantic holiday flight by rationalizing that his or her action does not really make any difference—the plane will fly with or without one more passenger (Kerr & Kaufman-Gilliland, 1997). Research has shown that people are sensitive to whether their contribution is likely to make a difference to a joint outcome and adjust their physical

effort (Kerr & Bruun, 1983) or financial contributions (De Cremer & van Dijk, 2002) accordingly. Recently, it has also been argued that the decreased sense to which one's action is perceived as making a difference explains why people behave more immorally in competitive markets than they would do individually (Falk & Szech, 2013).

From a theoretical perspective, group contexts provide a challenging test ground for assessing responsibility judgments. In most group contexts, joint outcomes are causally overdetermined: The same candidate would have won the election even if one of the voters had changed his or her mind, and the same team might have won the game even if one of their players had been injured. Nevertheless, we feel that each individual in the group has some degree of responsibility for the outcome. Hence, a simple test of whether the outcome would have been different but for a person's action fails as a model of responsibility attribution, which ostensibly provides additional evidence that counterfactual dependence is at best an imperfect guide to attributions.

Chockler and Halpern (2004), however, have developed an alternative account of responsibility attribution, which predicts that a person's responsibility reduces with the distance between the *actual situation* in which his or her action did not make a difference and a *possible situation* in which his or her action would have made a difference to the outcome. The distance between these two situations is based on the minimal number of changes that would be required to turn one situation into the other, given a causal representation of the situation (cf. Halpern & Pearl, 2005; Pearl, 2000). For example, Tom is predicted to feel very responsible when his vote alone could have changed the outcome of an election. In contrast, if the actual situation was such that several other voters would have had to change their minds to create the possible situation in which Tom's vote would have been pivotal, Tom is predicted to feel less responsible.

Several studies support the predictions of this counterfactual account of responsibility attribution (Gerstenberg & Lagnado, 2010, 2012; Lagnado, Gerstenberg, & Zultan, 2013; Zultan, Gerstenberg, & Lagnado, 2012). This research confirmed that responsibility attributions to individuals in groups are affected by both *prospective* and *retrospective* factors (cf. Hamilton, 1980; Schlenker et al., 1994). Prospectively, responsibility attributions increase the more critical a person's action is perceived for bringing about a positive group outcome. Retrospectively, attributions increase the closer a contribution was to having made a difference to the outcome.

Summary of Epoch 3

The emphasis on abnormal causes that was introduced in Epoch 2 continued into Epoch 3 but with a twist: In

Epoch 3, the abnormality of a causal candidate is determined after the fact by considering alternatives to surprising, and especially harmful, actions and outcomes. Early research in Epoch 3 was directed at investigating the factors—psychological closeness and abnormality in particular—that render an event surprising and promote the search for alternatives that would mutate the unusual event into its more normal state. Later, research focused more on the dissociations between diverging functions of causal and counterfactual explanations (Mandel, 2011). Although in early studies on counterfactual reasoning researchers tended to assume that these mental undoings reflected causal inferences of the form “if Joe hadn't done *X* (where *X* is typically an unusual action), then *Y* (where *Y* is typically a surprising outcome) wouldn't have occurred,” subsequent theory and research suggest that mentally transforming unusual events does not necessarily implicate *X* as the cause of *Y* or even as what one might spontaneously label as a cause of the outcome. People frequently engage in counterfactual thinking as a way of imagining how they might have personally controlled events in ways that would have been sufficient to prevent the harmful outcome, even if they do not believe that they were the principal cause of the harm.

More recent studies have shown further that both the extent to which a person's contribution is perceived to be critical for an outcome and the judged closeness to a situation in which things could have turned out differently are grounded in a causal representation of the situation. The causal representation of the situation provides the basis for mentally simulating what will happen or what would have happened if things had turned out differently (Goodman, Tenenbaum, & Gerstenberg, 2015) and illustrates the usefulness of explicating the role of such causal representations in understanding judgments of credit and blame in coordinated group activities. This approach promises to offer new insights into how people's moral evaluation of others is affected by causal and counterfactual reasoning.

Epoch 4: Personal Force and Moral Psychology

The metaphors governing the first three epochs—beginning with quasi-scientific covariation analyses, to a legalistic investigation of an actor's blameworthiness and responsibility for causing harm, and then to a reconstruction of events that seeks to explain how harm could have been avoided—suggest different epistemic orientations toward the attribution task. In Epoch 1, the orientation is to compare an actor's behavioral tendencies with that of others in similar situations to decide whether an observed behavior reflects a dispositional push or a situational pull. In Epoch 2, the orientation is to assess whether an actor

willfully, negligently, or recklessly caused a harmful outcome in the absence of an acceptable excuse or justification. Finally, in Epoch 3, the orientation is one of imagining how events, especially those with undesirable consequences, could have turned out differently. None of these orientations place any special emphasis on the emotions, psychological needs, or attitudes of the lay attributor. In Epoch 4, the analysis shifts to accommodate the ways in which causal and blame attributions are influenced by observers' moral, evaluative, and emotional reactions to the event and the people involved.

We noted that Heider's (1958) conception of personal causation, or what he called personal force, was unique in that it included intention as a requirement for personal causation. Legal and philosophical views of causation distinguish sharply between acting purposively and knowingly on the one hand, and actually bringing about an outcome on the other hand. These distinctions are required to account for the fact that people can intend outcomes that they fail to effectuate and cause outcomes accidentally or inadvertently. Nevertheless, Heider was prescient in recognizing that ordinary conceptions of "making something happen" by exerting personal force are based partly on aspects of events that are excluded from standard causal notions.

Evaluative tone

One of the most important "extraneous" influences on causal perception is the positivity or negativity—or evaluative tone—of the agent's motives, character, or the event's outcomes. In an early demonstration (Alicke, 1992), participants read the story of a driver, John, who had a car accident while he was on his way home to (a) hide an anniversary present from his parents or (b) hide a vial of cocaine from his parents. At an intersection with a stop sign, he encountered one of three environmental obstacles: (a) a tree branch that partly obscured the sign, (b) an oil slick in the road, or (c) another car that ran through the stop sign. John hit the other car on the driver's side, causing him serious injuries. Participants were asked to state what they thought was the primary cause of the accident, and their attributions were divided into whether they listed something about the driver or the environmental obstacle as the primary cause. Although causal citations varied with each environmental obstacle (e.g., the other driver was cited more frequently when he ran through the stop sign), the frequency with which John was cited as the primary cause in each condition was significantly greater when his motive was to hide a vial of cocaine than to hide an anniversary present.

Analogous findings have been obtained in studies in which the actor's general character (Alicke & Zell, 2009; Nadler, 2012) or the outcomes that occurred (Alicke &

Davis, 1989; Alicke & Rose, 2012; Alicke et al., 2011; Mazzocco & Alicke, 2004) were varied. According to the culpable control model of blame (CCM; Alicke, 1992, 2000, 2008; Alicke et al., 2011), spontaneous positive or negative evaluations of an event evoke a blame-validation mode of processing and judgment. In this mode, observers interpret evidence regarding the actor's control over the event's outcomes in a manner that tends to validate praising an actor who elicits positive evaluations and blaming an actor who arouses negative evaluations.

This issue has been revitalized recently in experimental philosophy. In a well-known study in this literature, Knobe and Fraser (2008) have described a situation in which administrative assistants and faculty members routinely take pens from a receptionist's desk. The administrators are allowed to take the pens, whereas the faculty members are not. As the story develops, an administrator and a faculty member each take a pen from the receptionist's desk, leaving her without a pen when an important message arrives. When participants rated the extent to which the administrator or the faculty member was the cause of the problem, they gave much higher causal ratings to the faculty member.

Hitchcock and Knobe (2009) interpreted these findings to indicate that causal judgments are amplified by norm violations—a view that accords with norm theory and with Hart and Honoré's (1959) emphasis on abnormal conditions. According to CCM, however, norms by themselves cannot explain heightened causal perceptions when actors engage in behavior that is accessible to praise or blame. Instead, positive or negative reactions to the actor, his or her behavior, or the outcomes of those actions are the engine that drives causal attributions.

The fact that counternormative behavior is usually bad behavior requires an experimental design in which normativity and goodness–badness are disentangled. In one study designed to make this distinction (Alicke et al., 2011, Study 2), participants read about a group of students who obtained a copy of their final exam in biology. All of these students either cheated or did not cheat on the test. One student, John Granger, went along with the group (normative behavior) or did not (counternormative behavior). There were, therefore, conditions in which Granger adhered to the norm by cheating (norm, bad), adhered to the norm by not cheating (norm, good), deviated from the norm by cheating (counternormative, bad), and deviated from the norm by not cheating (counternormative, good). Participants read that the class was graded on a curve such that an equal number of students received grades of A, B, C, and D. Granger was the last student in the class to receive an A. The next highest score was obtained by a premed student, who, as a result, missed the grade-point average cutoff she needed for medical school. Participants were asked to indicate the

degree to which Granger was the cause of the student missing the grade-point average cutoff. The norm violation approach predicts that causal attributions will be greater when a norm is violated than when it is not. The CCM predicts that causal attributions for a harmful outcome will be lessened for the praiseworthy behavior of not cheating and increased for the blameworthy behavior of cheating—regardless of whether the behavior was normative or counternormative. Results clearly support the CCM. In fact, the lowest causal ratings by far occurred in the condition in which Granger behaved counternormatively by refusing to cheat on the exam. According to the CCM, this occurs because participants refuse to assign causal responsibility for a negative outcome to a person who behaves heroically by asserting his or her principles (see also Turri & Blouw, in press).

Focal and peripheral effects

As we noted in Epoch 2, the causal process leading from intentional actions to their ultimate consequences is often complex and convoluted. When people enact goal-oriented behavior, both focal and peripheral effects can ensue. Focal effects represent an actor's primary objectives, whereas peripheral effects are outcomes that an actor either desires less or not at all but that he or she is willing to tolerate to achieve the focal effect.

An actively researched area in experimental philosophy concerns cases in which a person obtains a focal effect as anticipated but also causes foreseen, undesired peripheral effects. Examples include hoping vainly that someone will not be insulted by a critical comment and cooking a vegetarian dinner for guests that is unpalatable for the attending carnivores. A more serious example is the military euphemism of “collateral damage”—killing innocent civilians as an unintentional consequence of military operations. Research on this topic has focused on an experimental arrangement introduced by Knobe and Fraser (2008) in which the chief executive officer (CEO) of a company wants to initiate a program to increase profits (focal effect) but recognizes in one experimental condition that the program will have harmful consequences on the environment and in the other that it will have helpful effects (peripheral effects). The CEO states that he does not care whether the product helps/harms the environment; he is only concerned with profits. Research consistently shows that people ascribe greater intentionality to the CEO whose actions bring about negative versus positive peripheral consequences.

Various explanations for this “side effect” problem have been suggested (Alicke, 2008; Driver, 2007; Knobe, 2010; Nanay, 2010; Sripada, 2011; Uttich & Lombrozo, 2010), but interpretations are hindered by the looseness of the original paradigm, which fails to control the

numerous differences between the helping and harming conditions. These include the following: the important distinction between a CEO who is indifferent toward helping the environment and one who is indifferent toward harming it, the difference between acting on the certainty of bringing about helpful outcomes as opposed to harmful ones, and the fact that simply effecting positive outcomes leads to more favorable evaluations than effecting negative ones (Alicke & Davis, 1989; Alicke, Davis, & Pezzo, 1994; Alicke et al., 2011; Baron & Hershey, 1988; Mazzocco & Alicke, 2004). To unpack these factors, one would need a prohibitively large experimental design, and even this would not solve the inherent asymmetries in what it means, for example, to not care about helping versus harming the environment. Perhaps an even more problematic aspect of this CEO story is that in a free response format, very few participants characterize the CEO's behavior as intentional regardless of the peripheral outcomes it achieves (Guglielmo & Malle, 2010).

Despite the difficulties inherent in this particular experimental arrangement, Knobe and Fraser's (2008) thought experiment raises the important and underresearched question of how people evaluate the peripheral outcomes of focal objectives. Peripheral outcomes can be desired or undesired, can be brought about in foreseen or unforeseen ways, and can be accompanied by focal effects that also vary in desirability and foresight. These different combinations correspond to meaningful problems in attributions of personal force and blame that have been largely unexplored in previous studies.

Intuition versus deliberation

Whereas the CCM was originally developed to address issues in causal attribution, most research and theories that fall under the increasingly popular but amorphous rubric of moral psychology (Doris, 2010; Nadelhoffer, Nahmias, & Nichols, 2010; Sinnott-Armstrong, 2008a, 2008b, 2008c) address the more general question of whether moral judgments are made intuitively or deliberately (Haidt, 2001; see also Inbar & Pizarro, 2009). The place of causal judgment in moral psychology is as an essential criterion for assessing blame or praise for actions and their consequences. Causal judgments are usually tied to specific, concrete actions, such as deciding whether to blame someone for hurt feelings over an insulting comment or personal betrayal, but they can also be more general and abstract, such as when a child blames his or her parents for his or her low self-esteem. Even in this instance, however, the child implies that the parent did something to cause this unhappy consequence, even if he or she is unable to identify the specific precursors.

The intuitive, emotional, or evaluative element in causal judgment as it relates to praise and blame involves—as with the role of intuition in morality more generally—a conflation of automatic, intuitive, top-down with deliberate, evidence-driven, and bottom-up judgment processes. In particular, in assessing the evidence regarding an actor's causal role in a morally praiseworthy or blameworthy act (deliberate, bottom-up), the observer's attitudinal or emotional reactions (automatic, top-down) to the event and its consequences influence causal judgments.

In current moral psychology and experimental philosophy, the controversy about the extent to which this conflation between attitudinal reactions and evidential criteria occurs has been discussed more frequently with regard to Heider's (1958) other main component of personal force—intention—than it has with causation per se. In experimental philosophy, in particular, research on this topic has centered on the side effect problem discussed previously. Several investigators (e.g., Alicke, 2008; Driver, 2007; Nadelhoffer, 2004; Uhlmann et al., 2015) have suggested that the negative reactions or emotions that occur in response to a CEO who callously pursues profits while expressing complete indifference to environmental harm account for heightened intention judgments—an argument that is consistent with Tetlock's (2002) intuitive theologian metaphor. Conversely, explanations based on the assumption that people engage in more deliberate reasoning suggest that increased intent ascriptions are due to norm violations (Uttich & Lombrozo, 2010), to the fact that moral concepts are inherent in the way people understand intent (Knobe, 2006), and to dispositional inferences about the CEO (Sripada, 2011). As we noted, further developments in understanding how people make attributions for undesired but foreseen peripheral consequences of a focal goal await more adequate and diverse experimental paradigms.

Summary of Epoch 4

Epoch 4 follows rather naturally from the two preceding ones. Because the law is concerned with intentional, negligent, and reckless harm-doing, it is a short step to apply these concerns to moral judgment. The primary difference between the second and fourth epochs is that lawyers operating in the Anglo-American law tradition are bound by procedures and rules of evidence that do not apply to ordinary moral judgments. Similarly, research on counterfactual reasoning in Epoch 3 began with the observation that emotional reactions to harmful events spurred thoughts about how these events could have been avoided, which has implications both for identifying the cause of the event and the blameworthiness of the agent. Emotional reactions are also a large part of

moral judgment; in fact, philosophers have long debated whether moral judgments are anything more than emotional reactions (for a review, see Prinz, 2007). Research in Epoch 4 makes a strong empirical case for the effects of emotions on moral judgments and, more germane to our present concerns, for the effects of emotions and evaluative reactions on causal attributions. Whether these evaluative reactions are native to the process of ascribing cause and blame, or whether they represent occasional deviations from otherwise rational judgment tendencies, is a topic of ongoing debate (Alicke, 2014; Malle, Guglielmo, & Monroe, 2014).

Summary and Conclusions

If there is a primary difference among the attributional models of yesteryear and present day, it is a shift from the problem of inferring stable personal characteristics—attitudes, propensities, and abilities—to a concentration on evaluating social conduct and moral character. Whereas causal analysis in the early attribution theories of Epoch 1 primarily involved establishing stable dispositional or environmental tendencies, an actor's eligibility for responsibility and blame as a result of his or her causal participation in a harmful event became far more central in the ensuing three epochs.

The precise meaning of the causal construct varies across these different epochs. Epoch 1 relied on a particular version of Mill's (1843) methods of agreement and difference—covariation more generally (see Cohen & Nagel, 1934, for a description)—to establish whether an observed behavior covaried over different actors, situations, and occasions. Hart and Honoré's (1959) influential "Causation in the Law" guided Epoch 2 by emphasizing the difference between legal and scientific causal notions—namely, that legal causes typically favor human actions and abnormal conditions that are situation specific rather than universal generalizations. Epoch 3 also featured causal abnormality but in a somewhat different form. Here, imagined alternatives to actions that produce harmful outcomes, or to the outcomes themselves, impel post hoc causal assessments such that if the actor had done something different, the harmful outcome would not have occurred. Later research addressed the different functions served by causal judgment and by ruminative counterfactual thinking.

Epoch 4 completes the cycle from the original scientist metaphor to a depiction of people as moralists who are primarily concerned with assessing the moral goodness or badness of behavior. Judging peers and acquaintances in this manner—being "too judgmental"—is sometimes considered a significant character flaw. However, although it may be sensible to chastise people for judging their fellows precipitously, unfairly or inaccurately, it would be

futile to rebuke them for judging others on moral dimensions. As the extensive history of research on attitudes by social and personality psychologists has shown, judgments of objects on evaluative good–bad dimensions are ubiquitous and inevitable (Bargh & Chartrand, 1999; Fazio, Sanbonmatsu, Powell, & Kardes, 1986; Osgood, Suci, & Tannenbaum, 1957). People are evaluating creatures, and perhaps the most important things they evaluate are each other. The perception that an actor has been instrumental in causing a harmful outcome is one of the main criteria for blame and for unfavorable dispositional attributions (Epoch 2), and conversely, the desire to blame people affects perceptions of their causal participation in bringing about harmful consequences (Epoch 4).

Heider (1958) anticipated the primacy of moral evaluation by incorporating intention into personal causation, or what he called personal force. For Heider, meaningful attributions presuppose that the actor's behavior was self-willed and purposive. Whereas physical objects require an external force to set them in motion, animate objects—whatever one's position on free will and volitional action—are perceived to be able to set themselves in motion. Furthermore, humans, as animate objects par excellence, are seen not only to will their own actions but to do so for various reasons or purposes (Buss, 1978). When these purposes are especially laudatory or nefarious, or produce salutary or harmful outcomes, they arouse the approbation or disapprobation of the observers who view them. The question of causation in human action, therefore, is inextricably bound up with the evaluative meaning these actions have for observers. The tendency to conflate purposes and actions in causal assessment is so natural that philosophers, most notably David Hume and Thomas Reid (for a review of their positions on causation, see Robinson, 2002), have argued that the experience of human agency is the model that people use to understand physical causation. Therefore, whereas a scientist would explain Kepler's laws of planetary motion mechanistically, the ordinary observer, if planetary movement was shown under the right conditions, might say that Planets A and B were settling a score with Planet C. This, of course, is the point at which Heider in his studies with Marianne Simmel (1944) embarked on the study of causal perception.

It is no surprise, therefore, that Heider's (1958) conception of personal force integrated intention with more traditional causal conditions. As we have shown in this article, the multitudinous aspects of causal influence that were studied in the ensuing four epochs have essentially cycled back to a conception that is quite congenial to Heider's original one. In particular, the pervasive finding in studies during Epoch 4 that causal judgments are influenced by positive and negative evaluations of the actor's

behavior and its outcomes is consistent with Heider's recognition that folk conceptions of personal force entail more than scientific notions of causal influence imply.

At the same time, current researchers on blame and related attributions can benefit from reconsidering some of the major topics in earlier attributional formulations. For example, Kelley's variables of consensus, consistency, and distinctiveness are important for deciding in ambiguous cases whether a behavior was intentional or accidental, and if intentional, the immediate and extended reasons for which it was enacted. Furthermore, as we have discussed in Epoch 3, the discounting principle that Kelley discussed and the failure to discount situational pressures that constitute the basis of correspondence bias are quite relevant for considering the factors that mitigate or negate blame. The attributional problem of discounting dispositional inferences in light of situational pressures maps directly onto the problem of blame mitigation in Epoch 2 except that the situational pressures that underlie legal excuses are spelled out in greater detail, and "internal" pressures such as mental or emotional incapacities can also serve to discount intent and blame.

Counterfactual reasoning perspectives, especially the assumption that mental reconstructions of harmful events devolve on controllable antecedents of an outcome, are also, as we have noted, highly relevant for ascribing blame and assessing moral character. A main contribution of the counterfactual approach, in contrast to prior work on abnormal or necessary causes, is that features of the event that would have been unrecognized as causal at the outset accrue causal potency for various reasons, perhaps the most prevalent of which is their capacity to have prevented a harmful outcome. Thus, their perceived causal force is likely to be heightened, making them prime targets for blame.

In the absence of a grand Darwinian or Einsteinian world view, most psychological theories tend to focus on circumscribed aspects of behavior and cognition. This is true of theories of causal perception and personal force that address relatively specific problems, such as extracting dispositional qualities against the background of situational forces, identifying the prepotent cause among various competing causal forces, and assessing the conflation between causal impact and the badness or unusualness of behavior. We have traced the progression of these problems across the four research epochs and have drawn links among their treatments in each. Each of the topics identified and pursued in these research epochs are, of course, still being investigated. Although there are fewer studies in which researchers assess classic attributional models, studies directed at the personal force issues raised in Epochs 2–4 constitute as large a segment of the total research enterprise in psychology as they ever

have previously, and with the rise of moral psychology, there is every reason to believe that research on causal perception, personal force, and behavior evaluation will continue to elucidate these fundamental facets of social perception and judgment.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

References

- Alicke, M. D. (1992). Culpable causation. *Journal of Personality and Social Psychology*, *63*, 368–378.
- Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin*, *126*, 556–574.
- Alicke, M. D. (2008). Blaming badly. *Journal of Cognition and Culture*, *8*, 179–186.
- Alicke, M. D. (2014). Evaluating blame hypotheses. *Psychological Inquiry*, *25*, 187–192.
- Alicke, M. D., & Davis, T. L. (1989). The role of a *posteriori* victim information in judgments of blame and sanction. *Journal of Experimental Social Psychology*, *25*, 362–377.
- Alicke, M. D., Davis, T. L., Buckingham, J. T., & Zell, E. (2008). Culpable control and counterfactual reasoning in the psychology of blame. *Personality and Social Psychology Bulletin*, *34*, 1371–1381.
- Alicke, M. D., Davis, T. L., & Pezzo, M. V. (1994). A *posteriori* adjustment of a *priori* decision criteria. *Social Cognition*, *12*, 281–308.
- Alicke, M. D., & Rose, D. (2012). Culpable control and causal deviance. *Social & Personality Psychology Compass*, *6*, 723–725.
- Alicke, M. D., Rose, D., & Bloom, D. (2011). Causation, norm violation, and culpable control. *Journal of Philosophy*, *108*, 670–696.
- Alicke, M. D., & Yurak, T. J. (1995). Perpetrator personality and judgments of acquaintance rape. *Journal of Applied Social Psychology*, *25*, 1900–1921.
- Alicke, M. D., & Zell, E. (2009). Social attractiveness and blame. *Journal of Applied Social Psychology*, *39*, 2089–2105.
- American Law Institute. (1962). *Model Penal Code: Official draft and explanatory notes*. Philadelphia, PA: Author.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*, 191–215.
- Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, *54*, 462–479.
- Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology*, *54*, 569–579.
- Bayles, M. D. (1982). Character, purpose, and criminal responsibility. *Law and Philosophy*, *1*, 5–20.
- Bennett, M., & Earwaker, D. (1994). Victims' responses to apologies: The effects of offender responsibility and offense severity. *Journal of Social Psychology*, *134*, 457–464.
- Bierbrauer, G. (1979). Why did he do it? Attribution of obedience and the phenomenon of dispositional bias. *European Journal of Social Psychology*, *9*, 67–84.
- Bohner, G., Bless, H., Schwarz, N., & Strack, F. (1988). What triggers causal attributions? The impact of valence and subjective probability. *European Journal of Social Psychology*, *18*, 335–345.
- Brewer, M. B. (1977). An information-processing approach to attribution of responsibility. *Journal of Experimental Social Psychology*, *13*, 58–69.
- Buss, A. R. (1978). Causes and reasons in attribution theory: A conceptual critique. *Journal of Personality and Social Psychology*, *36*, 1311–1321.
- Cheng, P. W., & Novick, L. R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology*, *58*, 545–567.
- Cheng, P. W., & Novick, L. R. (1992). Covariation in natural causal induction. *Psychological Review*, *99*, 365–382.
- Chockler, H., & Halpern, J. Y. (2004). Responsibility and blame: A structural-model approach. *Journal of Artificial Intelligence Research*, *22*, 93–115.
- Cohen, M., & Nagel, E. (1934). *An introduction to logic and scientific method*. New York, NY: Harcourt, Brace & World.
- Collins, J., Hall, N., & Paul, L. A. (Eds.). (2004). *Causation and counterfactuals*. Cambridge, MA: MIT Press.
- Darley, J. M., Klosson, E. C., & Zanna, M. P. (1978). Intentions and their contexts in the moral judgments of children and adults. *Child Development*, *49*, 66–74.
- Darley, J. M., & Shultz, T. R. (1990). Moral rules: Their content and acquisition. *Annual Review of Psychology*, *41*, 525–556.
- Davis, C. G., Lehman, D. R., Silver, R. C., Wortman, C. B., & Ellard, J. H. (1996). Self-blame following a traumatic event: The role of perceived avoidability. *Personality and Social Psychology Bulletin*, *22*, 557–567.
- Davis, C. G., Lehman, D. R., Wortman, C. B., Silver, R. C., & Thompson, S. C. (1995). The undoing of traumatic life events. *Personality and Social Psychology Bulletin*, *21*, 109–124.
- De Cremer, D., & van Dijk, E. (2002). Perceived criticality and contributions in public good dilemmas: A matter of feeling responsible to all? *Group Processes & Intergroup Relations*, *5*, 319–332.
- DePaulo, B. M., & Friedman, H. S. (1998). Nonverbal communication. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 2, pp. 3–40). New York, NY: McGraw-Hill.
- Dhami, M. K., Mandel, D. R., & Souza, K. (2005). Escape from reality: Prisoner's counterfactual thinking about crime, justice and punishment. In D. R. Mandel, D. J. Hilton, & P. Catellani (Eds.), *The psychology of counterfactual thinking* (pp. 165–182). New York, NY: Routledge.
- Doris, J. (2010). *The moral psychology handbook*. New York, NY: Oxford University Press.
- Driver, J. (2007). Attributions of causation and moral responsibility. In W. Sinnott-Armstrong (Ed.), *Moral psychology* (Vol. 2, pp. 423–439). Cambridge, MA: MIT Press.
- Eisenberg, R., & Hans, V. P. (2009). Taking a stand on taking the stand: The effect of prior criminal record on the decision to testify and on trial outcomes. *Cornell Law Review*, *94*, 1353–1388.
- Epstude, K., & Roese, N. J. (2008). The functional theory of counterfactual thinking. *Personality and Social Psychology Review*, *12*, 168–192.

- Falk, A., & Szech, N. (2013). Morals and markets. *Science*, *340*, 707–711.
- 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.
- Fincham, F. D., & Jaspars, J. M. (1979). Attribution of responsibility to the self and other in children and adults. *Journal of Personality and Social Psychology*, *37*, 1589–1602.
- Fincham, F. D., & Jaspars, J. M. (1980). Attribution of responsibility: From man the scientist to man as lawyer. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 13, pp. 81–138). New York, NY: Academic Press.
- Fincham, F. D., & Jaspars, J. M. (1983). A subjective probability approach to responsibility attribution. *British Journal of Social Psychology*, *22*, 145–161.
- Försterling, F. (1989). Models of covariation and attribution: How do they relate to the analysis of variance? *Journal of Personality and Social Psychology*, *57*, 615–626.
- Frantz, C. M., & Bennis, G. (2005). Better late than early: The influence of timing on apology effectiveness. *Journal of Experimental Social Psychology*, *41*, 201–207.
- Gawronski, B. (2004). Theory-based correction in dispositional inference: The fundamental attribution error is dead, long live the correspondence bias. *European Review of Social Psychology*, *15*, 183–217.
- Gerstenberg, T., & Lagnado, D. A. (2010). Spreading the blame: The allocation of responsibility amongst multiple agents. *Cognition*, *115*, 166–171.
- Gerstenberg, T., & Lagnado, D. A. (2012). When contributions make a difference: Explaining order effects in responsibility attributions. *Psychonomic Bulletin & Review*, *19*, 729–736.
- Gigerenzer, G. (1991). From tools to theories: A heuristic of discovery in cognitive psychology. *Psychological Review*, *98*, 254–267.
- Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, *117*, 21–38.
- Gilovich, T., Griffin, D. W., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment*. New York, NY: Cambridge University Press.
- Giroto, V., Legrenzi, P., & Rizzo, A. (1991). Event controllability in counterfactual thinking. *Acta Psychologica*, *78*, 111–133.
- Goodman, N. D., Tenenbaum, J. B., & Gerstenberg, T. (2015). Concepts in a probabilistic language of thought. In E. Margolis & S. Lawrence (Eds.), *The conceptual mind: New directions in the study of concepts* (pp. 623–654). Cambridge, MA: MIT Press.
- Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, *106*, 148–168.
- Gray, K., & Wegner, D. M. (2008). The sting of intentional pain. *Psychological Science*, *19*, 1260–1262.
- Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry*, *23*, 101–124.
- Greenawalt, K. (1986). Distinguishing justifications from excuses. *Law and Contemporary Problems*, *49*, 89–108.
- Guglielmo, S., & Malle, B. F. (2010). Can unintended side effects be intentional? Resolving a controversy over intentionality and morality. *Personality and Social Psychology Bulletin*, *36*, 1635–1647.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*, 814–834.
- Halpern, J. Y., & Pearl, J. (2005). Causes and explanations: A structural-model approach. Part I: Causes. *The British Journal for the Philosophy of Science*, *56*, 843–887.
- Hamilton, V. L. (1976). Individual differences in ascriptions of responsibility, guilt, and appropriate punishment. In G. Berment, C. Nemeth, & N. Vidmar (Eds.), *Psychology and the law* (pp. 239–264). New York, NY: Lexington Books.
- Hamilton, V. L. (1978). Who is responsible? Toward a social psychology of responsibility attribution. *Social Psychology*, *41*, 316–328.
- Hamilton, V. L. (1980). Intuitive psychologist or intuitive lawyer? Alternative models of the attribution process. *Journal of Personality and Social Psychology*, *39*, 767–772.
- Hart, H. L. A. (1968). *Punishment and responsibility*. New York, NY: Oxford University Press.
- Hart, H. L. A., & Honoré, A. M. (1959). *Causation in the law*. Oxford, England: Oxford University Press.
- Hastie, R. (1984). Causes and effects of causal attribution. *Journal of Personality and Social Psychology*, *46*, 44–56.
- Hastie, R. (in press). Causal thinking in judgments. In G. Keren & G. Wu (Eds.), *The Blackwell handbook of judgment and decision making*. Oxford, England: Blackwell.
- Heider, F. (1958). *The psychology of interpersonal relations*. New York, NY: Wiley.
- Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, *57*, 243–259.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral & Brain Sciences*, *33*, 61–83.
- Hilton, D. J. (2007). Causal explanation: From social perception to knowledge-based attribution. In A. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (2nd ed., pp. 232–253). New York, NY: Guilford Press.
- Hilton, D. J. (2012). The emergence of cognitive social psychology: A historical analysis. In W. Stroebe & A. Kruglanski (Eds.), *Handbook of the history of social psychology* (pp. 45–80). Philadelphia, PA: Psychology Press.
- Hilton, D. J., McClure, J., & Sutton, R. M. (2010). Selecting explanations from causal chains: Do statistical principles explain preferences for voluntary causes? *European Journal of Social Psychology*, *40*, 383–400.
- Hilton, D. J., & Slugoski, B. R. (1986). Knowledge-based causal attribution: The abnormal conditions focus model. *Psychological Review*, *93*, 75–88.
- Hilton, D. J., Smith, R. H., & Kim, S. H. (1995). Process of causal explanation and dispositional attribution. *Journal of Personality and Social Psychology*, *68*, 377–387.
- Hitchcock, C., & Knobe, J. (2009). Cause and norm. *Journal of Philosophy*, *106*, 587–612.
- Hume, D. (2007). *An enquiry concerning human understanding* (P. Millican, Ed.). Oxford, England: Oxford University Press. (Original work published 1748)
- Inbar, Y., & Pizarro, D. A. (2009). Grime and punishment: How disgust influences moral, social, and legal judgments. *The Jury Expert*, *21*, 11–23.

- Jimenez-Leal, W. (2013). Causal selection and counterfactual reasoning. *Revista Colombiana de Psicología, 22*, 179–197.
- Johnson, J. T., & Drobný, J. (1987). Happening soon and happening later: Temporal cues and attributions of liability. *Basic and Applied Social Psychology, 8*, 209–234.
- Jones, E. E. (1979). The rocky road from acts to dispositions. *American Psychologist, 34*, 107–117.
- Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 220–266). New York, NY: Academic Press.
- Jones, E. E., Davis, K. E., & Gergen, K. J. (1961). Role playing variations and the informational value in person perception. *Journal of Abnormal and Social Psychology, 63*, 302–310.
- Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology, 3*, 1–24.
- Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93*, 136–153.
- Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. New York, NY: Cambridge University Press.
- Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201–208). New York, NY: Cambridge University Press.
- Kahneman, D., & Varey, C. A. (1990). Propensities and counterfactuals: The loser that almost won. *Journal of Personality and Social Psychology, 59*, 1101–1110.
- Kanazawa, S. (1992). Outcome or expectancy? Antecedent of spontaneous causal attribution. *Personality and Social Psychology Bulletin, 18*, 659–668.
- Karlovac, M., & Darley, J. M. (1988). Attribution of responsibility for accidents: A negligence law analogy. *Social Cognition, 6*, 287–318.
- Kashima, Y., Siegal, M., Tanaka, K., & Kashima, E. S. (1992). Do people believe behaviours are consistent with attitudes? Towards a cultural psychology of attribution processes. *British Journal of Social Psychology, 31*, 111–124.
- Kelley, H. H. (1960). The analysis of common sense. *PsycCritiques, 5*, 1–3.
- Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska symposium on motivation* (Vol. 15, pp. 192–240). Lincoln: University of Nebraska Press.
- Kelley, H. H. (1971). *Attribution in social interaction*. Morristown, NJ: General Learning Press.
- Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist, 28*, 107–128.
- Kelley, H. H. (1983). Perceived causal structures. In J. M. Jaspars, F. D. Fincham, & M. Hewstone (Eds.), *Attribution theory and research: Conceptual, developmental and social dimensions* (pp. 343–369). New York, NY: Academic Press.
- Kerr, N. L. (1989). Illusions of efficacy: The effects of group size on perceived efficacy in social dilemmas. *Journal of Experimental Social Psychology, 25*, 287–313.
- Kerr, N. L., & Bruun, S. E. (1983). Dispensability of member effort and group motivation losses: Free-rider effects. *Journal of Personality and Social Psychology, 44*, 78–94.
- Kerr, N. L., & Kaufman-Gilliland, C. M. (1997). "... and besides, I probably couldn't have made a difference anyway": Justification of social dilemma defection via perceived self-inefficacy. *Journal of Experimental Social Psychology, 33*, 211–230.
- Kim, J. (1973). Causes and counterfactuals. *Journal of Philosophy, 70*, 570–572.
- Kim, P. H., Dirks, K. T., Cooper, C. D., & Ferrin, D. L. (2006). When more blame is better than less: The implications of internal vs. external attributions for the repair of trust after a competence- vs. integrity-based trust violation. *Organizational Behavior and Human Decision Processes, 99*, 49–65.
- Knobe, J. (2006). The concept of intentional action: A case study in the uses of folk psychology. *Philosophical Studies, 130*, 203–231.
- Knobe, J. (2010). Person as scientist, person as moralist. *Behavioral & Brain Sciences, 33*, 315–365.
- Knobe, J., & Fraser, B. (2008). Causal judgments and moral judgment: Two experiments. In W. Sinnott-Armstrong (Ed.), *Moral psychology, volume 2: The cognitive science of morality* (pp. 441–447). Cambridge, MA: MIT Press.
- Kominsky, J. F., Phillips, J., Gerstenberg, T., Lagnado, D. A., & Knobe, J. (2015). Causal superseding. *Cognition, 137*, 196–209.
- Krauss, R. M., & Fussell, S. R. (1991). Perspective-taking in communication: Representations of others' knowledge in reference. *Social Cognition, 9*, 2–24.
- Lagnado, D. A., & Channon, S. (2008). Judgments of cause and blame: The effects of intentionality and foreseeability. *Cognition, 108*, 754–770.
- Lagnado, D. A., Gerstenberg, T., & Zultan, R. (2013). Causal responsibility and counterfactuals. *Cognitive Science, 47*, 1036–1073.
- Lerner, M. J. (1980). *The belief in a just world*. New York, NY: Plenum.
- Lerner, M. J., & Miller, D. T. (1978). Just world research and the attribution process: Looking back and ahead. *Psychological Bulletin, 85*, 1030–1051.
- Lloyd-Bostock, S. (1983). Attributions of cause and responsibility as social phenomena. In J. Jaspars, F. Fincham, & M. Hewstone (Eds.), *Attribution theory and research: Conceptual, developmental and social dimensions* (pp. 261–289). London, England: Academic Press.
- MacRae, C. N. (1992). A tale of two curries: Counterfactual thinking and accident-related judgments. *Personality and Social Psychology Bulletin, 18*, 84–87.
- Malle, B. F. (2004). *How the mind explains behavior: Folk explanations, meaning, and social interaction*. Cambridge, MA: MIT Press.
- Malle, B. F. (2011). Time to give up the dogmas of attribution: An alternative theory of behavior explanation. In J. M. Olson & M. P. Zanna (Eds.), *Advances in experimental social psychology* (Vol. 44, pp. 297–311). Waltham, MA: Academic Press.
- Malle, B. F., Guglielmo, S., & Monroe, A. E. (2014). A theory of blame. *Psychological Inquiry, 25*, 147–186.
- Mandel, D. R. (2003). Judgment dissociation theory: An analysis of differences in causal, counterfactual, and covariational reasoning. *Journal of Experimental Psychology: General, 137*, 419–434.
- Mandel, D. R. (2005). Counterfactual and causal explanation: From early theoretical views to new frontiers. In D. R.

- Mandel, D. J. Hilton, & P. Catellani (Eds.), *The psychology of counterfactual thinking* (pp. 11–27). New York, NY: Routledge.
- Mandel, D. R. (2010). Predicting blame assignment in a case of harm caused by negligence. *Mind & Society*, 9, 5–17.
- Mandel, D. R. (2011). Mental simulation and the nexus of causal and counterfactual explanation. In C. Hoerl, T. McCormack, & S. Beck (Eds.), *Understanding counterfactuals, understanding causation: Issues in philosophy and psychology* (pp. 147–170). Oxford, England: Oxford University Press.
- Mandel, D. R., & Dhami, M. K. (2005). “What I did” versus “What I might have done”: Effect of factual and counterfactual thinking on blame, guilt, and shame in prisoners. *Journal of Experimental Social Psychology*, 41, 637–645.
- Mandel, D. R., Hilton, D. J., & Catellani, P. (Eds.). (2005). *The psychology of counterfactual thinking*. New York, NY: Routledge.
- Mandel, D. R., & Lehman, D. R. (1996). Counterfactual thinking and ascriptions of cause and preventability. *Journal of Personality and Social Psychology*, 71, 450–463.
- Markman, K. D., & Tetlock, P. E. (2000). “I couldn’t have known”: Accountability, foreseeability and counterfactual denials of responsibility. *British Journal of Social Psychology*, 39, 313–325.
- Mazzocco, P. J., & Alicke, M. D. (2004). On the robustness of outcome bias: No constraint by prior culpability. *Basic and Applied Social Psychology*, 26, 131–146.
- McArthur, L. A. (1972). The how and the what of why: Some determinants and consequences of causal attribution. *Journal of Personality and Social Psychology*, 22, 171–193.
- McClure, J. L. (1998). Discounting causes of behavior: Are two reasons better than one? *Journal of Personality and Social Psychology*, 47, 7–20.
- McEleney, A., & Byrne, R. M. J. (2006). Spontaneous counterfactual thoughts and causal explanations. *Thinking & Reasoning*, 12, 235–255.
- Mill, J. S. (1843). *A system of logic, ratiocinative and inductive: Being a connected view of the principles of evidence, and methods of scientific investigation*. London, England: J.W. Parker.
- Miller, D. T., & Gunasegaram, S. (1990). Temporal order and the perceived mutability of events: Implications for blame assignment. *Journal of Personality and Social Psychology*, 59, 1111–1118.
- Miller, D. T., & McFarland, C. (1986). Counterfactual thinking and victim compensation: A test of norm theory. *Personality and Social Psychology Bulletin*, 12, 513–519.
- Miller, D. T., & Turnbull, W. (1990). The counterfactual fallacy: Confusing what might have been with what ought to have been. *Social Justice Research*, 4, 1–19.
- Miller, J. G. (1984). Culture and the development of everyday social explanation. *Journal of Personality and Social Psychology*, 46, 961–978.
- Morris, M. W., & Larrick, R. P. (1995). When one cause casts doubt on another: A normative analysis of discounting in causal attribution. *Psychological Review*, 102, 331–355.
- Morris, M. W., Nisbett, R. E., & Peng, K. (1995). Causal attribution across domains and cultures. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 577–612). Oxford, England: Clarendon Press.
- Nadelhoffer, T. (2004). On praise, side effects, and folk ascriptions of intentionality. *Journal of Theoretical & Philosophical Psychology*, 24, 196–213.
- Nadelhoffer, T., Nahmias, E., & Nichols, S. (2010). *Moral psychology: Historical and contemporary readings*. Malden, MA: Wiley-Blackwell.
- Nadler, J. (2012). Blaming as a social process: The influence of character and moral emotion on blame. *Law and Contemporary Problems*, 75, 1–31.
- Nadler, J., & McDonnell, M.-H. (2012). Moral character, motive, and the psychology of blame. *Cornell Law Review*, 97, 255–304.
- Nanay, B. (2010). Morality or modality? What does the attribution of intentionality depend on? *Canadian Journal of Philosophy*, 40, 28–40.
- N’gbala, A., & Branscombe, N. R. (1995). Mental simulation and causal attribution: When simulating an event does not affect fault assignment. *Journal of Experimental Social Psychology*, 31, 139–162.
- Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.
- Orvis, B. R., Cunningham, J. D., & Kelley, H. H. (1975). A closer examination of causal inference: The role of consensus, distinctiveness, and consistency information. *Journal of Personality and Social Psychology*, 32, 605–616.
- Osgood, E. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana: University of Illinois Press.
- Pearl, J. (2000). *Causality: Models, reasoning and inference*. Cambridge, MA: Cambridge University Press.
- Piaget, J. (1932). *The moral judgment of the child*. London, England: Routledge & Kegan Paul.
- Pizarro, D. A., & Tannenbaum, D. (2011). Bringing character back: How the motivation to evaluate character influences judgments of moral blame. In M. Mikulincer & P. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil* (pp. 91–108). Washington, DC: APA Press.
- Prinz, J. J. (2007). *The emotional construction of morals*. New York, NY: Oxford University Press.
- Pronin, E., Olivola, C. Y., & Kennedy, K. A. (2008). Doing unto future selves as you would do unto others: Psychological distance and decision making. *Personality and Social Psychology Bulletin*, 34, 224–236.
- Prosser, W. L., Wade, J. W., & Schwartz, V. E. (1982). *Cases and materials on torts*. Mineola, NY: Foundation Press.
- Pyszczynski, T. A., & Greenberg, J. (1981). Role of disconfirmed expectancies in the instigation of attributional processing. *Journal of Personality and Social Psychology*, 40, 31–38.
- Quayle, M., & Naidoo, E. (2012). Social risk and attribution: How considering the social risk of attributions can improve the performance of Kelley’s ANOVA model in applied research. *Journal of Applied Social Psychology*, 42, 1694–1715.
- Robinson, D. N. (2002). *Praise and blame: Moral realism and its applications*. Princeton, NJ: Princeton University Press.
- Roese, N. J. (1997). Counterfactual thinking. *Psychological Bulletin*, 121, 133–148.
- Roese, N. J., & Olson, J. M. (1995). Counterfactual thinking: A critical overview. In N. J. Roese & J. M. Olson (Eds.), *What*

- might have been: *The social psychology of counterfactual thinking* (pp. 1–55). Mahwah, NJ: Erlbaum.
- Ross, L., 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.
- Saks, M. J., & Spellman, B. A. (2015). *The psychological foundations of evidence law*. New York: NYU Press.
- Schlenker, B. R., Britt, T. W., Pennington, J., Murphy, R., & Doherty, K. (1994). The triangle model of responsibility. *Psychological Review, 101*, 632–652.
- Schlenker, B. R., Pontari, B. A., & Christopher, A. N. (2001). Excuses and character: Personal and social implications of excuses. *Personality and Social Psychology Review, 5*, 15–32.
- Sedlak, A. (1979). Developmental differences in understanding plans and evaluating actions. *Child Development, 50*, 536–560.
- Shaver, K. (1985). *The attribution of blame: Causality, responsibility, and blameworthiness*. New York, NY: Springer-Verlag.
- Shaw, M. E., & Sulzer, J. L. (1964). An empirical test of Heider's levels in attribution of responsibility. *Journal of Abnormal and Social Psychology, 69*, 39–46.
- Shen, F., Hoffman, M. B., Jones, O. D., Greene, J. D., & Marois, R. (2011). Sorting guilty minds. *New York University Law Review, 86*, 1306–1360.
- Shultz, T. R., & Schleifer, M. (1983). Towards a refinement of attribution concepts. In J. Jaspars, F. D. Fincham, & M. Hewstone (Eds.), *Attribution theory and research: Conceptual, developmental and social dimensions* (pp. 261–289). London, England: Academic.
- Shultz, T. R., Schleifer, M., & Altman, I. (1981). Judgments of causation, responsibility and punishment in cases of harm-doing. *Canadian Journal of Behavioural Science, 13*, 238–253.
- Shultz, T. R., & Wright, K. (1985). Concepts of negligence and intention in the assignment of moral responsibility. *Canadian Journal of Behavioural Science, 17*, 97–108.
- Shultz, T. R., Wright, K., & Schleifer, M. (1986). Assignment of moral responsibility and punishment. *Child Development, 57*, 177–184.
- Sinnott-Armstrong, W. (Ed.). (2008a). *Moral psychology, vol. 1. The evolution of morality: Adaptations and innateness*. Cambridge, MA: MIT press.
- Sinnott-Armstrong, W. (Ed.). (2008b). *Moral psychology, vol. 2. The cognitive science of morality: Intuition and diversity*. Cambridge, MA: MIT press.
- Sinnott-Armstrong, W. (Ed.). (2008c). *Moral psychology, vol. 3. The neuroscience of morality: Emotion, brain disorders, and development*. Cambridge, MA: MIT press.
- Sloman, S. A., & Lagnado, D. (2015). Causality in thought. *Annual Review of Psychology, 66*, 223–247.
- Snyder, C. R., & Higgins, R. L. (1988). Excuses: Their effective role in the negotiation of reality. *Psychological Bulletin, 104*, 23–35.
- Spellman, B. A. (1997). Crediting causality. *Journal of Experimental Psychology: General, 126*, 323–348.
- Spellman, B. A., Kincannon, A. P., & Stose, S. J. (2005). The relation between counterfactual and causal reasoning. In D. R. Mandel, D. J. Hilton, & P. Catellani (Eds.), *The psychology of counterfactual thinking* (pp. 28–43). New York, NY: Routledge.
- Sripada, S. S. (2011). Mental state attributions and the side-effect effect. *Journal of Experimental Social Psychology, 48*, 232–238.
- Taylor, S. E. (1991). Asymmetrical effects of positive and negative events: The mobilization-minimization hypothesis. *Psychological Bulletin, 110*, 67–85.
- Tetlock, P. E. (2002). Social functionalist frameworks for judgment and choice: Intuitive politicians, theologians, and prosecutors. *Psychological Review, 109*, 451–471.
- Thibaut, J. W., & Riecken, H. W. (1955). Some determinants and consequences of the perception of social causality. *Journal of Personality, 24*, 113–133.
- Trabasso, T., & Bartalone, J. (2003). Story understanding and counterfactual reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 904–923.
- Turri, J., & Blouw, P. (in press). A study in rule-breaking. *Philosophical Studies*.
- Uhlmann, E. L., Pizarro, D. A., & Diermeier, D. (2015). A person-centered approach to moral judgment. *Perspectives on Psychological Science, 10*, 72–81.
- Uttich, K., & Lombrozo, T. (2010). Norms inform mental state ascriptions: A rational explanation for the side-effect effect. *Cognition, 116*, 87–100.
- Van Overwalle, F. (1997). A test of the joint model of causal attribution. *European Journal of Social Psychology, 27*, 221–236.
- Walker, D. E., Smith, K. A., & Vul, E. (2015). The “fundamental attribution error” is rational in an uncertain world. In *Proceedings of the 37th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science. Retrieved from http://www.evullab.org/pdf/walker_cogsci2015_final.pdf
- Walsh, C. R., & Sloman, S. A. (2011). The meaning of cause and prevent: The role of causal mechanism. *Mind & Language, 26*, 21–52.
- Weiner, B., Amirkhan, J., Folkes, V. S., & Verette, J. A. (1987). An attributional analysis of excuse giving: Studies of a naive theory of emotion. *Journal of Personality and Social Psychology, 52*, 316–324.
- Wells, G. L., & Gavanski, I. (1989). Mental simulation of causality. *Journal of Personality and Social Psychology, 56*, 161–169.
- Wissler, R. L., & Saks, M. J. (1985). On the inefficacy of limiting instructions: When jurors use prior conviction evidence to decide on guilt. *Law and Human Behavior, 9*, 37–42.
- Wolff, P., Barbey, A. K., & Hausknecht, M. (2010). For want of a nail: How absences cause events. *Journal of Experimental Psychology: General, 139*, 191–221.
- Wong, P. T. P., & Weiner, B. (1981). Why people ask “why” questions, and the heuristics of attributional search. *Journal of Personality and Social Psychology, 40*, 650–663.
- Young, L., & Phillips, J. (2011). The paradox of moral focus. *Cognition, 119*, 166–178.
- Zechmeister, J. S., Garcia, S., Romero, C., & Vas, S. N. (2004). Don't apologize unless you mean it: A laboratory investigation of forgiveness and retaliation. *Journal of Social & Clinical Psychology, 23*, 532–564.
- Zultan, R., Gerstenberg, T., & Lagnado, D. A. (2012). Finding fault: Counterfactuals and causality in group attributions. *Cognition, 125*, 429–440.