Personality and Social
Psychology Bulletin
1–15
© 2016 by the Society for Personality
and Social Psychology, Inc
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0146167216663704
pspb.sagepub.com
SAGE

Grasping for Traits or Reasons? How People Grapple With Puzzling Social Behaviors

Joanna Korman[1] and Bertram F. Malle[1]

Abstract

Within social psychology, it is well accepted that trait inference is the dominant tool for understanding others' behavior. Outside of social psychology, a different consensus has emerged, namely, that people predominantly explain behavior in terms of mental states. Both positions are based on limited evidence. The trait literature focuses on trait ascriptions to persons, not explanations of behavior. The mental state literature focuses on explanations of ordinary behaviors (for which social scripts provide mental states), not of expectancy-violating behaviors. We examined the critical test case for the two opposing positions: explanations of expectancy-violating behaviors. Participants provided open-ended explanations of puzzling actions, which were content-analyzed for use of mental states, traits, and other causal background factors. Across four studies, three stimulus sets, and two subpopulations, people overwhelmingly offered mental states when explaining puzzling actions (compared with ordinary actions), while they struggled to generate traits and other background factors.

Keywords

causal attribution, trait inference, explanation, mentalizing, theory of mind

Received March 15, 2016; revision accepted July 2, 2016

A long tradition of social-psychological research has emphasized the critical role of stable traits in people's interpretation of others' behavior (Jones & Davis, 1965; Ross & Nisbett, 1991; see Moskowitz & Olcaysoy Okten, 2016, for a review). According to this tradition, trait attribution is the fundamental tool for understanding others' behavior partly because it affords social perceivers efficient and predictively useful knowledge about others (Hastorf, Schneider, & Polefka, 1970; Shaver, 1975). For example, if Jane is seen helping an old woman cross the street, the social perceiver infers that Jane is a *kind* person, or if she works late into the night on her homework, the social perceiver infers that she is *conscientious.* Both of these concise descriptions imply Jane's future behavior—other kind and conscientious acts she may perform.

In the past 20 years, however, a different consensus has emerged, particularly outside of social psychology. As originally suggested by Heider (1958), people indeed focus their explanations of everyday behavior on the *person*. But rather than stable traits, it is a person's mental states—beliefs, goals, and intentions—that constitute the "default" mode of understanding behavior (Malle & Holbrook, 2012; Moskowitz & Olcaysoy Okten, 2016; Reeder 2009). When people explain why a person performed a particular action, their appeal to mental states integrates aspects of the person (e.g., a goal in performing the action) as well as aspects of the situation (what the goal aims to achieve in the world). People make these kinds of mental state inferences spontaneously (Hassin, Aarts, & Ferguson, 2005), and they do so more easily and quickly than they are able to make trait inferences (Malle & Holbrook, 2012; Van Overwalle, Van Duynslaeger, Coomans, & Timmermans, 2012). In their verbal explanations of behavior, too, people primarily infer mental states, not stable traits (Malle, Knobe, & Nelson, 2007; McClure, 2002).

This new consensus crosses numerous disciplines. For half a century, philosophy of action has examined how intentional action is explained by subjective mental states (the agent's reasons) that are at the same time objective causes in the world (Sandis, 2009). Developmental psychology has offered strong evidence for infants inferring goals and preschoolers inferring a variety of other mental states (Bartsch & Wellman, 1995; Gergely, Nádasdy, Csibra, & Bíró, 1995), whereas trait inferences appear later in development (Kalish & Shiverick, 2004; Snodgrass, 1976). Finally, social neuroscience has revealed a network of brain regions specialized for mental

[1]Brown University, Providence, RI, USA

Corresponding Author:
Joanna Korman, Department of Cognitive, Linguistic, & Psychological Sciences, Brown University, Box 1821, 190 Thayer Street, Providence, RI 02912, USA.
Email: Joanna.Korman@gmail.com

state inferences, activated in response to observing human actions (Saxe, Carey, & Kanwisher, 2004).

The "mentalizing" consensus view is so widely held by researchers outside traditional social psychology that some even consider it a foregone conclusion that all behaviors elicit mental states: "Explicit explanations of *any* behavior . . . contain mental state reasons" (Young & Saxe, 2009, p. 1404; emphasis added). But this conclusion is almost certainly overstated, as much research in social psychology shows people's readiness to infer traits (Uleman, Saribay, & Gonzalez, 2008), and people clearly do not explain behavior solely with mental states (Malle, 2011). Furthermore, the mentalizing consensus relies on evidence grounded in a fairly limited scope of behaviors, namely ordinary, expectancy-consistent behaviors that are relatively easy to explain with common beliefs and goals. For example, Van Overwalle et al. (2012) and Malle and Holbrook (2012)—both demonstrating that people make goal inferences more quickly than trait inferences—used ordinary behaviors such as "After paying the bill, she left 5 euros on the table," and "The woman sweeps the floor in the apartment hallway." Similarly, work on people's free-response explanations has typically presented expectancy-consistent behaviors, such as, "Fred, who frequently went to expensive restaurants, went out for a meal at an expensive restaurant with his brother" (McClure & Hilton, 1998, p. 900) or "Why did Anne invite Ben for dinner?" (Malle, Knobe, O'Laughlin, Pearce, & Nelson, 2000, p. 314). These studies confirm that goals and mental states are frequently used for such everyday behaviors. However, the types of behaviors that are most likely to invite why questions or elicit explanatory activity are those that deviate from expectations (e.g., Weiner, 1985; Wong & Weiner, 1981), and there is little evidence that the dominant use of goals and mental states extends to such expectancy-violating behaviors. In fact, Ratcliffe (2007) suggests that the apparent prevalence of mental states as explanations for everyday behavior is due to easy accessibility of scripts and shared norms that imply those mental states. Therefore, when people explain *puzzling* behaviors (ones that violate scripts or norms), they will refer to traits and other causal background factors such as upbringing, habits, and social background—not mental states.

For just these puzzling, expectancy-violating behaviors, theories in the attribution tradition specifically predict that social perceivers are most likely to ascribe traits to an actor—because such behaviors violate "consensus" (Kelley, 1967) or "desirability" (Jones & Davis, 1965) and, thus, are diagnostic of the agent's unique characteristics (Reeder & Brewer, 1979; Sanbonmatsu, Mazur, Behrends, & Moore, 2015; Skowronski & Carlston, 1989).[1] To illustrate, an extreme violation such as robbing a bank is highly diagnostic of a trait such as *dishonesty*, whereas a mild violation such as lying about one's age is not as diagnostic. Likewise in the positive domain, volunteering 1 hr a day in the homeless shelter violates expectations and is highly diagnostic of a trait such as *compassion*, whereas thanking a grocery store cashier for a receipt is quite expected and, therefore, not as diagnostic of a trait such as *friendliness*. Thus, expectancy-violating behaviors are particularly prone to elicit trait attributions.

While current evidence for a hypothesized predominance of mental states is limited to ordinary behaviors, the evidence for a hypothesized predominance of traits is also limited. Evidence for this hypothesis comes from tasks that do not elicit explanations of behavior but invite participants to form an impression of the actor and to use response options restricted to trait descriptors (e.g., Trafimow, Bromgard, Finlay, & Ketelaar, 2005). But trait ascription and behavior explanation are distinct in both function and process (Hilton, Smith, & Kim, 1995; Johnson, Jemmott, & Pettigrew, 1984), and findings from one phenomenon do not necessarily generalize to the other. Whereas an observer's behavior explanation accounts for *why* an actor performed a particular behavior by citing specific characteristics of some person or entity involved in the behavior, trait ascriptions are a form of "belief updating": the observer "learn[s] more about the general characteristics of some person or entity" (Hilton et al., 1995, p. 378). So although traits are dominant in forming an impression of an actor who performs an expectancy-violating behavior, they may or may not be as dominant when it comes to actually explaining the expectancy-violating behavior itself.

When considering the phenomenon of explaining puzzling behaviors, evidence from the mentalizing and trait ascription hypotheses thus suggests competing predictions. Although evidence for the mentalizing hypothesis demonstrates that people explain ordinary behaviors with mental states, the evidence does not currently extend to explanations of *expectancy-violating* behaviors. And, although evidence for the trait hypothesis demonstrates the increased use of traits in the process of trait ascription for expectancy-violating behaviors, the evidence does not currently extend to *explanations* of such behaviors. Therefore, the goal of the present studies is to determine which of these two hypotheses best generalizes to the *explanation* of *expectancy-violating*, truly puzzling behaviors. Alongside the trait versus mental state hypotheses, we will also test a broader variant, according to which traits are only one kind of causal background factor (others being culture, norms, or social context; Malle, 2011), and that together these background factors will trump mental states in explanations of puzzling behaviors (Ratcliffe, 2007).

Methodological Approach

In four studies, we invited participants to provide open-ended explanations of a range of moderately to extremely puzzling behaviors. The studies' use of puzzling behaviors avoids limitations of previous explanation studies in which participants could use scripts and schemas to easily retrieve the agent's goals and beliefs. Instead, by driving participants to the edge of their ordinary explanatory habits, the present studies will reveal whether mental states or traits constitute people's default explanatory tool. Furthermore, the studies' use of open-ended explanations avoids limitations of previous trait ascription studies, which often narrowed potential inferences to a very small, pretested set (Winter & Uleman, 1984). In an open-ended task, participants can select whichever explanations (involving mental states, or traits, or other causes) they feel resolve the puzzle at hand.

For this more even-handed measurement context, we can reformulate the two alternative hypotheses developed above: Hypothesis M predicts that even when trying to explain puzzling behaviors, people will continue to routinely cite the agent's mental states. Hypothesis T predicts that people will increasingly cite explanations involving stable traits or causal background factors more generally. In testing these hypotheses, we focus on explanations of intentional actions, which people are most interested in explaining (Malle & Knobe, 1997) and for which both mental state inferences (Buss, 1978; Malle, 1999; McClure, 2002) and trait inferences (Jones & Davis, 1965; Shaver, 1975) have been claimed to be pivotal.

To sharpen the hypothesis tests, we draw on a theoretical framework of intentional action explanation that ascribes specific functions to such explanatory constructs as mental states, traits, and other causal background factors (Malle, 2004, 2011). According to this theory, people explain intentional actions by citing the agent's mental states (primarily beliefs and desires) as the *reasons* for which the agent acted—that is, the beliefs and desires in light of which and on the grounds of which the agent decided to act. For example, "Ben invited Sarah to dinner because he thought she liked him." This explanation refers to Ben's own reasons, his subjective mental states—he may well be wrong in thinking that Sarah likes him. The second main type of explanation people use to explain intentional behavior is referred to as a causal history of reason (CHR) explanation. Whereas reasons cite the specific mental states on the agent's mind before performing the action, CHRs cite background factors that may have led up to those reasons—factors literally in the causal history of those reasons. Consider a man who goes to the store. One might explain this action by saying that *he wants to pick up three large turkeys for Thanksgiving* (citing a desire reason). Or, one might refer to the background for his reason—the fact that *he has eight sons*. The fact that he has eight sons was not something the agent actively had on his mind but it accounts for why he would want three large turkeys in the first place; thus, the fact lies in the "causal history" of that reason. Such CHR explanations can cite a diverse array of causal background factors, including culture, contexts, and traits. (See supplementary material for additional examples.)

We used these theoretically derived distinctions and the validated classification system that assesses them (Malle, 1998) to measure (a) the frequency of inferred mental states (number of reason explanations), (b) the frequency of causal history explanations, and (c) among causal histories, the frequency of trait explanations.

Hypotheses M and T differ in a double comparison: in the relative number of *reasons versus trait/causal history explanations* in response to *puzzling versus ordinary behaviors*. We considered several designs for the latter comparison. A within-subject design would have drawn undue attention to the difference between puzzling and ordinary behaviors, making any pattern of results vulnerable to conversational demand accounts. A between-subjects design seemed preferable, but there are notable differences in explanation patterns across varying contexts of ordinary behavior explanations (e.g., spontaneous explanations in conversation vs. answers to why questions; experimenter-generated vs. participant-generated behaviors; Malle, 2006). To maximize generalizability across these and other variations, we took advantage of the substantial number of studies previously conducted that assessed explanations of ordinary (non-puzzling) behaviors—in particular, six studies from Malle et al. (2007). In the aggregate, these studies represent the currently most reliable comparison standard of ordinary behavior explanations, with a total sample size of $N = 732$. For example, together with a sample size of $N = 70$ in Study 1, it enabled us to detect effect sizes of $d \geq 0.35$ at $\beta = .2$ and $\alpha = .05$. Statistical power for all other studies is reported in the supplementary material.

Study 1

Method

Stimulus development. Actions may be puzzling in a variety of ways—they may be novel, misfit their context, or violate statistical, social, or moral expectations. We constructed our behaviors to be puzzling with respect to social perceivers' prior knowledge and expectancies about behavior in general, not about the behavior of any single individual. We all but eliminated individual-based expectancies because social perceivers stood in a zero-acquaintance situation with the (fictitious) target agents. Furthermore, we avoided stimuli that strongly violate moral expectations because they would introduce a potential confound, given the legal and everyday importance of mens rea for the moral domain (Cushman, 2008). Among morally largely neutral actions, we represented the variety of puzzling behaviors using four stimulus categories. These categories of puzzles are based on the idea that social scripts (Schank & Abelson, 1977) ground knowledge about the causal and temporal structure of behavioral events and that schemata (Rumelhart & Ortony, 1977) provide the types of agents and objects that are normally co-involved in such events. The first three categories of puzzling behavior derive from cases that break either a script or a schema or both a script and a schema.

Script-breaking, schema-compliant. The first category of stimulus sentences had elements that were semantically associated (schema-compliant) but violated a familiar script.

We created such script violations by altering the expected structure of events (actions denoted by verbs). For example, in the sentence "The garbage men dropped off bags of trash at the end of each driveway," the verb "drop off," though semantically compatible with the job of garbage men, violates the sequence of actions normally completed by garbage men. This sequence violation was achieved by replacing the phrase "pick up" with the phrase "drop off."

Schema-breaking, script-compliant. The second category of stimulus sentences had an intact script but violated a schema. We created schema violations by leaving the expected structure of events intact but replacing a noun denoting an acted-on object with an associatively unrelated noun. For example, in the sentence "He started the car with the ignition key," the phrase "ignition key" was replaced by an associatively distant word, such as "hairbrush."

Script-breaking and schema-breaking. The third category of stimulus sentences contained both script and schema violations. The actions in these sentences neither formed a familiar sequence of events nor did they bear a familiar associative relationship to one another: "The supermarket owner painted her scarecrow magenta." These sentences were constructed by starting with an initial, completely comprehensible sentence and replacing both nouns and verbs in the abovementioned ways.

Script-compliant, schema-compliant: Oversufficiency items. Finally, we created a fourth group of items by leaving both scripts and schemas intact but varying the sufficiency of a particular action to fulfill the goal specified by that action. This group of items includes actions that "overshoot" their apparent motivation. For example, "The vacationers brought six hundred cases of beer to the beach." Although these items should still seem puzzling to participants, they were designed to be less puzzling than the other three categories.

For detailed information on stimulus pretesting and selection, see supplementary material.

Participants. Seventy participants (34 female) completed the study online through Amazon Mechanical Turk in exchange for monetary compensation. One participant was excluded for providing an invalid ID number. The average participant was 30 years old. Fifty-nine percent of participants had completed a 2-year college degree or higher level of education.

Procedure. Participants were instructed that they would be reading a series of sentences. In response to each sentence, they were told to "add whatever sentences or phrases you think are needed to make sense of the sentence." Participants were instructed not to negate information in the original sentences but to "imagine a world in which the sentence is true." After adding their information they answered a follow-up question: "Taking into account the information you added, how much sense does the situation described in the original sentence make now?" Participants rated their responses on a scale from 1 (*no sense at all*) to 8 (*perfect sense*). We included satisfaction ratings to encourage participants to generate genuinely satisfying explanations for the task, but these ratings themselves could not serve as objective measures of explanation quality.

Out of the total of 16 items, each participant responded to four, selected according to a full Latin square design that generated 16 forms. The distribution of the items' strangeness was roughly equal across forms. One item from the schema-breaking category had to be excluded from data treatment and analyses because people's action interpretations and explanations revealed that it was ambiguous with respect to the agent who performed the behavior in question.

Data Treatment

Content coding of explanations. Participants' efforts to make sense of the puzzling stimuli were coded using the Folk Explanations (F.Ex) coding scheme (Malle, 1998, 2004). Among other things, this scheme categorizes explanations into reason explanations and CHR explanations; and within CHR explanations, it distinguishes between stable traits (dispositional properties such as personality, character, or attitudes) and non-traits (such as roles, norms, and culture). Reliability analyses for Study 1 and all subsequent studies are reported in the supplementary material.

Calculating explanation parameters. Using the F.Ex. system, participants' explanations were classified as either expressing a reason (a mental state) or a causal history factor (a non-mentalistic background cause) and, among the latter, a trait or non-trait. For each behavior that a person explained, we counted that person's number of mentioned reasons, causal history factors, and traits. For example, if, for a given behavior, a participant offered two reasons and one causal history (and it was a trait), that participant would have a score of 2 on the reason parameter, a 1 on the causal history parameter, and a 1 on the trait parameter.

Results and Discussion

Participants explained the puzzling behaviors in Study 1 on average with 0.89 reasons (SD = 0.40) per behavior, a rate that did not differ from the average of the six studies that represent our comparison standard, in which people explained ordinary behaviors (M = 0.98, SD = 0.64), Welch's $t(80.75) = 1.51$, $p = .13$, $d = -0.14$, 95% confidence interval [CI] = [−0.41, 0.13]. By contrast, participants offered 0.36 causal histories per behavior, which was lower than in the comparison studies (M = 0.71, SD = 0.75), Welch's $t(79.59) = 5.00$, $p < .001$, $d = -0.47$, 95% CI = [−0.74, −0.20]. To test Hypothesis T, we analyzed people's rate of trait explanations within the causal history category (which makes the test

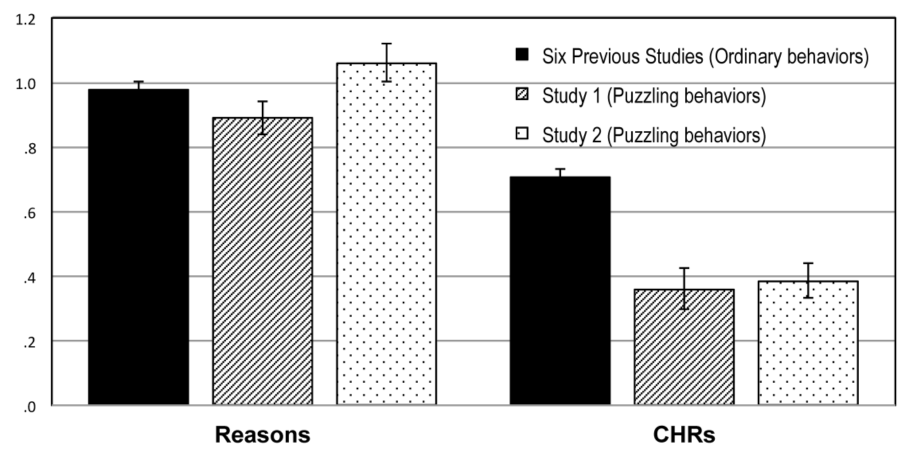

Figure 1. Number of reason explanations and causal history explanations (CHRs) for puzzling behaviors (in Studies 1 and 2) compared with ordinary behaviors (across six previous studies).

orthogonal to that of causal history explanations), including both enduring personality traits and dispositional mental states (e.g., "He has always dreamed of returning to his childhood home"). Participants gave 0.08 traits (and 1.18 non-traits) per causal history explanation, which was considerably smaller than the average across previous studies in which participants had explained ordinary behaviors (M = 0.50), $t(45.10) = 8.06$, $p < .001$, $d = -0.93$, 95% CI = [−1.30, −0.56].

Study 1 provided initial evidence for Hypothesis M—that people routinely search for an agent's reasons (mental states) even when trying to make sense of genuinely puzzling behaviors. People's rate of reason explanations for these puzzling behaviors was as high as the rate of reasons explanations for ordinary behaviors; and instead of resorting to traits (or other background factors), people actually gave fewer of these explanations. The low rate of trait explanations is especially noteworthy in light of the familiar prediction that particularly unusual (expectancy-violating) behaviors should elicit an increase in dispositional inferences (Jones & Davis, 1965; Ross & Nisbett, 1991). However, our participant population consisted of Amazon Turk workers who may be particularly motivated or adept at solving puzzles. We therefore sought to replicate the finding from Study 1 in a sample of participants recruited from the Providence, Rhode Island, community.

Study 2

Method

Participants. Participants were solicited via community advertisement and drawn from an existing database of non-student members of the Providence, Rhode Island, community. A sample of 54 participants (29 women) completed the study online. The average participant was 38 years old, and 61% of participants had completed a 4-year college degree or higher level of education.

Procedure and material. Participants received a URL for the study, whose procedure was identical to that of Study 1. The item that had to be removed from Study 1's analysis was replaced.

Results

Participants offered 1.06 reasons (SD = 0.41) in response to the puzzling behaviors in Study 2, a rate that, as in Study 1, did not differ from comparison studies in which participants explained ordinary behaviors (M = 0.98, SD = 0.64), Welch's $t(62.88) = -1.30$, p = .20, *ns,* d = 0.13, 95% CI = [−0.16, 0.42]. Study 2 participants also used fewer causal history explanations (M = 0.39, SD = 0.37) than did people in previous studies (M = 0.71, SD = 0.75), Welch's $t(75.15)$ = 5.29, $p < .001$, $d = -0.44$, 95% CI = [−0.73, −0.14]. Figure 1 summarizes the results from both Studies 1 and 2, consistently demonstrating that people explain puzzling behaviors, compared with ordinary behaviors, with as many reasons but fewer causal histories.

Traits. Once again, we analyzed trait explanations (per behavior explained by causal histories). As shown in Figure 2, participants offered 0.16 traits (and 1.00 non-traits) when explaining the puzzling behaviors, rates that are considerably lower than those in previous studies, when people explained ordinary behaviors (M = 0.50, SD = 0.46), $t(38.73)$ = 4.98, $p < .001$, $d = -0.76$, 95% CI = [−1.12, −0.39].

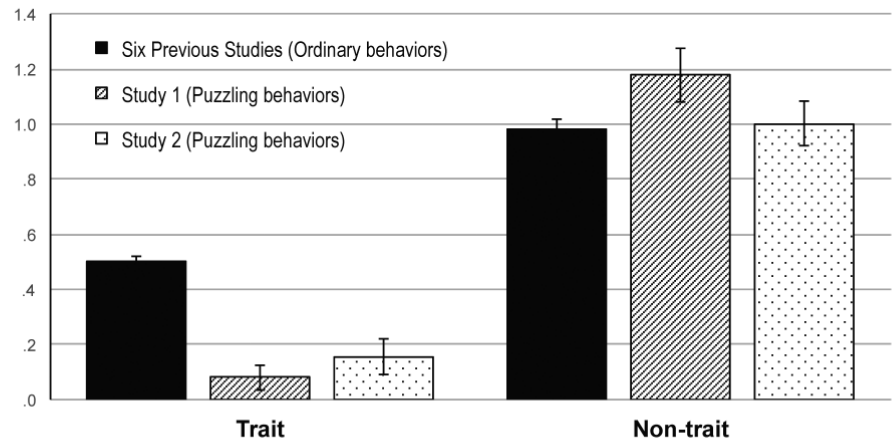

Figure 2. Number of trait (and non-trait) explanations given for puzzling behaviors (Studies 1 and 2) compared with ordinary behaviors (across six previous studies).

Discussion

Study 2 provided further evidence for Hypothesis M—that people routinely use mental state explanations even for puzzling behaviors—this time with a sample of community members. Participants persisted in providing the same number of reasons in response to puzzling behaviors as participants in previous studies did in response to ordinary behaviors. Furthermore, in both Studies 1 and 2, participants provided fewer causal history explanations and fewer trait explanations in response to puzzling behaviors than did previous participants when explaining ordinary behaviors. This result suggests that, when required, participants can and do go beyond their everyday scripts and schemas to actively search for and generate reason explanations, but they are actually less willing or able to offer traits and other causal history explanations when faced with puzzling behaviors. Why might this be?

One possible account is that, lacking a clear knowledge structure for the puzzling stimuli, participants could no longer generate traits and other causal history explanations. This account of knowledge-based generalizations is supported by previous work on explanations of group behavior (O'Laughlin & Malle, 2002). When people explained the behavior of groups for which there is general, stereotyped knowledge available (e.g., members of a particular university's track team), they made ample use of causal history explanations. By contrast, when people explained the behavior of abstract groups about which they had no general knowledge (e.g., "Group A"), they decreased their use of causal history explanations and appealed to reasons instead. A lack of knowledge structures may also have posed specific problems for trait explanations in our first two studies. According to Jones and Davis (1965), inferences about people's dispositions are "correspondent inferences." That is, behaviors are seen as manifestations of more general traits that correspond in content or meaning to the specific behavior (e.g., "John helped the old lady cross the street; he did that because he's generous"). But because the strange behaviors in Studies 1 and 2 do not have a clear meaning, they are not easily understood as manifestations of any particular trait. Thus, participants may have struggled to generate traits.

Study 3

To address the concern that the strangeness of actions in our first studies negated common knowledge structures, we constructed new stimuli for Study 3, ones that are puzzles but nonetheless activate meaningful knowledge structures. Instead of puzzling *actions*, we presented participants with intelligible actions that were paired with puzzling *reasons* for those actions, a stimulus that more closely approximates the type of contextualized puzzles people may encounter in real life. For example, people rarely encounter an action such as "The garbage men dropped off bags of trash at the end of each driveway" in isolation. Instead, they might first see the garbage truck passing by on its usual route and, within this familiar context, notice that the garbage man's behavior seems to pursue an unusual goal. A stimulus sentence such as "The garbage men drove their truck all around the city to drop off a can full of trash at the end of each driveway," provides both familiar context while also presenting a clear puzzle.

In addition, the presence of a puzzling reason explanation—rather than a puzzling action—may invite explainers to explain the reason itself by way of causal history factors, such as stable traits. This prediction is consistent with Jones and Davis (1965), who argue that traits are inferred directly from information about the actor's mental states underlying an action (see also Read, Jones & Miller, 1990). Thus, if people frequently respond to puzzling behaviors with traits, but only when meaningful action or mental state information is available, Study 3 should capture this prerequisite.

Method

Stimulus development. Study 3 stimuli contained an initial clause that described an easily intelligible action (e.g., "The man went to the toy store . . . ") that was paired with a puzzling reason for that action (e.g., " . . . because he wanted to get some watermelons."). Stimulus reasons were either beliefs ("she thought that . . . , " "she knew that . . . , ") or desires ("He wanted to . . . , " "so that he could . . . "). Thus, the puzzles in Study 3 stemmed not from any strangeness of the action itself (as in Study 1) but from the incompatibility between the action and its stated reason. Pretest ratings of such "action-reason incompatibility" provided a graded index of how puzzling each stimulus was. As in Study 1, we constructed four puzzle types, violating either schemas, scripts, or both, or making the action (in light of its reason) seem oversufficient. Sixteen items of each type yielded a total of 64 items. (See supplementary material for a further discussion of stimulus creation and pretesting.)

In the experimental task, each participant was presented with two practice trials and 32 stimulus sentences. From the total 64 items, eight distinct forms of 32 sentences were drawn, each including eight items from each of the four puzzle types, four with belief reasons, and four with desire reasons. Items within each form had similar distributions of action-reason compatibility.

Participants. Forty-one members of the Providence, Rhode Island, community participated in exchange for monetary compensation. They were recruited for study participation via direct telephone solicitation, bulletin board ads, and online advertising. Of the 40 participants who reported demographic information, 58% were female, and the mean age was 36 years. Sixty percent of participants had completed a 4-year college degree, and 95% had completed at least some college. One participant who demonstrated insufficient comprehension of the experimental materials was eliminated from the analyses. A second participant whose accent was difficult to understand in the audio recording was also eliminated.

Procedure. Participants sat at a computer and wore a headset with microphone. The experimenter left the room and the participant read the instructions for the task on the screen. The experimenter then re-entered the room and confirmed the participant's understanding of the instructions. The participant then completed two practice trials and, after the experimenter left the room, began the main task.

In each trial, participants read a stimulus sentence at the top of the screen. After 4 s, a text prompt instructed them to "Add whatever sentences or phrases you think are needed to make sense of the sentence." Participants then had 45 s to provide their additional information by speaking. As in Studies 1 and 2, participants were instructed to "imagine a world in which the sentence is true" and not to change or negate any details of the stimulus sentence but only to add information that helps make the sentence make better sense. After 30 s, they received a 15-s warning to finish up their spoken response for a given item. At the end of each trial, they rated how well the original sentence now made sense in light of their added information (1 = *no sense*, 5 = *some sense*, and 9 = *perfect sense*).

Results

To evaluate the rates of (spoken) explanations for puzzling behaviors, we again aggregated a set of previous studies to serve as a reliable comparison standard for explanations of ordinary behaviors, this time for spoken explanations (Malle et al., 2007).[2] The number of reason explanations for puzzling behaviors in Study 3 ($M = 1.59$, $SD = 0.64$) was indistinguishable from the comparison standard of reason explanations for ordinary behaviors ($M = 1.44$, $SD = 0.68$), Welch's $t(57.46) = -1.32$, $p = .19$, *ns*, $d = 0.23$, 95% CI = [−0.12, 0.57]. Likewise, the number of causal history explanations in Study 3 ($M = 0.72$, $SD = 0.44$) was indistinguishable from the comparison standard of causal history explanations for ordinary behaviors ($M = 0.60$, $SD = 0.54$), Welch's $t(65.15) = -1.52$, $p = .13$, *ns*, $d = 0.23$, 95% CI = [−0.11, 0.58] (see Figure 3).

Traits. Among causal histories, the number of trait explanations ($M = 0.38$, $SD = 0.25$, vs. 1.01 non-traits) for the puzzling behaviors in Study 3 did not differ from the comparison standard of trait explanations for ordinary behaviors ($M = 0.31$, $SD = 0.29$), Welch's $t(78.61) = -1.61$, $p = .11$, $d = 0.28$, 95% CI = [−0.09, 0.65] (see Figure 4).

Having available at least modest numbers of trait explanations, we examined whether the number of offered traits varied as a function of how puzzling the behaviors were (indexed by the action-reason incompatibility ratings). Because the 64 items were distributed across eight forms of 32 items each, not every subject received every item. We therefore averaged the number of trait explanations for a given item across all participants who had received that item. The more strongly incompatible (puzzling) an action-reason pair was, the fewer traits people offered for it ($r = -.25$, $p = .05$).

Discussion

As in Studies 1 and 2, participants in Study 3 persisted in providing reason explanations even for newly developed puzzling behaviors that contained additional information

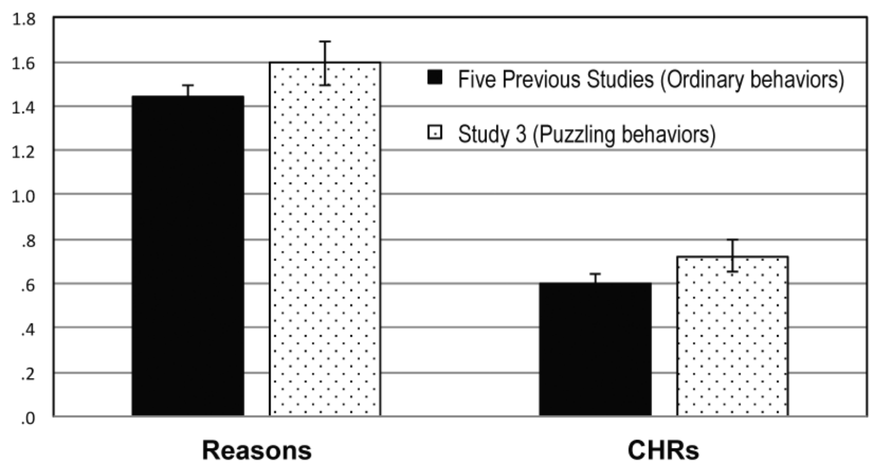

Figure 3. Number of reason explanations and causal history explanations (CHRs) given for puzzling behaviors (Study 3) compared with ordinary behaviors (across five previous studies).

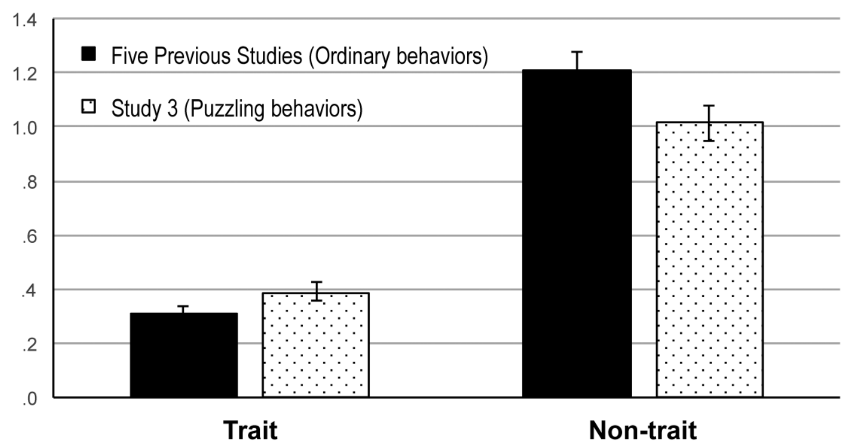

Figure 4. Number of trait (and non-trait) explanations given for puzzling behaviors (Study 3) compared with ordinary behaviors (across five previous studies).

about the agent performing the behaviors. However, participants were able to generate relatively more causal histories than in the first two studies, supporting our conjecture that the single-sentence puzzling actions in Studies 1 and 2 had provided insufficient knowledge structures to afford causal history explanations. Even so, Study 3 participants offered no more causal history explanations than did the comparison samples when explaining ordinary behaviors.

Study 3 participants also increased their trait explanations relative to Studies 1 and 2, but once more, they offered no more traits than people normally do in response to ordinary behaviors. Moreover, the tendency to offer traits was negatively correlated with the degree to which a stimulus sentence was puzzling, thus further weakening the hypothesis that people select trait explanations for the purpose of clarifying puzzling behaviors.

Study 4

In light of Studies 1 through 3, evidence is mounting for the hypothesis that people routinely search for an agent's reasons (mental states) even when trying to make sense of genuinely puzzling behaviors. However, while Study 3 addressed some limitations of Studies 1 and 2, it came with one limitation of its own: the presence of a reason explanation in the stimulus sentence itself. We had designed this type of stimulus (locating the puzzle in the agent's mental state) to provide sufficient information for inferences about causal background and to

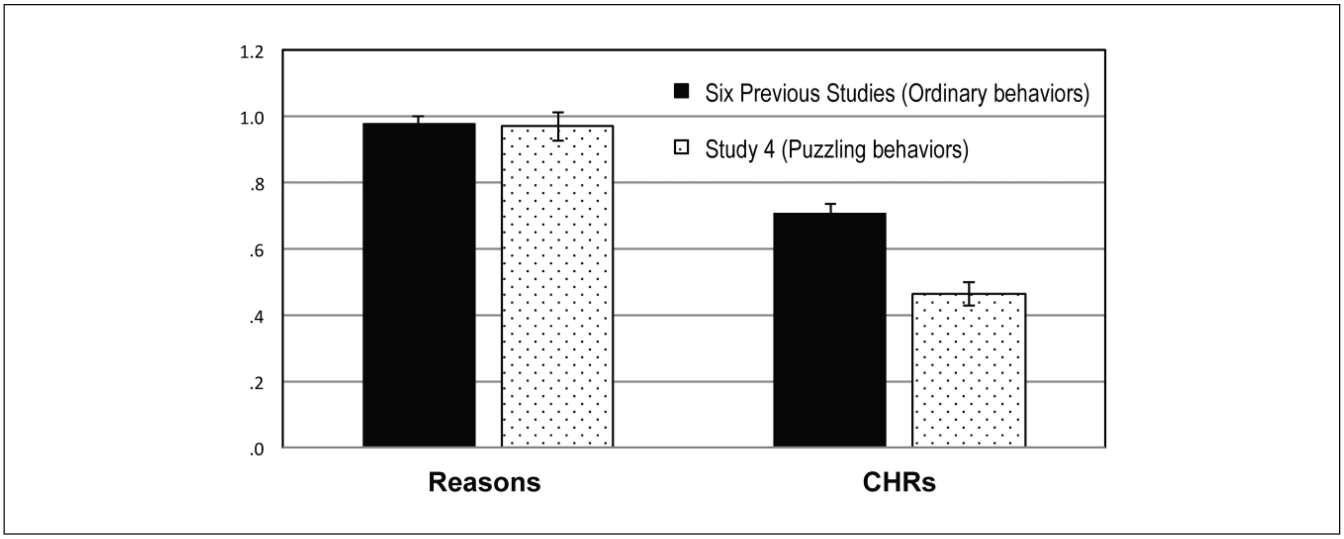

Figure 5. Number of reason explanations and causal history explanations (CHRs) for puzzling behaviors (Study 4) compared with ordinary behaviors (across six previous studies).

perhaps even invite causal history explanations as accounting for the very reason that was puzzling. It is possible, however, that participants were effectively primed to provide more reason explanations because they encountered reasons already present in the stimuli. This account would neither elucidate why reason explanations were frequent already in Studies 1 and 2 (neither of which featured reasons in the stimulus) nor would it elucidate why CHR explanations increased from Studies 1 and 2 to Study 3. But one might suspect that CHR explanations might further increase, and reason explanations decrease, if the stimulus behaviors contained incompatible CHR explanations. To address this possibility, we presented participants in Study 4 with a new set of stimulus sentences that paired actions with puzzling causal history explanations.

Method

Participants. One hundred participants (43 female) completed the study online through Amazon Mechanical Turk in exchange for monetary compensation. Participants' mean age was 35 years, and 60% of them had completed a 2-year college degree or higher level of education.

Stimulus development. In naturally occurring explanations, causal history explanations can appear either by themselves or in tandem with reasons (Malle, 2004). We wanted to represent both variants in this experiment and therefore created two categories of stimulus sentences. In the first, an intelligible action was paired with a single causal history explanation (single-CHR items) that was incompatible with the action. For example, "The journalist shot photos of the crime scene [action] *because he always puts them on his nightstand* [causal history explanation.]" In the second category, an intelligible action was paired with both a reason that was incompatible with the action and a causal history explanation that functioned as a causal history of that particular reason (reason + CHR items). For example, "She went to the plant nursery *[because she wanted] to pick up some video games for her son* [incompatible reason]; *she was an indulgent parent* [causal history of that reason]." See supplementary material for a detailed description of stimulus construction and pretesting.

For the final stimulus pool, we selected four items from the single-CHR category and eight items from the reason + CHR category, including four that were researcher generated (to reflect the theoretical meaning of this explanation type; Malle, 1999) and four that were participant-generated (drawn directly from explanations in Study 3). All reason + CHR items were formulated in two versions: mentioning the reason first or mentioning the causal history first.

Procedure. Instructions and procedure followed those of the previous studies. Each participant responded to one half of the item pool: two single-CHR items and four reason + CHR items (two generated by the researchers, two generated by previous participants), with order of item type counterbalanced across participants.

Results

As shown in Figure 5, the number of reason explanations for puzzling behaviors in Study 4 (M = 0.97, SD = 0.43) did not differ from the number of reason explanations for ordinary behaviors in comparison studies (M = 0.98, SD = 0.64), Welch's $t(161.01)$ = 0.15, p = .88, *ns,* d = −0.01, 95% CI = [−0.22, 0.20]. However, as in Studies 1 and 2, the number of causal history explanations for puzzling behaviors in

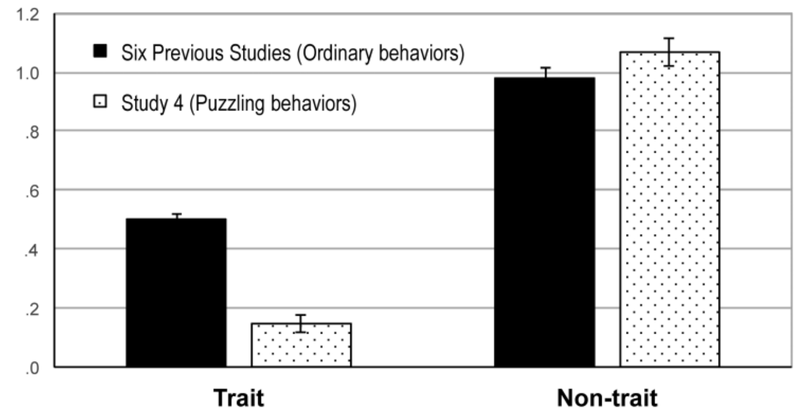

Figure 6. Number of trait (and non-trait) explanations for puzzling behaviors (Study 4) and ordinary behaviors (across six previous studies).

Study 4 ($M = 0.46$, $SD = 0.37$) was lower than the number of causal history explanations for ordinary behaviors in the past ($M = 0.71$, $SD = 0.75$), Welch's $t(227.78) = 5.22$, $p < .001$, $d = -0.34$, 95% CI = [−0.55, −0.12].

Puzzling behaviors with a single causal history explanation and puzzling behaviors with both a reason and a causal history explanation behaved exactly the same, showing lower causal history rates than ordinary behaviors elicited in past studies: for single-CHR items, Welch's $t(162.35) = 4.05$, $p < .001$, $d = -0.31$, and for reason + CHR items, Welch's $t(157.38) = 4.16$, $p < .001$, $d = -0.33$.

As shown in Figure 6, the number of trait explanations for puzzling behaviors in Study 4 ($M = 0.15$, $SD = 0.29$, vs. 1.07 non-dispositions) was less than the corresponding number of trait explanations for ordinary behaviors in comparison studies ($M = 0.50$, $SD = 0.46$), $t(178.21) = 9.15$, $p < .001$, $d = -0.81$, 95% CI = [−1.05, −0.57]. This difference held across both types of puzzling behaviors (single-CHR: $d = -0.72$; reason + CHR: $d = -0.86$).

Discussion

This study addressed the possibility that stimulus behaviors in Study 3 may have primed participants to provide reason explanations simply because each stimulus already presented a reason (that was incompatible with the behavior). We therefore constructed a new stimulus set in which causal histories (alone or in combination with a reason) were presented as part of the puzzling stimulus. But, confirming the patterns of all three previous studies, participants persisted in providing the same number of reasons (mental states) as people in the comparison sample did in response to ordinary behaviors. Moreover, despite being exposed to causal history explanations in the stimuli, participants did not offer any more such explanations in their responses.

In addition, even though the new stimuli were designed to provide information about the agent from which trait inferences could be constructed, the puzzling behaviors in Study 4 elicited fewer trait explanations than did the ordinary behaviors in the comparison sample and fewer trait explanations than the puzzling behaviors in Study 3. This latter difference may derive from the difference between spoken and written explanations. When participants speak their explanations out loud as they did in Study 3, they provide more explanations overall (Malle et al., 2007) and, as a result, they also include somewhat more causal history explanations. In contrast, Study 4 elicited explanations in written form, and causal history explanations returned to their lower rate we had seen in Studies 1 and 2. The impact of communication mode (spoken vs. written) is just a small aspect of a larger phenomenon—the significant role that communicative forces play in shaping explanations (Hilton, 1990). For example, in the present studies the communicative audience was an unfamiliar experimenter, but systematic audience variations can lead to systematic variations in explanation types (Slugoski, Lalljee, Lamb, & Ginsburg, 1993). Whether there are any audiences who would substantially increase causal history explanations (perhaps sociologists or psychoanalysts) is a question for future research.

General Discussion

Across four studies, three unique stimulus sets, and online as well as local community samples, people overwhelmingly persisted in providing reason explanations in the face of puzzling actions, and, contrary to predictions of the classic trait attribution model (e.g., Jones & Davis, 1965), people seem to offer no more traits or causal histories in response to puzzling actions than in response to ordinary actions. In fact, especially in the studies in which participants responded in the written medium, they struggled to generate trait explanations. This surprising pattern of results for trait explanations is illustrated in Figure 7, which places the means for traits from the present four studies in

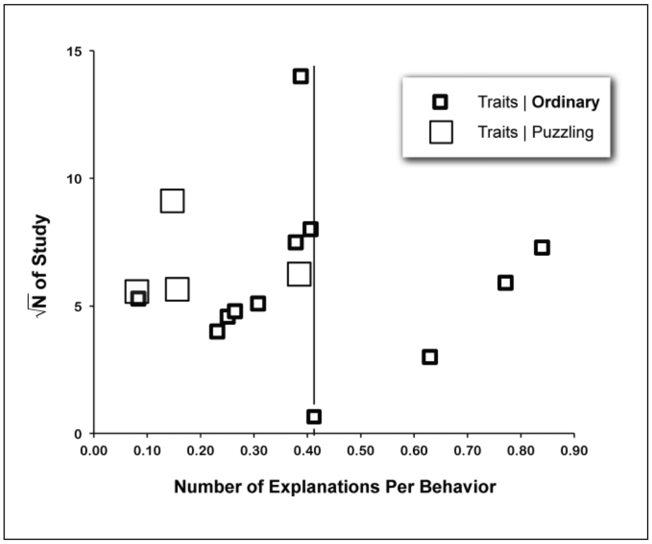

Figure 7. Entries show average number of trait explanations for puzzling behaviors in present four studies and for ordinary behaviors in comparison studies, including five spoken and six written explanation sets.

Note. Following practices in meta-analysis, the *y*-axis indicates a weight for each mean (square root of the sample size on which the mean is based). The aggregated comparison standard (sample-size weighted mean across previous samples) is marked by a vertical line.

the context of averages from past studies that made up our comparison standards. (The supplementary material contains a formal meta-analysis of effect sizes comparing the present with the previous studies, for each of the three explanation types.)

The consistent use of reason explanations even in the face of seriously puzzling behaviors supports Hypothesis M—which claimed that reasons (mental states) are the routine and default explanation mode for any intentional behavior, ordinary or puzzling. By contrast, the results do not support Hypothesis T—the claim that people would explain genuinely puzzling behaviors increasingly with traits (and other causal background factors). It also calls into question its corollary—that people would abandon explanations using mental states when familiar scripts and schemas (as provided by ordinary behavior) are no longer available. In fact, the lower prevalence of trait and causal history explanations in response to puzzling behaviors suggests that it may be these explanations that rely heavily on familiar, general knowledge and are more difficult to generate when scripts and schemas are violated. By contrast, people had no difficulty constructing reason explanations under these informationally sparse conditions.

Frequency and Function of Traits in Behavior Explanations

The infrequent use of dispositional explanations in response to puzzling behaviors suggests that dispositions may be used less commonly than previously thought (here, they made up 7.5% of all explanations across the four studies). In particular, their low prevalence in the face of puzzling behaviors contradicts the common hypothesis that when behaviors deviate from expectations or normality, social perceivers are particularly inclined to make trait inferences (Gilbert & Malone, 1995; Jones & Davis, 1965; Ross & Nisbett, 1991). The literature considers this hypothesis to be very well established (Fiske & Taylor, 1991); how can we reconcile our finding with the previous literature?

We offer one theoretical and one methodological point of reconciliation. The theoretical point is that trait *explanations* are psychologically distinct from trait *ascriptions*. The latter are likely to occur when the perceiver is motivated to form an impression of a target person (Hamilton, 1998) and especially when such an impression is to be communicated to others. Impressions are particularly useful as summary information, and people are prone to use them for broad predictions (e.g., of job performance over an extended time). But such general person impressions, and the trait inferences that constitute them, are not as useful in the domain of action explanation, when a social perceiver is trying to make sense of a particular behavior that someone performed. Making sense of puzzling behaviors is just what people in our studies were asked to do and, as in previous research on ordinary behavior explanations, they explained those puzzling behaviors predominantly with mental states. It is quite possible, of course, that a study with similar stimuli but a task that emphasizes understanding the whole *person*—for example, a processing goal to form an impression (Hamilton, 1998) or anticipate a future interaction (Devine, Sedikides, & Fuhrman, 1989)—could produce a different result: a greater reliance on traits. However, there was nothing in the present task or instructions that *prevented* people from mentioning traits. The rarity with which they did so suggests that behavior explanation and trait ascription are two distinct processes.

The methodological point is that people's trait ascriptions may have been augmented in past studies by selective stimuli (specifically designed to elicit trait inferences; Uleman et al., 2008) and selective response options (rating scales of dispositions only; for example, Jones & Harris, 1967). Under these circumstances, studies document high levels of dispositional attributions, whereas more naturalistic studies (without tailored stimuli or limited response options) find surprisingly low levels of dispositional attributions (Lewis, 1995). The present studies did not rely on stimuli tailored to elicit particular explanations but employed a full range of stimuli: actions alone, actions with reasons, actions with causal histories, and actions with both reasons and causal histories. Moreover, the present studies left response options unconstrained: people offered explanations in their own words (written or spoken), with no instructions except to make sense of the behaviors. Under such conditions, we learned, people tend not to spontaneously provide traits but readily offer mental states.

The Unique Function of Reason Explanations

What can reason explanations accomplish that trait explanations (and causal history explanations more generally) cannot? Explaining another person's behavior by appealing to a causal history explanation such as a trait almost always explains an action by subsuming it under a broader pattern that holds for a type of agent or for a type of action. In contrast, reasons are individuated along three dimensions: They are specific to a particular agent (the agent who possessed the mental states), a particular context (the reasons were on the agent's mind at that time and in that situation), and a particular action (they serve as the rational grounds for that particular action). Using a causal history explanation we might, for example, explain one man's saving a cat from a tree by saying that "he is a firefighter." This serves as an explanation in virtue of the knowledge that *saving cats from trees is something firefighters tend to do.* But as we have seen, when such generalizations are absent or contradictory, people search primarily for mental states to make sense of others' behaviors. In contrast to causal histories, reasons provide information about the particular person performing the particular behavior in the particular context. Rather than answering questions such as, "Why might a person save an animal?" reason explanations answer questions such as "Why did *this person* save *this cat* from the tree *today?*" Perhaps he knew that the cat belonged to his grandmother's friend (reason), and he figured that because it was a slow day at the firehouse, he could use his equipment to help (reason).

Although causal histories derive their explanatory power from invoking familiar generalizations, reasons are more flexible, enabling novel combinations of knowledge about firefighters, grandmothers, a slow day at work, and even the less plausible combinations of social scripts and roles found in our stimuli. For example, in response to the stimulus sentence, "She put on her flip flops because she wanted to improve her cattle-herding skills," one response was, "She figured a pair of comfortable shoes would make her more relaxed and possibly cause injury. [reason] By using flip flops she had to be more aware of her surroundings and take each step carefully. [reason]" Knowledge-based generalizations are not available as tools for the explainer in such a unique, novel case. And, neither is social projection—another common tool to understand other people's behavior (Clement & Krueger, 2000)—that requires some meaningful basis for similarity between the explainer and the agent. The only readily available tool in such a case is a *simulation* of the agent's mind (Nichols & Stich, 2003). In a simulation, the explainer considers the situation in which he himself might *wear flip flops to go cattle herding* and how flip flops might possibly improve his cattle-herding skills if he did. It is only through special consideration of this individuated counterfactual scenario that the explainer is able to find a novel connection between uncomfortable shoes and a challenging cattle-herding experience. Because simulations can easily individuate information along all three dimensions discussed earlier (agent, context, and action), they are the perfect tool to construct reason explanations, especially for puzzling behaviors.

We should caution, however, that our studies have highlighted the types and frequency of explanations people offer in response to puzzling behaviors; they have not addressed the quality of these explanations. In two related studies (Korman & Malle, 2016), we have explored this issue by turning the present participants' explanations into experimental stimuli and presenting them to a new group of lay perceivers. In keeping with the finding that reasons are of fundamental importance, these independent perceivers judged reason explanations as enhancing understanding better than causal history explanations.

Action Explanations and Other Kinds of Explanations

The predominance of particularized explanations for novel, puzzling instances appears to be unique to lay explanations of human action. Both philosophical treatments of scientific explanation (Hempel, 1966; Kitcher, 1989) and empirical evidence from children and adults' explanations of nonbehavioral events (Lombrozo, 2009; Walker, Lombrozo, Legare, & Gopnik, 2014) suggest that explanations that appeal to a general pattern (Williams & Lombrozo, 2010) are the most highly favored. This may be in part due to their function of facilitating predictions about similar phenomena in the future. In everyday life, however, a person's most immediate concern about another person's puzzling behavior is not forming a predictive generalization (the domain of causal history explanations) but understanding a particular present behavior (the domain of reason explanations) and planning the proper response to it.

Simulation, Reasons, and Their Limits

The simulation-based reason-giving our participants engaged in is not an inevitable response to everyday behaviors. As previous research suggests (Epley, Keysar, Van Boven, & Gilovich, 2004; Lin, Keysar, & Epley, 2010), people do not engage in simulation and reason-giving all the time; this would be cognitively expensive and often plain unnecessary (in many cases, scripts and norms do a very fine job). But in response to many significant puzzles they face in their everyday lives, people are able to go beyond their knowledge base and invest in simulation-based mental state inference. So future work is needed to examine exactly when this more cognitively expensive processing is turned on and off and what immediate payoffs it has in dynamic social interaction.

Moreover, some puzzles may activate the simulation machinery without producing reasons, such as excessively altruistic self-sacrifice, suicide, or extreme violence. People struggle to understand, for example, why a family might go so far as to adopt 20 children:

> Some people thought they were saints; but others thought they were publicity-seekers, or weirdos, or had some kind of psychological disorder. Some thought they were addicted to acquiring kids to fill some need, the way others were addicted to shopping. (MacFarquhar, 2015)

All these are causal history explanations—vague generalizations borne of the inability to simulate the agents' actual reasons, as even people who "thought they were saints couldn't understand why they did it" (MacFarquhar, 2015). Likewise, commenting on the perpetrator of the Sandy Hook massacre, Solomon (2014) wrote that even if we discovered that Adam Lanza had suffered from schizophrenia or pedophilia, or had been abused as a child, people "still wouldn't know why he acted as he did"—that is, they would not know his reasons. Nonetheless, finding themselves at the limits of their own explanatory capacity, people reveal their unbending tendency to grasp for *reasons*, even when none are in reach.

Acknowledgment

Thanks to Corey Cusimano, who served as second coder on three studies.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Supplemental Material

The online supplemental material is available at http://pspb.sagepub.com/supplemental.

Notes

1. Throughout the article, we focus on expectancies or priors that social perceivers have about others' behaviors *in general* (e.g., people expect others to be polite or to eat sitting down while at a fancy restaurant). Social perceivers may also have expectancies about a particular individual (e.g., that Jared will behave kindly because he is known to be a kind person). However, if such individual-based expectancies are violated, theories such as Kelley's (1967) predict a decreasing use of traits (the case of "low consistency"). To specifically test the prediction of an increasing use of traits, we focus on violated expectancies about people, contexts, and behaviors (the case of "low consensus").
2. Three of these studies were reported in detail in this article; two of them were included in its meta-analysis.

References

Bartsch, K., & Wellman, H. M. (1995). *Children talk about the mind*. New York, NY: Oxford University Press.

Buss, A. R. (1978). Causes and reasons in attribution theory: A conceptual critique. *Journal of Personality and Social Psychology*, *36*, 1311-1321. doi:10.1037/0022-3514.36.11.1311

Clement, R. W., & Krueger, J. (2000). The primacy of self-referent information in perceptions of social consensus. *British Journal of Social Psychology*, *39*, 279-299. doi:10.1348/014466600164471

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, *108*, 353-380. doi:10.1016/j.cognition.2008.03.006

Devine, P. G., Sedikides, C., & Fuhrman, R. W. (1989). Goals in social information processing: The case of anticipated interaction. *Journal of Personality and Social Psychology*, *56*, 680-690.

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology*, *87*, 327-339. doi:10.1037/0022-3514.87.3.327

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (2nd ed.). New York, NY: McGraw-Hill.

Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, *56*, 165-193. doi:10.1016/0010-0277(95)00661-H

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, *117*, 21-38. doi:10.1037/0033-2909.117.1.21

Hamilton, D. L. (1998). Dispositional and attributional inferences in person perception. In J. M. Darley & J. Cooper (Eds.), *Attribution and social interaction* (pp. 99-114). Washington, DC: American Psychological Association.

Hassin, R. R., Aarts, H., & Ferguson, M. J. (2005). Automatic goal inferences. *Journal of Experimental Social Psychology*, *41*, 129-140. doi:10.1016/j.jesp.2004.06.008

Hastorf, A. H., Schneider, D. J., & Polefka, J. (1970). *Person perception*. Reading, MA: Addison-Wesley.

Heider, F. (1958). *The psychology of interpersonal relations*. New York, NY: Wiley.

Hempel, C. G. (1966). *Philosophy of natural science*. Englewood Cliffs, NJ: Prentice Hall.

Hilton, D. J. (1990). Conversational processes and causal explanation. *Psychological Bulletin*, *107*, 65-81.

Hilton, D. J., Smith, R. H., & Kim, S. H. (1995). Processes of causal explanation and dispositional attribution. *Journal of Personality and Social Psychology*, *68*, 377-387. doi:10.1037/0022-3514.68.3.377

Johnson, J. T., Jemmott, J. B., & Pettigrew, T. F. (1984). Causal attribution and dispositional inference: Evidence of inconsistent judgments. *Journal of Experimental Social Psychology*, *20*, 567-585. doi:10.1016/0022-1031(84)90044-1

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. 219-266). New York, NY: Academic Press.

Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology*, *3*, 1-24. doi:10.1016/0022-1031(67)90034-0

Kalish, C. W., & Shiverick, S. M. (2004). Children's reasoning about norms and traits as motives for behavior. *Cognitive Development*, *19*, 401-416.

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.

Kitcher, P. (1989). Explanatory unification and the causal structure of the world. In P. Kitcher & W. Salmon (Eds.), *Scientific explanation* (Vol. 8, pp. 410-505). Minneapolis: University of Minnesota Press.

Korman, J., & Malle, B. F. (2016). *The folk concept of rationality*. Unpublished data, Brown University.

Lewis, P. T. (1995). A naturalistic test of two fundamental propositions: Correspondence bias and the actor-observer hypothesis. *Journal of Personality*, *63*, 87-111.

Lin, S., Keysar, B., & Epley, N. (2010). Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology*, *46*, 551-556. doi:10.1016/j.jesp.2009.12.019

Lombrozo, T. (2009). Explanation and categorization: How "why?" informs "what?" *Cognition*, *110*, 248-253. doi:10.1016/j.cognition.2008.10.007

MacFarquhar, L. (2015, August 3). The children of strangers. *The New Yorker*. Retrieved from http://www.newyorker.com/magazine/2015/08/03/the-children-of-strangers

Malle, B. F. (1998). *F.Ex: Coding scheme for people's folk explanations of behavior* (Latest version 4.5.7, 2014). Retrieved from http://research.clps.brown.edu/SocCogSci/CodingSchemes.html

Malle, B. F. (1999). How people explain behavior: A new theoretical framework. *Personality and Social Psychology Review*, *3*, 23-48. doi:10.1207/s15327957pspr0301_2

Malle, B. F. (2004). *How the mind explains behavior: Folk explanations, meaning, and social interaction*. Cambridge, MA: MIT Press.

Malle, B. F. (2006). The actor-observer asymmetry in attribution: A (surprising) meta-analysis. *Psychological Bulletin*, *132*, 895-919. doi:10.1037/0033-2909.132.6.895

Malle, B. F. (2011). Time to give up the dogmas of attribution: A new theory of behavior explanation. In M. P. Zanna & J. M. Olson (Eds.), *Advances in experimental social psychology* (Vol. 44, pp. 297-352). San Diego, CA: Academic Press.

Malle, B. F., & Holbrook, J. (2012). Is there a hierarchy of social inferences? The likelihood and speed of inferring intentionality, mind, and personality. *Journal of Personality and Social Psychology*, *102*, 661-684. doi:10.1037/a0026790

Malle, B. F., & Knobe, J. (1997). Which behaviors do people explain? A basic actor-observer asymmetry. *Journal of Personality and Social Psychology*, *72*, 288-304. doi:10.1037/0022-3514.72.2.288

Malle, B. F., Knobe, J., & Nelson, S. E. (2007). Actor-observer asymmetries in explanations of behavior: New answers to an old question. *Journal of Personality and Social Psychology*, *93*, 491-514. doi:10.1037/0022-3514.93.4.491

Malle, B. F., Knobe, J., O'Laughlin, M. J., Pearce, G. E., & Nelson, S. E. (2000). Conceptual structure and social functions of behavior explanations: Beyond person-situation attributions. *Journal of Personality and Social Psychology*, *79*, 309-326. doi:10.1037/0022-3514.79.3.309

McClure, J. (2002). Goal-based explanations of actions and outcomes. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 12, pp. 201-235). New York, NY: Wiley.

McClure, J., & Hilton, D. J. (1998). Are goals or preconditions better explanations? It depends on the question. *European Journal of Social Psychology*, *28*, 897-911.

Moskowitz, G. B., & Olcaysoy Okten, I. (2016). Spontaneous goal inference (SGI). *Social & Personality Psychology Compass*, *9*, 644-661.

Nichols, S., & Stich, S. P. (2003). *Mindreading: An integrated account of pretence, self-awareness, and understanding other minds*. New York, NY: Oxford University Press.

O'Laughlin, M. J., & Malle, B. F. (2002). How people explain actions performed by groups and individuals. *Journal of Personality and Social Psychology*, *82*, 33-48. doi:10.1037/0022-3514.82.1.33

Ratcliffe, M. (2007). *Rethinking commonsense psychology: A critique of folk psychology, theory of mind and simulation*. New York, NY: Palgrave Macmillan.

Read, S. J., Jones, D. K., & Miller, L. C. (1990). Traits as goal-based categories: The importance of goals in the coherence of dispositional categories. *Journal of Personality and Social Psychology*, *58*, 1048-1061. doi:10.1037/0022-3514.58.6.1048

Reeder, G. D. (2009). Mindreading: Judgments about intentionality and motives in dispositional inference. *Psychological Inquiry*, *20*, 1-18. doi:10.1080/10478400802615744

Reeder, G. D., & Brewer, M. B. (1979). A schematic model of dispositional attribution in interpersonal perception. *Psychological Review*, *86*, 61-79. doi:10.1037/0033-295X.86.1.61

Ross, L., & Nisbett, R. E. (1991). *The person and the situation*. New York, NY: McGraw-Hill.

Rumelhart, D. E., & Ortony, A. (1977). The representation of knowledge in memory. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the acquisition of knowledge* (pp. 99-135). Hillsdale, NJ: Lawrence Erlbaum.

Sanbonmatsu, D. M., Mazur, D., Behrends, A. A., & Moore, S. M. (2015). The role of the baserate frequency of correspondent behavior and trait stereotypes in attribution: Building on Rothbart and Park (1986). *Social Cognition*, *33*, 255-283. doi:10.1521/soco.2015.33.4.255

Sandis, C. (Ed.). (2009). *New essays on the explanation of action*. New York, NY: Palgrave Macmillan.

Saxe, R., Carey, S., & Kanwisher, N. (2004). Understanding other minds: Linking developmental psychology and functional neuroimaging. *Annual Review of Psychology*, *55*, 87-124. doi:10.1146/annurev.psych.55.090902.142044

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, NJ: Lawrence Erlbaum.

Shaver, K. G. (1975). *An introduction to attribution processes*. Cambridge, MA: Winthrop.

Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin*, *105*, 131-142. doi:10.1037/0033-2909.105.1.131

Slugoski, B. R., Lalljee, M., Lamb, R., & Ginsburg, G. P. (1993). Attribution in conversational context: Effect of mutual knowledge on explanation-giving. *European Journal of Social Psychology*, *23*, 219-238. doi:10.1002/ejsp.2420230302

Snodgrass, S. R. (1976). The development of trait inference. *Journal of Genetic Psychology*, *128*, 163-172.

Solomon, A. (2014, March 17). The reckoning. *The New Yorker* Retrieved from http://www.newyorker.com/magazine/2014/03/17/the-reckoning

Trafimow, D., Bromgard, I. K., Finlay, K. A., & Ketelaar, T. (2005). The role of affect in determining the attributional weight of immoral behaviors. *Personality and Social Psychology Bulletin*, *31*, 935-948.

Uleman, J. S., Saribay, S. A., & Gonzalez, C. M. (2008). Spontaneous inferences, implicit impressions, and implicit theories. *Annual Review of Psychology*, *59*, 329-360. doi:10.1146/annurev.psych.59.103006.093707

Van Overwalle, F., Van Duynslaeger, M., Coomans, D., & Timmermans, B. (2012). Spontaneous goal inferences are often inferred faster than spontaneous trait inferences. *Journal of Experimental Social Psychology*, *48*, 13-18. doi:10.1016/j.jesp.2011.06.016

Walker, C. M., Lombrozo, T., Legare, C. H., & Gopnik, A. (2014). Explaining prompts children to privilege inductively rich properties. *Cognition*, *133*, 343-357. doi:10.1016/j.cognition.2014.07.008

Weiner, B. (1985). "Spontaneous" causal thinking. *Psychological Bulletin*, *97*, 74-87.

Williams, J. J., & Lombrozo, T. (2010). The role of explanation in discovery and generalization: Evidence from category learning. *Cognitive Science*, *34*, 776-806. doi:10.1111/j.1551-6709.2010.01113.x

Winter, L., & Uleman, J. S. (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. *Journal of Personality and Social Psychology*, *47*, 237-252. doi:10.1037/0022-3514.47.2.237

Wong, P. T. P., & Weiner, B. (1981). When people ask "why" questions, and the heuristics of attributional search. *Journal of Personality and Social Psychology*, *40*, 650-663.

Young, L., & Saxe, R. (2009). An fMRI investigation of spontaneous mental state inference for moral judgment. *Journal of Cognitive Neuroscience*, *21*, 1396-1405.