CAUSES AND CAUSAL ATTRIBUTIONS

Questions Raised by Dave Hamilton and Spontaneous Trait Inferences

James S. Uleman

Causality is complex, to say the least. The online *Stanford Encyclopedia of Philosophy* (http://plato.stanford.edu/) has 308 entries on this topic. These include Aristotle’s four kinds of causes—material, formal, efficient, and final (Falcon, 2012)—as well as Bertrand Russell’s well-known assertion that, “The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm” (Russell, 1913, p. 1). For those who prefer mathematical formulations to verbal theories, Pearl (2000) presents a masterful account, with a helpful non-mathematical epilogue (or see Sloman, 2005). The point is that there are many conceptions of causality, both in terms of the processes by which people identify causes, and the meanings and implications of calling something a cause.

Social psychologists are most likely to think of the various formulations of causal attribution theory, particularly Jones and Davis’s (1965) on inferring dispositions from actions, and Kelley’s (1967) ANOVA model of how we infer the locus of causality. These textbook staples are not without their critics (e.g., Hilton, 2007; Malle, 2004). But they and the paradigms that support them provide some core meanings of causality in social psychology.

With this seminal work in mind, this chapter examines whether or in what ways spontaneous trait inferences (STIs) might be causal inferences, and the inferred traits might be causes. (We treat *traits* and *dispositions* as synonymous throughout.) This question arises because STIs occur outside the usual context of explaining events or answering questions—in fact, occur outside of awareness. STIs arise without intentions, simply in the course of comprehending text that describes trait-implying behaviors. Participants in STI studies are asked to memorize or familiarize themselves with text such as “The secretary solved the mystery half-way through the book.” A wealth of evidence shows that most people infer
clever and associate that concept with the secretary and/or her photo, without realizing they have done so (Uleman, Rim, Saribay, & Kressel, 2012). The evidence includes study participants falsely recognizing clever as having been in the sentence with the photo (Todorov & Uleman, 2002), learning to explicitly associate clever with the secretary’s photo more readily (Carlston & Skowronski, 1994), explicitly rating the secretary as more clever than they would have otherwise (Skowronski, Mae, Carlston, & Crawford, 1998), and more readily recalling the behavior description when cued by clever (Winter & Uleman, 1984).

But all this evidence begs the question of what clever means. Does it mean that the secretary’s action was clever, or that she is a clever person, or that her cleverness caused her action? Wittgenstein famously said that, “the meaning of a word is in its use in the language” (Brenner, 1999). But in this case, there is no “use in the language.” The concept clever is activated, as seen in various kinds of reaction time data. But it is never “used” until a dependent variable measure probes for it in one way or another: “did you see this word?” or “how clever is this person?” Most words can be used in many ways, as seen in their multiple dictionary definitions. Even when the part of speech is specified, trait terms are ambiguous, even polysemous. Just as a paper can be (a) this chapter you’re reading, or (b) the material on which it may be printed, or (c) a newspaper delivered to your front porch, a trait can describe a behavior, a person in the moment, or an enduring characteristic. See Uleman (2005) for more examples, and research on the ways that people use and misuse this ambiguity of traits.

Hamilton (1998) raised this question of meaning in an insightful chapter, arguing more broadly for a distinction between “dispositional and attributional inferences.” I will review his argument as it bears on STIs; raise questions about it; describe recent evidence that traits are causes in several senses; and finally attempt to reconcile these positions.

**Hamilton’s Argument: Descriptions Versus Causal Attributions**

Hamilton’s (1998) chapter is a commentary on the work of Trope (1986, 1998) on intentional dispositional inferences. He notes that in Trope’s theory, “the first step involves the identification of the observed behavior in disposition-relevant terms. This identification then serves as an hypothesis that is tested in the second, or inference, stage . . . the identification stage is essentially automatic, whereas the inference stage is resource demanding” (Hamilton, 1998, p. 102). Although different in several respects, Gilbert’s (1989) model of intentional dispositional inferences also contains early automatic stages (in which the behavior is categorized in trait-relevant terms and the person is similarly characterized), followed by a more deliberate and often insufficient correction stage that takes account of the situation. Hamilton notes that the automaticity of Trope’s early identification stage (and Gilbert’s categorization and characterization stages) is supported by research suggesting that STIs are largely
automatic (e.g., Newman & Uleman, 1989). Thus STIs are equated with the identification of behaviors in disposition-relevant terms, i.e., correspondent inferences (Jones & Davis, 1965).

Hamilton (1998) then turns to the powerful role of perceivers' expectancies in impression formation, particularly to differences in how information is processed when expectancies are confirmed versus disconfirmed. His central thesis is that unambiguous disconfirmation of prior expectancies initiates "qualitatively different inference processes" (p. 104) from those that Trope and Gilbert describe, and in which STIs are central. "Consider first the case in which... the person's... behavior is entirely consistent with expectancy... The behavior itself is easily interpreted" (pp. 104-105). The process of interpretation yields essentially what STIs produce. "In contrast... [when] the behavior is clearly inconsistent with expectancy... a different set of processes is likely to be invoked" (p. 105).

He supports the existence of this different set of processes by citing research showing that (a) perceivers take more time to process inconsistent than consistent information; (b) perceivers "spontaneously retrieve from memory other information they have already processed about the target person" when behaviors are inconsistent; and (c) "when the behavior does not fit with the perceiver's impression of the actor, the perceiver tries to explain why that behavior occurred" (Hamilton, 1998, p. 105, including relevant research references). In short, perceivers ask why this inconsistent behavior occurred. Thus "attributorial analysis—the analysis of causation—is at the heart of those differences... attributorial thinking is most likely to occur precisely under those conditions when correspondent inferences are least likely to be made, specifically, when the behavior violates the operative expectancy" (p. 106).

In short, STIs such as clever are mere descriptions (of behaviors and/or persons) in trait-relevant terms, but are not causes. "Causes" are only inferred by an attributorial analysis, triggered by such events as learning inconsistent information about a target person (or by direct causal questions, of course). "Given an expectancy-consistent behavior, perceivers tend to make simple, spontaneous, non-analytic dispositional inferences—moving from acts to dispositions, making correspondent inferences... the perceiver is not engaging in an analysis of why... When behavior conforms to expectancy, there's no reason to ponder why it happened... the process of making these correspondent inferences does not seem to involve much attributorial thinking" (p. 106).

At least two points are noteworthy in Hamilton's (1998) treatment. First, Hamilton does not ask why behaviors are characterized "in disposition-relevant" rather than other terms. He is not alone in this omission (e.g., see Gilbert, 1989; Trope, 1998), but it remains an important issue. Second, Hamilton's view is that a disposition or trait concept itself (e.g., clever) cannot be identified as causal per se. Only the cognitive process by which it arises identifies it as causal. If it is a response to an explicit or implicit why question, often triggered by inconsistent information, it is causal; otherwise it is merely descriptive. (As in much of person
perception, issues of accuracy are not addressed. All that matters is the perceiver’s perception and how it was arrived at.)

Carlston and Skowronski (2005) Argued That STIs Are Attributions

A different view of STIs’ attributional status emerges from contrasting STI with STT, spontaneous trait transference. Skowronski, Carlston, Mae, and Crawford (1998) discovered that under some circumstances, the concept clever becomes associated not with the person who enacts the behavior (the secretary in this example) but with another person (a communicator or informant) who describes that behavior by someone else but does not perform it. So if the office manager says that the secretary solved the mystery half-way through the book, we may later believe that the office manager is clever, rating her higher on the trait, and relearning an association between office manager and clever more readily in a paired associates learning task (Carlston & Skowronski, 1994). Interestingly, STTs occur only when representations of the actors, e.g., photos, are absent at encoding. Actor representations block STT (Crawford, Skowronski, & Stiff, 2007; Crawford, Skowronski, Stiff, & Leonards, 2008; Goren & Todorov, 2009).

Carlston and Skowronski (2005) argued that STIs and STTs arise from different processes, and cited several lines of evidence in support of this. First, STIs show larger effect sizes than STTs, suggesting deeper and more elaborate processing for STIs (e.g., Goren & Todorov, 2009; Skowronski et al., 1998; Todorov & Uleman, 2004). Second, STIs but not STTs often show the negative halo effects characteristic of attributions, apparently not through valenced trait associations, but through the informant’s perceived likeability when conveying negative information (Carlston & Skowronski, 2005; Crawford, Skowronski, & Stiff, 2007; Crawford, Skowronski, Stiff, & Scherer, 2007; Skowronski et al., 1998; Wells, Skowronski, Crawford, Scherer, & Carlston, 2011). Third, telling participants that photos and behaviors were randomly paired reduces STI but not STT, as attributional logic would suggest. Fourth, early research showed that cognitive load reduces STI but not STT, suggesting that STT is more automatic than STI (Skowronski et al., 1998). (However, more recent work by Wells et al. (2011) shows that both STI and STT require working memory capacity.)

Crawford, Skowronski, Stiff, and Scherer (2007) also showed that asking participants to “determine whether each informant was lying” (p. 681), rather than asking them to merely familiarize themselves with the stimuli, reduces STI effects but has no effect on STT effects (using both a savings in relearning and a trait rating task). This, too, is consistent with an attributional account of STI and a less complex associationist account of STT.

So does all this evidence show that STIs are attributions? It depends on what is meant by “attributions.” None of this research uses Hamilton’s (1998) criterion of posing causal questions to participants, or asking them why? All of the inferences
are "spontaneous," that is, unintentional and largely unconscious. Carleton and Skowronski (2005) argue that STIs are attributions because they (a) "involve deeper and more elaborative processing [than STTs], (b) which depends more on processing goals and available cognitive capacity and (c) which produces stronger and more enduring person–trait associations (d) that involve labeled links, indicating that the trait is a property of the person, and (e) . . . [are] especially true for traits implied by negative behaviors" (p. 886). They certainly share more properties with attributions than do STTs.

But are these properties unique to causal attributions, versus other kinds of deliberations? For example, aren't they also true of arithmetic calculations, where the answer is a property of the calculation requested (e.g., multiplication or subtraction), or deliberative choices, where negative information carries more weight? It seems the most we can conclude is (a) that STI is more attributional than STT and (b) that none of this evidence addresses whether or not STIs are causal attributions. STIs can be seen as either attributional (Carleton & Skowronski, 2005) or not attributions (Hamilton, 1998), depending on what they are contrasted with (STTs, or answers to why questions, respectively).

One thing that is missing from this discussion is agreement on what the criteria of attributions might be. Hamilton (1998) is relatively clear, in citing Jones and Davis (1965), Trope (1986), and Gilbert (1989). Carleton and Skowronski (2005) are less so. Part of the problem is that there are many attribution theories. For a good overview of the range of phenomena and theories that "attributions" encompass, with no mention of STI, see Hilton (2007). That chapter emphasizes linguistic and conversational influences on attributions, and suggests additional ways of thinking about Hamilton's position.

Three Other Observations Prompted by Hamilton (1998)

First, as noted earlier, the phrase disposition-relevant terms occurs in all the theories of trait attributions that Hamilton reviews. Behaviors are described (categorized, identified) in disposition-relevant terms. The foundation of the correspondence bias is that there are such correspondent terms in the lexicon, terms that are equally applicable to describing actions or actors. Why should this be so, and how general is it? Is the default to explain things as the ancients did in terms of "the nature of things," so that things float "because" they are buoyant rather than because of the weight of the liquid they displace, or things fall "because of" their weight rather than gravity? With regard to people, part of the answer may lie in the development of people's theory of mind. Even in the first year of life, infants distinguish between animate and inanimate movement. And they make goal-like, and trait-like, inferences about animate but not inanimate actors. For example, infants anticipate different patterns of behavior by human arms than by robotic arms. See Apperly (2011) for a recent review of this extensive, fascinating, and contentious literature.
Furthermore, there is good evidence that the development of theory of mind depends on, and drives language development (Baird & Astington, 2005). For social psychologists, perhaps it is enough to observe that we have a different set of terms to describe people and inanimate things. In Malle's (2004) terms, inanimate things do not perform intentional actions and only exhibit unintentional behavior. Or rather, intentions are not relevant to inanimate things. Traits can be used to describe both intentional (skilled) and unintentional (clumsy) acts, and their meanings (uses) differ and depend on the situation and the decades of each individual's cognitive and linguistic development. There is much to learn about the development of and many uses of trait terms.

Traits are not the only concepts that describe behavior. So a related question is, why infer traits and not goals or behavioral gists? In fact, people do spontaneously infer goals (Hassin, Aarts, & Ferguson, 2005) and gists (Winter, Uleman, & Cunniff, 1985) from trait-implying behavior descriptions. Goals and gists are inferred simultaneously (Uleman & Moskowitz, 1994). So reading "The doctor severs the patient's jugular vein during the operation" produces spontaneous inferences of both incompetent (a trait) and surgery (a gist). Finally, the same behavior can produce simultaneous inferences of both actors' traits and characterizations of the situation (Ham & Vonk, 2003; Todd, Molden, Ham, & Vonk, 2011). Reading that "Mary got an A+ on the calculus exam" yields simultaneous spontaneous inferences of smart and easy.

Thus identifying behaviors in "disposition-relevant terms" is more complex than it might seem, because many different characterizations can be and are simultaneous inferred from behaviors (Reeder, Vonk, Ronk, Ham, & Lawrence, 2004). Only some of them are disposition-relevant. What selects the disposition-relevant ones? More theory and research are needed.

Second, there is evidence of spontaneously inferred causes that are not traits. Hassan, Aarts, and Ferguson (2005) showed spontaneous goal inferences—goals that are usually regarded as causes. And Hassan, Bargh, and Uleman (2002) showed spontaneous causal inferences outside the domain of mental events. In three cued-recall studies, participants read short scenarios such as "After spending the day exploring beautiful sights in the crowded streets of New York, Jane discovered that her wallet was missing" and rated the scenarios on "how interesting they are." Although pickpocket was the implied cause of this event, it was not implied by the control sentence, "Before leaving home for a day of exploring beautiful sights in the crowded streets of New York, Jane discovered that her wallet was missing." Then participants spent 5 minutes filling in the names of states on an outline map of the United States and then completed a surprise recall test for the scenarios. Cues were either the implied causes or important words from the scenarios (repetition cues such as sights). Causes were more effective retrieval cues for causal scenarios than noncausal scenarios, whereas repetition cues showed no difference. Participants showed no awareness of making causal inferences or their role in recall.
Hassin et al. (2002) pointed out that this evidence of spontaneous causal inferences complements the work of McKoon and Ratcliff (1986) showing spontaneous effect ("predictable") inferences. They used scenarios such as "Superman grabbed the prison bars that enclosed him and started to pull slowly" and, as the control for this, "Superman grabbed the villain and put him behind prison bars." Predicted events (i.e., effects), such as bend, cued recall of the predicting scenarios more effectively than the control scenarios. These two papers provide good evidence that both causes and effects are spontaneously inferred.

Hamilton's (1998) thesis is that causal attributions only arise after a causal question is posed. Perhaps the stimuli in Hassin et al. (2002) pose causal questions implicitly if not explicitly. Most of the scenarios they used described unexpected or unusual events with causes such as flu, drowning, flood, fire, food poisoning, and earthquake. Often the cause of the outcome was omitted and had to be inferred from its effects. Thus why? was often implicit. But rather than following from the kind of causal attribution processes described by Jones and Davis (1965) or Kelley (1967), these answers to why? more often seem to arise from filling in the blanks in story schemata (Graesser & Bower, 1990; Kintsch & van Dijk, 1978; Schank & Abelson, 1977) in order to comprehend the text. They are nouns rather than adjectives, and there is no "correspondence" between descriptions of behaviors and causes. Nor is there information about covariation. Inferring that Jane lost her wallet to a pickpocket, or that "Joe's house in Los Angeles suffered severe damage" because of an earthquake, requires retrieving causal schema embedded in world knowledge, rather than the kinds of attributional inferences classically considered by social psychologists.

Thus the question why? may often arise implicitly in comprehending events. But inferring the answer depends on accessing world knowledge which is already organized in causal terms. Comprehension involves selecting which of many possible schemata best fit the information given. "Correspondent inferences" seem to be a special case of causal inferences that depend on the existence of particular lexical terms which enable correspondent labeling.

Third, causal schemata can be elicited by, or are inherent in the meaning, of single words. The well-known work on causality implicit in verbs (Rudolph & Forsterling, 1997) illustrates this. Learning that "John impresses Mary" or "John bores Mary" typically implies that John is impressive or boring; whereas learning that "John admires Mary" or "John pities Mary" implies that Mary is admirable or pitiful. Thus "action" and "state" verbs, respectively, imply that the cause lies in something about John in the first case, or Mary in the second (Semin & Fiedler, 1988). A causal schema is part of the meaning of the verb. Notice that what is implied by the causality implicit in verbs is the locus of causality, that is, the person who caused the event. The cause itself, within the person, is often a cognate of the verb. Impressive and boring are more accessible than impressionable and bored; admirable and pitiful are more accessible than admiring and pitying. Whatever the preferred explanation for these effects (see the review by Rudolph &
Forsterling, 1997), the extensive evidence on causality implicit in verbs illustrates the more general principle that words’ meanings often include schemata that describe the contexts in which they may be used and what they apply to, as well as their implications, causal and otherwise.

Are these implications attributional, descriptive, or definitional? How do they depend on the conversational context? In any case, this research on the causality implicit in verbs asked participants for intentional explicit inferences and causal judgments, so they are not spontaneous. Nevertheless, it illustrates the idea that causality can be inherent in the meanings of words.

**McCarthy and Skowronski (2011) Showed That STIs Are Predictive**

Hassin et al. (2002) showed spontaneous causal inferences by targeting events that most people would agree are causes, for example, illness or natural disasters. But in the absence of such general agreement, what kind of evidence could be cited that traits are causal. One criterion might be that they are predictive of future behavior. Isolating dispositional causes within people allows one to predict their behavior across situations that may be very different from each other, but which also engage that disposition.

McCarthy and Skowronski (2011) investigated the possibility that STIs are causes by assessing the extent to which participants would rely on STIs in predicting actors’ behaviors in situations quite different from those in which the traits were inferred. In other words, the only things linking the “observed” or described behaviors and the predicted behaviors were the particular traits they evidenced. McCarthy and Skowronski (2011) don’t give any examples, but they are easy to imagine. For example, those who read that “She returned the wallet with all the money in it” (implying honest) should believe this same actor is more likely to “finally tell him that he has bad breath,” even though the events have little in common except honesty. And in their first two studies, that is what they found. In Study 1, participants familiarized themselves with pairs of actors and behaviors, most of which implied traits. Then they matched actor photos with new behaviors that they thought the actors were most likely to do. In Study 2, they estimated the likelihood that actors would perform specific new behaviors. Results suggested that STIs guide both the predictive matches and the likelihood ratings. These effects occurred independent of valence (halo effects) and did not depend on recall of behaviors.

Study 3 was designed to replicate these effects; to strengthen the case for STIs’ causal roles by including a condition meant to interfere with STI and hence with prediction; and to examine how much automatic processes contribute to the predictions, through an adaptation of Jacoby’s (1991) process dissociation procedure (PDP). Participants familiarized themselves with the stimuli, as before, and then rated the likelihood of actors engaging in various behaviors. The basic
findings of Studies 1 and 2 replicated. Under instructions that interfere with STI (lie detection), these effects were significantly reduced. And the PDP analyses suggested that this reduction in STI effects on predictions occurred through a change in the automatic, rather than the controlled component of this effect, as one would expect. The authors identify these components with “different sources of memory that could affect behavior predictions.” The automatic component is “an implicit inference that exerts an influence outside of awareness” (p. 330). McCarthy and Skowronski (2011) conclude

that spontaneous inferences behave in ways resembling the intentional judgments studied by researchers examining intentional inferences, as occurred in explorations of Correspondent Inference Theory (Jones & Davis, 1965). . . . Trait inferences act as if they have causal force, in that they are used to construct predictions of future behavior.

(p. 330)

So if by “causal” one means that traits provide the basis for making predictions of behavior in situations with little surface resemblance to those on which the initial trait inferences were based, these studies provide good evidence that STIs are causal.

Of course it is possible that the traits were not encoded as causes, but were merely associated with the actors in these studies and then used in a causal manner when prediction questions were asked. These studies do not rule out this possibility, although it’s unclear what being spontaneously “not encoded as causes” might mean or how it might be tested. This possibility is made less plausible by recent evidence that traits themselves are inherently causal.

Kressel’s (2011) Evidence That Trait Terms Are Inherently Causal

Links among concepts in semantic memory are not merely associative. The links themselves, apart from the concepts, may carry different meanings. Superordinate links (“is a”) reflect hierarchical relationships (a sparrow “is a” bird). Part-object links describe structure (windows are “part of” a building); antonym relations link opposites whereas synonym relations link concepts with the same meanings. There are functional relations (knife-cut), habitat relations (bird-nest; bear-cave), and innumerable others (see Collins and Loftus, 1975; Dosher, 1983; Moss, Ostrin, Tyler, & Marslen-Wilson, 1995; Quillian, 1967).

Causal relations (links) have not been much investigated, until recently. Fenker, Waldmann, and Holyoke (2005) showed that such links exist in semantic memory and produce interesting asymmetries in response times (RTs). Participant in their studies judged “yes or no,” as quickly as possible, “whether the concept described by the first word caused or is caused by the concept described by the second word.”
That is, are they causally related or not? One word of a pair was shown for 1000 ms, followed immediately by the second word of the pair. The pairs included concepts that were causally related (sunshine—freckles, acid—corrosion) as well as ones that were not. When the causally related concepts were presented in the predictive order (cause—effect), participants responded with yes faster than when they were presented in the diagnostic order (effect—cause). Fenker et al. (2005) speculated that this occurs because cause and effect are experienced in that order; we become aware of causes before effects. When concepts are presented in the reverse order, participants must reverse them to match them with the order stored in semantic memory, and this takes extra time.

Study 1 of Kressel's (2011) dissertation asked whether this would be true of traits (causes) and actions/verbs (effects). On one hand, the same asymmetry should occur if traits are causes of actions (e.g., friendly—greet; aggressive—dominate). On the other hand, it should not occur because (a) traits are never observed, so they cannot be said to occur before actions; (b) decades of STI research have shown that people readily (almost automatically) infer traits from actions, but not vice versa; and (c) people are more likely to infer traits from actions than vice versa (Maass, Cadinu, Taroni, & Masserini, 2006). So she adapted the paradigm of Fenker et al. (2005) to see what would happen.

Following Fenker et al. (2005), trait—action (adjective—verb) pairs were chosen such that the associations between them were weak and symmetrical (according to the Southern Florida Word Association Norms; Nelson, McEvoy, & Schreiber, 1998), and the estimated conditional probabilities of one given the other were symmetrical. For the diagnostic conditional probability, pretest participants estimated, "If 100 people [action], how many are [trait]?" For predictive conditional probabilities, pretest participants estimated, "If 100 people are [trait], how many [action]?" These trait—action pairs were then used in a social version of the Fenker et al. (2005) paradigm, and nonsocial cause—effect pairs from the Fenker et al. study were used in a nonsocial replication. What happened?

The RT asymmetry was the same for the social and the nonsocial tasks. Participants detected causal relations 72 ms faster when the social words appeared in the predictive (friendly—greet) rather than the diagnostic (greet—friendly) order. And they detected causal relations 77 ms faster when the nonsocial words appeared in the predictive rather than the diagnostic order. There was no interaction with type of task. This strongly suggests that trait concepts, in and of themselves, are causal. See Kressel and Uleman (2010) for a more complete account of Study 1.

Study 2 used a relation priming paradigm to overcome one drawback of Study 1, that causality was asked about explicitly. Making causality explicit might have primed causal relations in general, links that might not have been activated otherwise. (Of course, to be primed by the instructions, such links have to be present or applicable in the first place.) If traits and actions are causally linked in semantic memory, it should be possible to prime the link itself without ever
mentioning causality explicitly. McKoon and Ratcliff (1995) showed that semantic links between pairs of concepts can be primed by task contexts. They presented participants with a lexical decision task in which pairs of "words" (letter strings) occurred, and the task was to decide as quickly and accurately as possible whether the second string was an English word. When pairs of opposites (hot–cold) were embedded in a list of opposites rather than synonyms (close–near), lexical decision RTs for the second word were shorter than when they were embedded in a list of synonyms. The nature of the link (opposites) was never mentioned. So the question is, would such a list-context effect occur with causal links and trait-action pairs. This was examined in Study 2 (Kressel, 2011).

Trait–action pairs of words were embedded in causal list contexts and in associated list contexts. Each list contained 14 to 28 word–nonword pairs, 54 cause–effect pairs (e.g., sunshine–freckles) or 54 associated pairs (e.g., print–type), 4 trait–action pairs, and 4 unrelated word–action pairs. Participants made speeded lexical decisions for the second word in each pair, for four lists, two causal and two associated. On each trial, a fixation cross was shown for 250 ms, followed by the first word for 250 ms, and then the target word until a response (pressing "word" or "nonword") occurred. Within causal lists, actions primed by traits were recognized as words 78 ms faster than unprimed actions (520 ms vs. 598 ms, \( p < .01 \)). Within associated lists, there was no difference between actions primed by traits and those primed by other words (569 ms vs. 583 ms, \( p > .40 \)). The Prime × List interaction was significant (\( p < .05 \)).

So without ever mentioning causes or causality, this study shows that there are causal links between traits and behaviors. Priming them via the list context speeds lexical decisions, that is, speeds access to the primed words.

Study 3 had two purposes. One was to explore an alternative hypothesis to that of Fenker et al. (2005): that the RT asymmetry of their study and Kressel’s (2011) first study is due to people observing causes before effects. This is clearly not true of traits and behaviors because traits are never observed. So might the effect be due to the mental template the task invokes, that is, a predictive template (cause \( \rightarrow \) effect) rather than a diagnostic template (effect \( \rightarrow \) cause)? The second purpose was to link the RT asymmetries to STIs. If STIs are inferences about causes, then those who see traits as more causal (as evidenced by larger RT asymmetries) may be more likely to make STIs. STIs were measured with the false recognition paradigm introduced by Todorov and Uleman (2002). First, participants completed the STI task. Then they made relational decisions to pairs of words, as in Study 1. In Study 3, two different relational judgments were made, by different groups of participants. Finally, they completed several measures of explicit causal theories about traits.

First, consider RT asymmetries in relational judgments. One group of participants got the instructions of Study 1, to indicate (yes/no) whether the pair of concepts were causally related, that is, whether one concept causes or is caused by the other. Another group of participants were asked to "determine whether
one concept explains or is explained by the other concept” (Kressel, 2011, p. 29). The idea was that the causal instructions invoke a predictive template, in which causes naturally precede effects. But these explanatory instructions should invoke a diagnostic template, in which effects naturally precede causes:

Two concepts have an explanatory relationship when one of the concepts explains, gives mechanism, or gives reason for the other concept. For example, boiling water can be explained by the application of heat; thus the concepts of boiling and heat are explanatorily related. Similarly, shortness of breath can be explained by exercise; thus the concepts of shortness of breath and exercise are explanatorily related. Finally, a stomach ache can be explained by overeating; thus the concepts of stomach ache and overeating are explanatorily related. (Kressel, 2011, p. 35)

If the RT asymmetry is due to the template participants use in making their judgments, rather than the way the concepts “are stored” in semantic memory or the order in which they are normally experienced, then the explanatory instructions should reverse the asymmetry found in Study 1.

As hypothesized, those in the predictive condition (with causal instructions) correctly identified causal relations among predictive pairs (trait → act) faster (1060 ms) than among diagnostic pairs (act → trait; 1139 ms). This replicated the findings of Study 1. More importantly, those in the diagnostic condition (with explanatory instructions) correctly identified explanatory relations among diagnostic pairs (act → trait) faster (1010 ms) than among predictive pairs (trait → act; 1094 ms); ps < .05.

Note that both causal and explanatory relations involve causes. They only differ in whether one reasons from cause to effect (predictive) or effect to cause (diagnostic). Both produce asymmetries in recognition RTs, which favor word pairs presented in the order of reasoning implied by the instructions template. And both demonstrate that trait concepts function as causes, in both predictive and diagnostic frameworks.

Second, STIs were measured with the false recognition paradigm, specifically via differences in error rates between recognition of implied (but not explicitly presented) traits when paired with the target photos about whom they were implied, and error rates for implied traits paired with different target photos. This measure showed that STIs occurred in Study 3. But the important question was whether individual differences in forming STIs are related to the asymmetries in RTs. Additionally, Kressel looked at possible effects of explicit theories of traits’ utility for predicting behaviors (e.g., traits’ temporal stability, consistency across situations) and of situations’ predictive utility (Church et al., 2003; Church, 2009). These two scales are only weakly correlated.

In the predictive framework, STIs and recognition RT asymmetries correlated +0.26 (p = .068), but explicit trait theories did not correlate with STIs. A
simultaneous multiple regression analysis that also included the nonsocial relation recognition RT asymmetries revealed essentially the same thing. That is, participants with larger causal relation recognition RT asymmetries tended to produce more STIs; explicit theories about traits' predictive utility made no difference; and nonsocial relation recognition RTs were unrelated. This suggests that the strength of implicitly viewing traits as causal (as seen in RT asymmetries) tended to predict the likelihood of making STIs, again suggesting that STIs are causal.

In the diagnostic framework, STIs and recognition RT asymmetries correlated −0.33 ($p = .019$), and explicit trait theory correlated with STIs +0.37 ($p = .009$). A simultaneous multiple regression analysis, identical to the one above, revealed that these two effects were independent of each other. That is, participants with larger explanatory relation recognition RT asymmetries produce fewer STIs. And independent of this, those with stronger explicit beliefs (that traits can predict behavior) produced more STIs. Even though the direction of the RT versus STI correlation was unexpected, it provides further evidence that the degree to which traits are implicitly understood as referring to causality (specifically, to explanation) is related to the likelihood of forming STIs. And the results of both studies suggest that these effects are unrelated to explicit theories about traits' predictive utility.

It is not clear why the predictive asymmetry was positively correlated, and the diagnostic asymmetry was negatively correlated with STI. The original idea was that these asymmetries were indicative of the strength of predictive and diagnostic links between traits and behaviors, and that each of them should correlate positively with STI. This simple hypothesis was not supported, and a clearer explanation of the findings must await future research. However, there are two differences between relation recognition RTs and STIs. First, relation recognition is a judgment task, whereas the STI task is a more complex comprehension or extraction-and-generation-of-meaning and recognition task. Second, and consistent with this difference, relation recognition occurs much more quickly (about 1,000 ms) than STI (with typical exposure times of 6 to 8 s for participants and minutes between encoding and recognition). Although these exposure times do not tell us the minimum times it takes STIs to develop, they do allow for more complex processes that mere relation recognition. These findings merit replication and further examination.

In both template conditions of Study 3, participants also performed nonsocial relation recognition tasks using the nonsocial stimuli of Fenker et al. (2005). RT asymmetries on these tasks were unrelated to STIs. This suggests that social and nonsocial causality and causal explanation are understood and function differently, as many (e.g., Keil, 1989; Malle, 2004) have suggested.

In short, Kressel (2011; Kressel & Uleman, 2010) showed that traits are causal, not merely descriptive concepts. This was shown in two paradigms (relation recognition and relation priming), and replicated with relation recognition in Study 3. In addition, Study 3 showed that two kinds of causal relations (predictive
and diagnostic) link traits and behaviors and that each of these is related to STI, independent of explicit theories about traits (Church et al., 2003). Even though traits can be used descriptively, causality is inherent in their meaning.

Conclusions

Must a trait inference be activated by a causal question ("Why did they do that?") or embedded in a causal frame ("They did that because . . .") for it to be a cause? When we spontaneously (unintentionally and unconsciously) infer that the secretary is clever because she solved the mystery half-way through the book, is clever merely descriptive and not causal? This seems unlikely for several reasons.

First, inferring traits depends on world knowledge, which is often organized in terms of causal schemata. This is clearly shown in theory and research on text comprehension (e.g., Graesser & Bower, 1990; Kintsch & van Dijk, 1978; Schank & Abelson, 1977), especially narrative text (Wyer, 2004) of the kind used in STI work. Such reasoning depends on extensive knowledge of likely and unlikely scenarios, counterfactuals, logical possibilities, social norms, and so on, that is, far more than the few simple factors identified in classical attribution theory. These fall under the "knowledge structures" approach outlined by Hilton (2007). Such knowledge is structured in part by assumptions about what can and might cause what. Traits seem to be one class of concepts that can be used as causes.

Second, STIs show some of the features of intentional impression formation, which is clearly attributional by any definition. These have been documented clearly by Carlson and Skowronski (2005) in their work contrasting STI with STT. Although they do not address the issues of causality or of description vs. attribution, and they frame their concerns in terms of association vs. attribution, there is good evidence that STIs involve more complex processes than STT, identified as associative. So at least relative to simple associations, STIs are "attributional."

Third, impersonal, inanimate nontrait causes can be spontaneously inferred, as shown with the same paradigms used to study STIs (e.g., Hassin et al., 2002). Although this work does not provide a formal definition of what counts as a cause versus a description (the distinction at the core of Hamilton’s (1998) argument), the "causes" inferred in this research hardly seem like mere descriptions. There are few "correspondent inferences" here because the lexicon has few terms that can be used as both causes and descriptions of these events. Pickpockets, food poisoning, and earthquakes were implied causes but they do not describe the events from which they were inferred. Similarly, McKoon and Ratcliff’s (1986) demonstration that predictable events (effects rather than causes) are inferred in the course of text comprehension also illustrates the spontaneous use
of causal schemata in nontrait domains, where there is little to no “correspondence” of lexical terms for describing events and causes.

Fourth, causality is implicit in verbs (Rudolph & Forsterling, 1997). Decades of research on this phenomenon support not only the ubiquity of causal schemata in our understanding of most events, but also the fact that single words (verbs in this case) convey causal relations. That is, causal schemata are part of the meanings of these words. So these serve as existence demonstrations for the possibility that trait words may also convey causality. This is not to say that these are the only meanings that such words can have, or the only ways that they can be used. Nor is it to minimize the complex processes that mediate the translation of experienced events into linguistic descriptions. But it is to say that there is much more to the meanings of words and concepts than is found in a dictionary, and that causal schemata are often central.

Fifth, McCarthy and Skovronski (2011) have shown that STIs are used in making predictions. This would seem to be one of the defining features of causes, and strongly suggests that STIs are causes.

Sixth, isolated personality trait terms behave much like nonsocial causes do, at least in relation recognition and relation priming paradigms. The relation recognition paradigm of Fenker et al. (2005) showed that causal relations between pairs of words, sequentially presented, are identified more quickly if cause precedes effect than if effect precedes cause. Fenker et al. (2005) used nonsocial causes and effects, such as sunshine and freckles. Kressel and Uleman (2010) showed the same thing for traits (causes) and behaviors (effects). Furthermore, Kressel (2011, Study 2) showed that these causal links from traits to behaviors can be activated by lists of nonsocial concepts that are causally linked, thus facilitating lexical decision times without any explicit reference to causality. Study 3 showed that explanatory (diagnostic) relations—relations in which effects precede causes—produce a similar relation recognition asymmetry in which recognition of explanatory relations is faster if behaviors (effects) precede traits (causes). This study also showed that both of these relation recognition asymmetries are correlated with the likelihood of making STIs, independent of explicit theories about traits. Taken together, these findings strongly support the idea that trait concepts, in and of themselves, are causal concepts.

So Hamilton’s (1998) suggestion 15 years ago, that STIs are merely descriptions but not causes, has done what all good theories do. It challenged current thinking with cogent arguments and stimulated new research that produced new insights.

In retrospect, one might wonder how the meaning of a concept can depend on the cognitive processes (as distinct from the explicit questions and their conversational contexts) by which it is activated. Is it reasonable that answering an attributional why question confers a different meaning on the answer than priming the same concept? What other examples are there of two different processes activating the same concept, yet conferring different meanings in each
case? As a counterexample, 5 is 5 whether it is arrived at by adding 2 + 3 or taking the square root of 25 or subtracting 10 from 15. Do we then distinguish the sum 5 from the square root 5 or the difference 5? Does the concept of clever mean something different when activated through STI versus through answering why?

One answer is that, in discourse, the question itself constrains or preordains the meaning. Thus, “Why did that happen?” sets up a different discourse framework from “What is that like?” But an STI has no discourse framework. And the evidence reviewed earlier (esp. McCarthy & Skowronski, 2011; Kressel, 2011) suggests that the dominant or default meaning of traits includes being the causes of behaviors.

What does this imply about Hamilton’s (1998) central argument that Jones and Davis’s correspondent inferences (and perhaps also Gilbert’s characterizations, or Trope’s inferences) are not causal attributions because they are not answers to why questions? One reason to doubt this is that traits’ meanings are inherently, by default, causal. The other reason to doubt this is that when inferences are implicit or spontaneous, they often seem to take place in contexts in which there are also implicit questions such as “Why did this happen?” or “Who is this person?” These implicit questions are part of an implicit discourse that shapes the meanings of concepts, just as it shapes the meanings and impact of primes (Loersch & Payne, 2011). Just what these implicit questions are in any particular context is then an empirical question, assuming that methods are available to tap them. Spontaneous inferences may provide one such method. The spontaneous inferences that occur in particular contexts may provide evidence of what implicit questions occur in that context, for these people in that culture. This suggests an exciting direction for future research.

Meanwhile we have Dave to thank for articulating the distinction between descriptions and causes, which is perhaps uniquely obscure in the case of correspondent inferences. He noticed the distinction and posed his challenging question 15 years ago in such clear and scholarly terms that it could not be ignored. He has done for this for many other interesting phenomena central to our field, as evidenced by many of the chapters in this book. We look forward to 15 more years of the same.

References


SOCIAL PERCEPTION FROM INDIVIDUALS TO GROUPS

Edited by Steven J. Stroessner and Jeffrey W. Sherman