

The texture of causal construals: Domain-specific biases shape causal inferences from discourse

Brent Strickland¹ · Ike Silver² · Frank C. Keil²

Published online: 2 December 2016
© Psychonomic Society, Inc. 2016

Abstract We conducted five sets of experiments asking whether psychological and physical events are construed in broadly different manners concerning the underlying textures of their causes. In Experiments 1a–1d, we found a robust tendency to estimate fewer causes (but not effects) for psychological than for physical events; Experiment 2 showed a similar pattern of results when participants were asked to generate hypothetical causes and effects; Experiment 3 revealed a greater tendency to ascribe linear chains of causes (but not effects) to physical events; Experiment 4 showed that the expectation of linear chains was related to intuitions about deterministic processes; and Experiment 5 showed that simply framing a given ambiguous event in psychological versus physical terms is sufficient to induce changes in the patterns of causal inferences. Adults therefore consistently show a tendency to think about psychological and physical events as being embedded in different kinds of causal structures.

Keywords Domain specificity · Causal reasoning · Event representation

People of all ages ubiquitously infer causal structure from observation. For example, in a classic case, Dr. John Snow was able to show, from multiple observations of cholera spreading through London in 1854, that its transmission was caused by contaminated drinking water. Similarly, in the

psychological domain, people infer causal structure from everyday observations. For example, upon learning that a manager of Manchester United recently decided to invest 36 million pounds in an unproven 19-year-old player, fans could infer that the decision to buy at this particularly high price was brought about by several factors: a positive evaluation of the player’s potential, a long-term strategy to build for future (as opposed to immediate) success, a desire to secure the transfer before a deadline, and the selling club’s reluctance to sell for a lower price.

Just how such causal inferences are achieved has been an area of intense investigation in cognitive psychology. The general view of causal induction (i.e., the postulation of causal relationships from observed data) adopted here resembles the “theory-based causal induction” view developed by Griffiths and Tenenbaum (2009). According to this view, at the computational level of analysis (Marr, 1982), causal induction can be seen as the product of “domain-general statistical inference guided by domain-specific prior knowledge” (Griffiths & Tenenbaum, 2009, p. 661).

The first facet of the theory involves the tracking of statistical information in a domain-general way, and the incorporation of such statistical information into causal inferences. Evidence for such an ability is robust. In making causal inferences without much background knowledge concerning the particular types of entities involved, people are able to analyze the frequency with which a cause and a potential effect co-occur, and systematically deploy such information to form judgments about causal relationships (Cheng, 1997; Jenkins & Ward, 1965; Shanks, 1995) in a manner that may be supported by a process resembling Bayesian analysis (Sobel, Tenenbaum, & Gopnik, 2004; Tenenbaum & Griffiths, 2001). To successfully differentiate causation from mere correlation, people also apply a number of domain-general strategies or heuristics in support of such inferences. These include

✉ Brent Strickland
stricklandbrent@gmail.com

¹ Département d’Études Cognitives and Institut Jean Nicod, UMR 8129, Pavillon Jardin, École Normale Supérieure, CNRS, PSL University, 29 rue d’Ulm, F-75005 Paris, France

² Yale University, New Haven, CT, USA

intervention (e.g., Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003), reasoning about temporal sequences (e.g., reasoning that if x generally occurs before y , then x is a likely cause of y ; Lagnado & Sloman, 2006), and reasoning about abrupt transitions (e.g., whether an initial change in X co-occurs more often with a change in Y , or vice versa; Rottman & Keil 2012; Rottman, Kominsky, & Keil, 2014).

The second facet of the theory involves the influence of domain-specific prior knowledge on how statistical information is used to draw causal conclusions. Thus, infants and adults categorize objects into types based on their causal properties (Gopnik, Sobel, Schulz, & Glymour, 2001; Tenenbaum & Niyogi, 2003), and these resulting categories can carry with them expectations about the strength of causal relationships (Kemp, Goodman, & Tenenbaum, 2007). Moreover, prior expectations about mechanisms have been postulated to constrain causal inferences from otherwise identical covariational data (e.g., Ahn & Kalish, 2000; Schlottmann, 1999). To give just one example, when presented with identical covariational data regarding (temporally asynchronous) color changes in two balls, participants who were told that the two balls were connected by a hidden wire were more likely to judge that one ball changed the color of the other than were participants who did not receive information about the mechanism (Wolff, Ritter, & Holmes, 2014).

In the present article, we investigate the possibility of a type of domain-specific bias that may be at play in causal reasoning: biases in the relative numbers of causes that would be postulated to bring about physical versus psychological events. Our focus is less on specific structures, such as common cause, common effect, and feedback loops, and more on what we might call “metastructural” expectations—namely, biases about the relative density and broad types of causal patterns associated with different ontological classes of entities. From a broader point of view, such structural expectations could serve to constrain domain-general processes of causal induction across a wide range of contexts.

Domain specificity in psychological versus physical reasoning

There are strong reasons to suspect that the psychological and physical ontological domains may differ with respect to people’s broad expectations regarding the density and types of causal patterns that serve to bring about events in these different domains.

From early in development, humans have structured expectations regarding how psychological states come about and influence behavior. This is sometimes known as “theory of mind.” Thus, infants expect social agents (but not physical objects) to have not only goals that help organize their actions (Hamlin, Hallinan, & Woodward, 2008; Olineck & Poulin-

Dubois, 2005; Woodward, 1998), but also desires (Repacholi & Gopnik, 1997), beliefs (Onishi & Baillargeon, 2005), and rational means–end reasoning (Gergely & Csibra, 2003). Additionally, young infants possess the ability to reason about emotional states, and are thus sensitive to the congruence between emotions and actions (Hepach & Westermann, 2013).

Infants’ expectations about the behavior of physical objects contrast with their expectations regarding social agents possessing psychological states. Whereas they know that social agents can intentionally create order, they do not expect this of inanimate physical objects (Newman, Keil, Kuhlmeier, & Wynn, 2010). They understand that for an entity to cause a physical object (but not a mindful social agent) to move, that entity must come into direct contact with the object (Leslie & Keeble, 1987). And they understand that physical objects, in contrast to social agents, are not capable of goal-driven self-propelled motion (Saxe, Tenenbaum, & Carey, 2005; Spelke, Phillips, & Woodward, 1995).

Consistent with these differences in infancy, one also finds important domain contrasts in reasoning about physical versus psychological events at later stages of development and into adulthood. As Paul Bloom (2006, pp. 211–212) puts it:

People universally think of human consciousness as separate from the physical realm. Just about everyone believes, for instance, that when our bodies die, we will survive—perhaps rising to heaven, entering another body, or coming to occupy some spirit world. And just about everyone believes in free will. At both a phenomenological level and an intellectual level, we experience ourselves as free agents. While our bodies are physical, and can be affected by physical things, we have choice.

Accordingly, young toddlers explicitly distinguish certain types of psychological states from brain-based, physical actions. So, for example, young children accept that certain actions, such as solving a math problem, require a brain, whereas other psychological states/events, such as loving one’s brother or pretending to be a kangaroo, do not (Bloom, 2004).

Although the above evidence suggests a deep divide between the psychological and physical domains for the purposes of causal reasoning, such a divide does not necessarily show that people will have broadly different metastructural expectations about the relative numbers and types of causes that serve to bring about events (which is at issue in the present article). This more precise prediction derives from studies on adult and childhood reasoning.

Specifically, naive ascriptions of free will suggest that the causes leading up to our current mental states and actions are multiple and nondeterministic (Nichols, 2004). On the other hand, physical events may intuitively be expected to stem from linear, deterministic processes and to result from a

handful of easily identifiable causal difference makers (Danks, 2007; Strevens, 2008). For example, when presented with a description of a specific event, young children by around 5 years of age expect that human agents but not physical objects are free and “could have done otherwise” (Nichols, 2004). In related work, Walsh and Byrne (2007) had adults reason about alternatives in reason–action sequences (e.g., pulling into a lane and missing a turn when trying to avoid heavy traffic) or cause–effect sequences typically not involving salient psychological states as causes (e.g., traffic being diverted onto a different route because of a fallen tree). The authors found that people tended to think “if only” about actions in a reason–action sequence, but tended to think “if only” about causes in a cause–effect sequence. This result is compatible with an interpretation whereby participants tacitly believe that (physical) causes inevitably lead to their effects (and thus, to change the outcome, one must change the cause). However participants may infer that the relationship between reasons and actions is not determinate in the same way.

Thus, taken together the above findings are suggestive that we may intuitively believe that the causes responsible for mental states and actions are multiple and nondeterministic, whereas the causes for physical events may intuitively be expected to stem from more simple linear and deterministic processes.

Present experiments

Here we conducted five sets of experiments asking whether psychological and physical events are construed differently in terms of the patterning of their causal inputs. We predicted that people would attribute relatively fewer causes to physical than to psychological events, because physical causal chains are more likely to be construed as simplistic, linear, and deterministic, whereas psychological causal chains (e.g., those leading to mental states, thoughts, etc.) are more likely to be thought of as complex and nondeterministic. Although the definition of the physical events used here was straightforward (i.e., events in which no social agent was involved), our operational definition of “psychological events” merits clarification. We focused mainly on changes in mental states as opposed to the taking of intentional actions, which has been the focus of other work (e.g., looking at reason–action sequences, in Walsh & Byrne, 2007). Nevertheless, the changes in psychological states studied here sometimes did heavily imply an action (e.g., making a decision to do x), and thus intersected with the previous areas of study.

In addition to testing for effects of domain on inferences about causes and causal structure, we also tested for domain differences in the numbers of estimated effects that people predicted would result from particular events. Although we had no structured hypotheses concerning these results, their

inclusion potentially served as an informative comparison condition for two reasons. First, it would provide information with respect to whether any observed effects were specific to reasoning about causes or might reflect more general biases that would extend to other types of judgments. Second, this comparison could prove informative to the growing literature examining potential asymmetries in how people reason about causes versus effects (Ahn & Nosek, 1998; Fernbach, Darlow, & Sloman, 2010; Waldmann & Holyoak, 1992).

Experiment 1a

Our participants viewed descriptions of simple events that were either psychological (e.g., “A teacher becomes depressed.”) or physical (e.g., “A house burns down.”) in nature. The physical events were designed to involve an inanimate object undergoing a change of state. The psychological events involved a human being undergoing a purely psychological change of state (e.g., a teacher becoming depressed), undergoing a change of mental state that implied an accompanying action (e.g., a politician changing her mind about a policy), or performing an action that strongly implied an underlying psychological state (e.g., bursting into tears),

Participants were then asked either to estimate the total number of effects that would follow from the event or to estimate the total number of causes that led to the event.

Method

Participants A total of 18 adults were recruited with Amazon’s Mechanical Turk and were compensated a token amount. The number of participants was matched with that of pilot results and served to fix to within one or two participants the number of participants for Experiments 1b–1d below (different sample sizes occurred due to differences between the numbers of participants requested and the numbers of responses actually received). For Experiments 1–5, the sample sizes were decided in advance, and optional stopping was never employed. In cases in which we received more or fewer participants than we requested via Mechanical Turk, we always report all available data.

Workers were restricted to those having a 95 % or higher hit approval rate and coming from the United States. Compensation amounts were similar across all experiments (1a–5) reported here, varying between \$.15 and \$.4. The selection criteria of a 95 % or higher hit approval rate and being in the US were also constant across all experiments. We did not collect demographic information for the workers from the particular studies carried out here, but previous large-scale analyses (see Mason & Suri, 2012, for a review) have shown that the majority of US respondents are female (55 % vs. 45 %

male), with an average age of 32, and earn roughly USD 30,000 per year.

Design The experiment had a 2×2 design with both conceptual domain (physical or psychological) and estimation type (causes and effects) as repeated measures.

Materials and procedure Each participant read ten different sentences online. Five referred to simple physical events, and five referred to simple psychological events (as described above). See the [Appendix](#) for a full list of the stimuli and the by-item means for Experiments 1a–1d.

Participants were asked to estimate on a scale of 0–100 how many specific things were likely either to have caused the event (“cause” condition) or to result from the event (“effect” condition). The cause and effect questions were presented in blocks (with a block containing only cause or only effect questions), and the order of presentation of the blocks was randomized between participants. Within each block, the order of all items was randomized.

Results

Each participant’s average estimation was calculated for the following four conditions: physical cause, psychological cause, physical effect, and psychological effect. A repeated measures 2×2 analysis of variance (ANOVA) revealed a significant interaction between conceptual domain and estimation type [$F(1, 17) = 15.42, p = .001, \eta_p^2 = .48$].¹ In a first within-participants two-tailed planned contrast, we found that the estimated number of causes for the physical items was fewer than the estimated number of causes for the psychological items (27.87 vs. 48.96) [$t(17) = 5.17, p < .001, \eta_p^2 = .61$]. The estimated numbers of effects, however, failed to differ between conditions (38.77 vs. 38.37) [$t(17) = .07, p = .95, \eta_p^2 = .00$].

Discussion

The results of Experiment 1a support the hypothesis that participants show important domain differences in causal estimation. In particular, the results suggest a specific pattern whereby the number of estimated causes is systematically lower for physical than for psychological events, whereas the estimated number of effects is not.

¹ A power analysis revealed that at 18 participants, we had a statistical power of .97 to detect this interaction.

Experiment 1b

In Experiment 1a, we picked a relatively arbitrary scale (0–100) for our estimation task. It is unlikely that people would be able to actually generate on the order of 20–40 causes for a given event, instead our account suggests that these estimations reflect broad domain-specific expectations about the relative densities of causes in physical versus psychological events. We would hypothesize that such expectations could be modified to fit various contexts.

To address this, in Experiment 1b we asked whether the results above would generalize to a new scale (i.e., 0–10 instead of 0–100). The underlying idea was that by specifying a scale (0–10 vs. 0–100), participants can likely adjust the level of granularity of their causal expectations to fit this scale. For example, a participant might implicitly or explicitly reason that if 100 is the maximum value of the scale, then he or she should be thinking about relatively fine-grained causes or effects (of which there would be many). If 10 were the maximum, however, they should be thinking about relatively coarse-grained causes or effects.

If the results from Experiment 1a were due to metastructural expectations about the relative density (as opposed to the brute number) of causes associated with different ontological classes of entities, then one would expect that the effects from Experiment 1a should replicate, regardless of how coarse- or fine-grained the implied causal structure was. Thus, we would predict that those effects would replicate even on a 0–10 scale.

Method

Participants A total of 18 adults participated using Amazon’s Mechanical Turk.

Design, materials, and procedure Experiment 1b was identical to Experiment 1a, except that the scale was changed from 0–100 to 0–10.

Results

A repeated measures 2×2 ANOVA again revealed a significant interaction between conceptual domain and estimation type [$F(1, 17) = 5.24, p = .035, \eta_p^2 = .24$]. In a first within-participants two-tailed planned contrast, we found that the estimated number of causes for physical items was fewer than the estimated number of causes for the psychological items (3.71 vs. 4.79) [$t(17) = 2.13, p = .048, \eta_p^2 = .21$]. The estimated numbers of effects, however, failed to differ between conditions (6.2 vs. 5.8) [$t(17) = 1.18, p = .26, \eta_p^2 = .08$].

Discussion

These results are consistent with the view that the domain-specific asymmetries in estimating the numbers of causes versus effects are due to differences in metastructural expectations pertaining to the relative numbers of causes in the physical versus psychological domains. One possible account of these results is that the scale (i.e., 0–100 vs. 0–10) sets the participants' expectations about the size and scale of causes and effects that are relevant for the task, and an expectation about the number of causes is then applied to the relevant scale according to the domain. This suggests that any estimation effects found here (and in other experiments) would be more likely to reflect broad expectations about causal density, which can be adapted to context, than to reflect something about the specific causes that a person brought to mind when diagnosing various events.

Experiment 1c

Experiment 1c extended the findings above by testing a new stimulus set (on a scale of 0–100). The primary goal was to ensure that the effects observed above were not a consequence of the particular stimuli that we chose to study and were likely to be robustly generalizable.

We created new sets of physical as well as psychological items. Both stimulus sets were subdivided into complex and simple events. For the physical items, the simple events involved the functioning of a single artifact (e.g., “A computer starts.”). The complex events involved a natural weather phenomenon (e.g., “A hurricane formed.”), which typically covers a wider physical area than that covered in an event related to a physical artifact and typically involves one or many physical substances, such as water, air, or lava (as in “A volcano erupted.”). For the psychological items, the simple events were changes in the psychological state of a single individual (e.g., “A professor changes his mind.”), whereas the complex psychological events involved changes in the psychological states of organizations of individuals (e.g., “A corporation becomes interested in making computers.”)

We again predicted fewer estimated causes for the physical than for the psychological events, for both simple and complex events. On the basis of the previous results, we expected that the estimated number of effects would not be less for physical than for psychological events.

Method

Participants A total of 19 adults participated using Amazon's Mechanical Turk (18 requested).

Design, materials, and procedure Experiment 1c was identical to Experiment 1a, except that it employed ten psychological events and ten physical events. None of the sentences had appeared in the previous experiments.

Results

An initial 2 (physical vs. psychological conceptual domain) \times 2 (complex vs. simple) \times 2 (cause vs. effect estimation) repeated measures ANOVA revealed a number of findings.

It first revealed a main effect of complexity [$F(1, 18) = 45.77, p < .001, \eta_p^2 = .21$], with simple events receiving fewer overall estimated causes/effects than did complex events (23.41 vs. 44.44). This factor failed to interact with conceptual domain [$F(1, 18) < 1, p = .37, \eta_p^2 = .05$], and there was no three-way interaction between conceptual domain, judgment type, and complexity [$F(1, 18) < 1, p = .91, \eta_p^2 = .001$]. These results thus served as a validation of our manipulation of complexity, certifying that complex events were indeed perceived as being more complex (in terms of their causes and effects) and that the (perceived) differences between complex and simple events were similar across the physical and psychological domains.

We again replicated the two-way interaction (found in the other experiments) between conceptual domain and judgment type [$F(1, 18) = 15.25, p = .001, \eta_p^2 = .001$]. A first planned contrast between simple physical and psychological events revealed a significant difference in the numbers of estimated causes (12.94 vs. 40.25) [$t(18) = 3.52, p = .002, \eta_p^2 = .41$]. A second planned contrast revealed a significant difference in the numbers of estimated causes for complex physical versus psychological events (29.41 vs. 52.95) [$t(18) = 4.89, p < .001, \eta_p^2 = .571$].

However, the simple physical versus psychological events did not differ significantly with respect to their estimated effects (17.01 vs. 23.45) [$t(18) = 1.8, p = .09, \eta_p^2 = .15$], nor did the complex events (46.99 vs. 48.40) [$t(18) = 0.26, p = .80, \eta_p^2 = .004$].

Discussion

Experiment 1c suggested that the effects discovered in Experiments 1a–1b were broadly generalizable, given that they replicated in an entirely new stimulus set and held across differing levels of baseline complexity.

Experiment 1d

Although the results in Experiments 1a–1c were consistent with our predictions, unwitting experimenter bias in creating our stimuli might still have unfairly weighted the results in favor of our predictions. Effects of biasing by knowledgeable

experimenters in stimulus creation have been demonstrated in similar online contexts (Strickland & Suben, 2012), and we were eager to avoid any such limitations in the present studies. To address this concern, we had online participants who were blind to our hypotheses first create sentences referring to physical versus psychological events, and then tested those stimuli in the estimation task.

Method

Participants

Stimulus creation A group of ten adults were recruited using Amazon's Mechanical Turk.

Main experiment A group of 20 adults were again recruited using Amazon's Mechanical Turk (18 requested).

Design, materials, and procedure

Stimulus creation Participants were asked to create five sentences referring to physical events and five sentences referring to psychological events. The participants received the following verbatim instructions (for the physical vs. psychological conditions):

Physical sentence generation:

Please write in 5 different PHYSICAL sentences. All of these sentences must have a PHYSICAL object as the grammatical subject of the sentence, and must describe a purely physical event, which are characterized by a change in the physical world. An example sentence would be "A volcano erupts." Another example sentence would be "An airplane lands." Note that none of the sentences may have a person or an animal as their grammatical subject. So a sentence like "A thirsty man drinks water" would not be acceptable because it has the word "man" as the grammatical subject. Similarly, a sentence like "A small dog barks" would be unacceptable because it has an animal as the grammatical subject.

Psychological sentence generation:

Please write in 5 different PSYCHOLOGICAL sentences. All of these sentences must have a person as the grammatical subject of the sentence, and must describe some psychological event, which is characterized by a change to a person's mental states. An example sentence would be "A person decides to believe in God." Another example sentence would be "A criminal decides to be a better person."

It was decided in advance that we needed 20 stimuli from the physical domain and 20 from the psychological domain for our main experiment. We wanted to increase the overall number of

items being tested while still allowing the experiment to be completed in a reasonable amount of time by online participants. Given that we were unsure how well participants would be able to generate stimuli for the task, we decided to be cautious by overestimating the number of total stimuli that we received. Thus, we gathered ~100 participant-generated stimuli, of which we planned to use 40.

To decide which 40 to use for the main task, the underlying goal was to get a fair spread across the participants. Thus, we numbered the physical (and psychological) stimuli such that the first participant's first item would be labeled "1," the second participant's first item "2," and so forth. Once we reached the tenth participant, the second item produced by the first participant would be labeled "11," the second item produced by the second participant "12," and so forth. However, any stimulus that did not conform to the instructions was eliminated in this process (e.g., the sentence "The lion roars" was eliminated as a physical item because it violated the rule about not having an animal as a grammatical subject). We then simply selected Stimuli 1–20 from each conceptual domain. This procedure generated a total of 40 items, with at least three items selected from each participant.

Main experiment Experiment 1d was identical to Experiment 1a, with the exception that there were now 40 total items (20 physical and 20 psychological).

Results

The results of Experiment 1d broadly replicated the pattern of results found in Experiment 1a. The interaction between conceptual domain and estimation type was again significant [$F(1, 19) = 24.82, p < .001, \eta_p^2 = .57$]. A first planned contrast revealed that the estimated number of causes for physical items was lower than the estimated number of causes for the psychological items (20.47 vs. 32.33) [$t(19) = 2.16, p = .04, \eta_p^2 = .20$]. The estimated number of physical effects was actually greater than the estimated number of psychological effects (34.89 vs. 26.44) [$t(19) = 2.54, p = .02, \eta_p^2 = .25$].

Discussion

The pattern of results found in Experiment 1a was replicated, even in a stimulus set generated by a set of participants blind to the present hypothesis. Thus, in Experiment 1d the number of estimated causes was again systematically lower for physical than for psychological events, whereas the estimated number of effects was not. The stimuli used in generating these results were unlikely to have been influenced by unwitting experimenter bias, and likely also had the advantage of being ecologically valid in the sense that they were representative of the types of events that people would naturally consider in their everyday lives.

Experiment 2

The goal of our present research is to test for potential domain-specific expectations that may apply across a range of experimental contexts and dependent variables. Toward that goal, Experiment 2 asks whether the pattern of results we previously observed applies beyond our basic estimation task. This time, we asked participants to actively produce hypothetical causes and effects for various events and tested whether the same domain-biasing effect would result.

Method

Participants A total of 42 adults were recruited using Amazon's Mechanical Turk (40 requested). We approximately doubled the standard sample size from Experiments 1a–1d, given that we were unsure how participants would perform in this task and how many participants would complete the task (given that it was more time-intensive and more demanding).

Of these participants, nine failed to complete more than 75 % of the survey and were thus excluded from all further analyses.

Design, materials, and procedure Experiment 2 was identical to Experiment 1b, except that participants were asked to list (instead of estimate) as many causes versus effects as possible for the event. Participants were provided with 12 blank slots for each item in which to enter their responses. We deliberately restricted the pragmatically relevant set of causes/effects to 12 blank slots, primarily to ensure that the task would be doable in a short amount of time (given that we expected any domain-specific expectations to be relative as opposed to scale-specific, this change did not affect our ability to ask the primary theoretical question of interest).

Results

Each participant's average number of responses was calculated for each of the experimental conditions. The results mirrored those found in Experiments 1a–1d. The interaction between conceptual domain (i.e., physical vs. psychological) and judgment type (i.e., cause vs. effect) was significant [$F(1, 32) = 31.85, p < .001, \eta_p^2 = .50$].

Planned contrasts revealed that participants generated fewer hypothetical causes for physical than for psychological events (2.50 vs. 3.65) [$t(32) = 6.22, p < .001, \eta_p^2 = .55$]. On the other hand, participants generated roughly equal numbers of effects for physical versus psychological events (2.31 vs. 2.43) [$t(32) = 1.44, p = .16, \eta_p^2 = .06$].

Redundancy ratings We also wished to ensure that the results above were not driven by redundancy in the responses. For example, perhaps participants generated more psychological

than physical causes, but the psychological causes were mostly redundant. This would mean that the numbers of *different* causes listed might not differ between the domains.

Thus, three independent coders were instructed to rate each participant's responses for redundancy. First we explained how the basic task worked, and then we instructed them as follows:

We are trying to understand whether participants listed any “redundant” answers. That is, we want to know whether any row contains multiple causes (or effects) that are exactly the same in meaning. We are asking that you read each row left to right and count the number of redundant causes or effects contained in that row only (for many rows of answers, this number may well be zero). Please consider answers to be redundant only if *they are the same in meaning*. Answers that are only *similar* in structure or meaning should *not* be considered redundant unless they are by and large the same in meaning.

Each rater provided a number of responses equivalent to the total number of items that were shown to participants (2,640 responses per rater). The average pairwise percentage agreement between and raters was 95.81 %, ranging between 96.17 % and 95.42 %. Thus, the raters showed a high level of agreement.

For each experimental participant, we averaged across the raters to compute the percentage of redundant responses that the relevant experimental participant provided for each of the four experimental categories. We then averaged across participants to compute a mean percentage of redundant responses. These were as follows: physical diagnosis = 4.31 %, psychological diagnosis = 3.23 %, physical prediction = 3.28 %, psychological prediction = 3.99 %. A 2×2 repeated measures ANOVA revealed no main effect of conceptual domain [$F(1, 32) < 1, p = .72, \eta_p^2 = .004$]. Similarly, we observed no main effect of judgment type [$F(1, 32) < 1, p = .94, \eta_p^2 = .00$] and no significant interaction between conceptual domain and judgment type [$F(1, 32) = 3.40, p = .08, \eta_p^2 = .096$]. Finally, t tests mirroring the planned contrasts from the main experiment revealed a lack of significant differences between the conditions. Thus there was no significant difference between the physical and psychological diagnoses [$t(32) = 1.28, p = .21, \eta_p^2 = .05$], and there was also no significant difference between the physical and psychological predictions [$t(32) = 1.34, p = .19, \eta_p^2 = .05$].

Discussion

The results of Experiment 2 supported the conclusion that the asymmetries in causal reasoning between physical and psychological events are due to general cognitive tendencies that

apply not only in the specific estimation tasks studied in Experiments 1a–1d, but also in a different task type involving a different dependent variable (i.e., the production of causes and effects).

Experiment 3

In Experiment 3, we examined in detail one particular factor that may (at least partially) account for the reduced numbers of estimated and imagined causes in physical events: a greater expectation in the physical domain of simple linear causal chains, as opposed to multiple converging factors combining to bring about an outcome.

Method

Participants A total of 42 adults were recruited using Amazon’s Mechanical Turk (40 requested). We were again uncertain how participants would respond to this task, and thus conservatively opted to (approximately) double the sample size relative to Experiment 1d (i.e., the experiment that was most similar to the present experiment).

Design, materials, and procedure Experiment 3 was identical to Experiment 1d, except that participants were asked to choose the diagram that best illustrated the causes or the effects of the event. One diagram presented a linear chain, whereas the other presented multiple converging causes/multiple diverging effects (see Fig. 1).

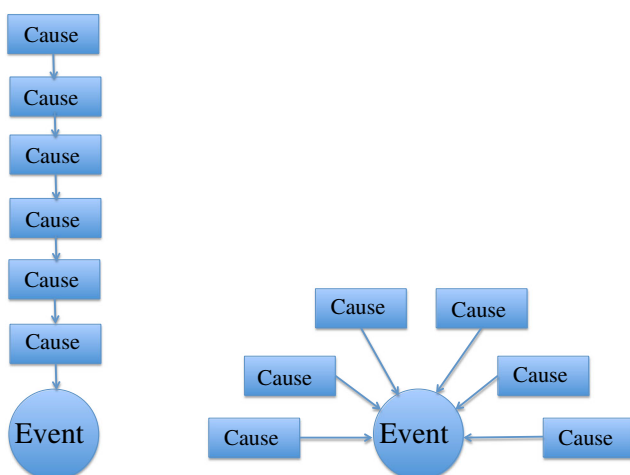


Fig. 1 For each event, the participants in the “cause” condition were presented with a choice between a linear causal structure and a causal structure depicting multiple converging causes. The participants in the “effect” condition saw similar structures, except that the directionality of the causal arrows in the diagrams was reversed

Results

Each participant’s percentage of linear choices was calculated for each of the four experimental categories. As in our previous experiments, the interaction between conceptual domain and judgment type was marginally significant [$F(1, 41) = 3.87, p = .056, \eta_p^2 = .06$].

Participants displayed a significant preference for linear causes in physical relative to psychological events (64.86 % vs. 38.90 %) [$t(41) = 4.21, p < .001, \eta_p^2 = .3$]. They also showed a significant preference for linear effects in physical relative to psychological events (54.14 % vs. 38.84 %) [$t(41) = 2.57, p = .01, \eta_p^2 = .14$], but this preference was smaller than in the cause condition, thus creating the (marginally) significant interaction.

Discussion

The results of Experiment 3 supported the hypothesis that people have a greater expectation for linear causal chains in the identification of causes for physical as opposed to psychological events. Even though participants had no practice or training in matching such diagrams to events, they showed a differentiation in the kinds of diagrams that applied to physical and psychological events.

Experiment 4

In Experiment 4, we further probed the relative preferences for linear chain causality in physical versus psychological events, by asking whether such linear chains are related to imputed deterministic processes. Here we employed participants’ conditional probability judgments (i.e., the probability of an effect given a cause) after participants had chosen a diagram depicting either linear or multiple converging causes for a given event.

Method

Participants A total of 214 adults were recruited (200 requested) using Amazon’s Mechanical Turk, allowing for approximately ten participants per passage per condition.

Design The present experiment was based on a two-factor design (psychological vs. physical), with two dependent variables (choice of causal structure and conditional probability judgment).

Materials and procedure Participants were randomly assigned to read about a single event, taken from a list of 20 possible events. The list of possible events was composed of descriptions of ten physical events and ten psychological

events taken from a subset of the items generated in Experiment 1d above (see the asterisked items in the Appendix for Exp. 1d). Upon reading their assigned event stimulus, participants were first asked to choose the diagram that best illustrated the causes of the event. The diagrams used were identical to those employed in the “cause” condition from Experiment 3. Thus, one diagram presented a linear chain, whereas the other presented a diagram of multiple converging causes (in a manner identical to the “cause” condition in Exp. 3). After making their choice, participants were then presented with a new diagram matching their preferred causal diagram. For example, if the participant had previously chosen a linear option, that participant then saw a new linear diagram that was identical to the previous one, with the exception that one of the nodes was highlighted in red (selected at random). The participant was asked to indicate the probability that the event would occur given the presence of this cause.

Results

Each item’s percentage of linear choices was calculated for each of the two experimental categories. We performed a by-item analysis, as opposed to a by-participant analysis (as we had done in the previous experiments), because each participant saw only a single item, thus making it impossible to average individual participants’ means for individual conditions. An independent-samples *t* test revealed that, as in Experiment 3, the percentages of linear choices differed significantly between the physical and psychological domains (51.33 % vs. 22.86 %) [$t(18) = 3.53, p = .002, \eta_p^2 = .41$].

For each item, we then calculated the average conditional estimated probability (without distinguishing multiple converging from linear causal types). An independent-samples *t* test revealed that the estimated probability of an effect given a physical cause was seen as being higher than the estimated probability of an effect given a psychological cause (50.55 vs. 36.30) [$t(18) = 4.12, p = .001, \eta_p^2 = .49$].

For each conceptual domain, we separately calculated the average linear estimated conditional probability as well as the average converging-causes estimated conditional probability. The means were as follows: physical/linear = 55.69, physical/converging = 41.32, psychological/linear = 46.91, psychological/converging = 33.37. A 2×2 ANOVA with Judgment Type as a within-items factor (one item from the psychological domain was excluded from this analysis because it received no linear responses) and Conceptual Domain as a between-items factor revealed a nonsignificant interaction [$F(1, 17) < 1, p = .87, \eta_p^2 = .002$].

There was, however, a main effect of causal type, whereby the conditional probabilities in the linear chains were judged to be higher than those for the converging causes (51.3 vs. 37.66) [$F(1, 17) = 10.21, p = .005, \eta_p^2 = .38$].

Discussion

The results of Experiment 4 replicated and extended those from Experiment 3, by showing a relative preference for linear causal chains leading to an event in the physical as compared to the psychological domain. These results also yielded two further insights into the mechanisms of domain-specific causal reasoning. First, linear chains are conceived of in more deterministic terms, with individual causal nodes having more power to bring about a given effect (i.e., there is a higher estimated probability of an effect, given a cause). Second, averaged across causal schemas (i.e., linear or converging), the physical domain is considered to be more deterministic than the psychological domain. The fact that physical events are more readily associated with simple, deterministic chains may play a role in reducing the expected number of causes for physical events (observed in Exps. 1a, 1b, 1d, 2, and 5).

Experiment 5

Experiments 1–4 suggested differing expectations in causal reasoning for the physical and psychological domains. One might therefore expect that simply framing an ambiguous phenomenon as being physical versus psychological would bring about significant changes in reasoning about its causes versus effects. In this manner, the same phenomenon could be construed quite differently when it was immersed in a different set of inferred causal structures. Experiment 5 tested this prediction.

Experiment 5 also addressed a minor design issue present in Experiments 1a–1d. Whereas in those experiments estimation type (cause or effect) was varied within participants (thus introducing the possibility that one type of judgment might influence the other), here we eliminated this possibility by varying estimation type between participants.

Method

Participants A total of 192 adults were recruited (200 requested) using Amazon’s Mechanical Turk for the primary experiment. This requested sample size was set on the basis of a pilot experiment with a similar sample size (but that tested only a single item).

Design The experiment was based on a 2×2 mixed design with conceptual domain (physical vs. psychological) as a repeated measure and estimation type (causes and effects) as a between-subjects variable.

Materials and procedure Participants were shown a series of ten texts like the following:

Consider the phenomenon of having low self-esteem. Modern research has begun to show that that low self-esteem is a purely PHYSICAL phenomenon. That is, having low self-esteem is really just a PHYSICAL process in the brain. Despite the fact that many people think of low self-esteem as being inherently psychological, most research shows that this isn't the case at all. Now imagine that someone you know has low self-esteem. Given that low self-esteem is a physical phenomenon in the brain, how many specific things do you think are likely to have CAUSED their low self-esteem?

Roughly half the participants (random assignment) were asked to estimate on a scale of 0–100 how many specific things were likely to have caused the event (“cause” condition, exemplified above), whereas the other half were asked to estimate how many specific effects were likely to result from the event (“effect” condition).

Our items consisted of mental conditions that could plausibly be conceptualized as being either inherently physical (i.e., brain-based) or psychological. These were low self-esteem, political conservatism, anxiety, bulimia, depression, obsessive compulsive disorder, antisocial personality disorder, anorexia, compulsive gambling, and posttraumatic stress disorder.

For each participant, exactly half of the items were described as being inherently physical (i.e., brain-based) phenomena, as in the example above. The other half of the items were described as being inherently psychological (i.e., mind-based) phenomena.

We pseudorandomized the particular pairings of which five items were described as physical and which were described as psychological by randomly generating two separate lists. On the first list, the following items were physical: low self-esteem, political conservatism, anxiety, bulimia, depression, and obsessive compulsive disorder. The rest were psychological. On the second list, this was reversed. These lists were identical across the cause and effect conditions. Participants were randomly assigned to one of two lists. All items were presented in a randomized order to participants.

Results

The individual participant averages were calculated for each of the experimental conditions. We first tested for effects of list (from the pseudorandomization) and observed no main effects or significant interactions with either conceptual domain (psychological vs. physical) or estimation type (cause vs. effect). We thus collapsed across lists for all further analyses.

A 2×2 ANOVA with Judgment Type as a between-participants factor and Domain Framing as a within-participants factor revealed a significant interaction [$F(1, 190) = 6.88, p = .009, \eta_p^2 = .035$].

This interaction was driven by a pattern of results that was similar to those in the previous experiments. Two planned contrasts revealed that the estimated number of causes was significantly lower for physical than for psychological events (31.89 vs. 39.80) [$t(96) = 3.38, p = .001, \eta_p^2 = .11$]. On the other hand, the estimated numbers of effects failed to differ significantly between physical and psychological events (47.77 vs. 48.16) [$t(94) = 1.08, p = .24, \eta_p^2 = .001$].

Discussion

Experiment 5 supports the hypothesis that differing tendencies in reasoning about causes (from a given effect) may be due to biases that are specific to the cognitive domains, whereas such biases in reasoning about effects (from causes) are not present. Thus, people will estimate different causal densities for relatively ambiguous but well-known phenomenon depending on whether they are framed as being inherently physical or psychological.

General discussion

Whether it is simply estimating numbers of causes and effects (Exps. 1a–1d and 5), listing hypothetical causes and effects (Exp. 2), matching abstract causal structures depicted in diagrams to events (Exps. 3 and 4), or generating conditional probabilities (Exp. 4), adults consistently think about psychological and physical events as being embedded in different kinds of causal structures. This tendency is so strong that it is found even when the same well-known phenomenon is simply framed in psychological versus physical terms (Exp. 5).

Domain specificity

In particular, when estimating the number of things that have caused a given effect, participants consistently estimated that a lower number of causes were likely to have brought about physical than psychological events. However, no such domain-specific effects consistently held for the estimation of effects from a given cause, thus suggesting an asymmetry in diagnostic reasoning (i.e., reasoning from effects to causes) versus predictive reasoning (reasoning from causes to effects), consistent with other such observations (Ahn & Nosek, 1998; Fernbach et al., 2010; Waldmann & Holyoak, 1992). This asymmetry also suggests that the domain effects found here are specific to causal estimation and do not reflect a general response bias (e.g., to indicate lower numbers for physical events) that would be obtained in any type of judgment task.

Our results also point to a potential mechanism explaining the decreased estimates of causes for physical versus psychological events: Causes of physical events are more likely to be

conceptualized as deterministic and simple linear causal chains than are the causes of psychological events. On the other hand, psychological events are more likely than physical events to be seen as resulting from multiple, converging causes in a nondeterministic fashion (possibly due to naive intuitions about free will that are associated with psychological events). It may be that these domain-specific qualitative differences in the complexity of the imputed causal structures translate into differing quantitative estimates of the numbers of likely causes.

This result is compatible with previous work suggesting that people conceive of relationships between nonpsychological causes and effects differently than those between psychologically imbued reason–action sequences (Walsh & Byrne, 2007). People may have different default expectations in the two domains because the causal relationship between reasons and actions is not typically as stable as the causal relationship between non-psychologically-driven causes and effects (Juhos, Quelhas, & Byrne, 2015; Walsh & Byrne, 2007). Thus, a single action may be thought of as resulting from multiple causes, as when one drives down a street to achieve multiple goals (e.g., going to the grocery store, picking up children from work, and dropping off something at a post office). On the other hand, a person may perform multiple actions to achieve a single goal. For example, to become a better athlete, one may lift weights, train more often, run, swim, and read books. But Byrne and colleagues hypothesized that these “many-to-one and one-to-many mappings of reasons to actions are uncharacteristic of causal relations, which tend to have a simpler one-to-one mapping of causes to effects” (Juhos et al., 2015, p. 58).

Our work may provide further insight into this theoretical perspective. First, it suggests that the many-to-one and one-to-many mappings found in previous work may not simply be inherent to reason–action sequences. Instead, such mappings may extend further to cover a far larger range of psychological event types (including but not limited to the causes and effects of reasons and emotions). Second, our results suggest that in the physical domain (which consists only of cause–effect sequences), the simplicity of mappings is asymmetric. Although there does appear to be a simpler inferred mapping from effects to causes than in the psychological domain, there seems to be no systematic difference between psychology and physics when reasoning from causes to effects.

There are, however, some important limitations to note with regard to our general conclusions. First, at best, our results would show that imputed linearity and determinism (associated with physical events) are correlated with lower estimates for causes. Such a correlation, even if established, would not demonstrate the further point that linear

assumptions actually produce a reduction in causes. More work would be needed to show this.

Second, there are question marks regarding the specific interpretations of Experiments 3 and 4. In Experiment 4, participants gave higher conditional probability ratings for causes leading to physical events than for causes leading to psychological events. This may reflect, as we suggested, a greater sense of determinism in the physical than in the psychological domain. Alternatively, however, people may have interpreted our request to estimate the probability of the effect given the cause as a request to estimate the probability of the effect given *only* the cause (known as a “causal power judgment”; Cheng, 1997). This possibility would be in line with recent work by Cummins (2014a), showing that people indeed have a general tendency to misinterpret the test question in this way. In this case, our results would be indicative of domain-specific biases regarding causal power that may be partially or entirely independent of intuitions regarding determinism.

Relatedly, it is well established that in diagnostic inference (i.e., reasoning from effects to causes), alternative causes spontaneously come to mind, whereas in predictive causal inference (i.e., reasoning from causes to effects) disablers (i.e., causes that might prevent the event in question from occurring) spontaneously come to mind (Byrne, 1989; Cummins, 2014b; Cummins, Lubart, Alksnis, & Rist, 1991; Markovits, 1986). This could not straightforwardly explain the interaction between judgment type and conceptual domain in Experiment 3 or the differences between physical and psychological diagnoses observed in Experiments 3 and 4. Nevertheless, conceptual domain may interact with the activation of spontaneous causes and disablers, and this may account for some of the variance observed here. Follow-up studies examining this possibility could prove informative.

Computational role

What role do these construal biases play in causal induction? As we described earlier, causal induction can be viewed as a domain-general process supplemented by domain-specific biases (Griffiths & Tenenbaum, 2009). Future research might address the interaction between these two facets of causal reasoning. For example, the bias to attribute fewer causes to physical than to psychological events may guide information search, choices of intervention, or evaluation of alternative hypotheses in tasks that more directly look at causal induction.

The present findings may also connect with the literature on the “illusion of explanatory depth” (Mills & Keil, 2004; Rozenblit & Keil, 2002), in which participants initially rate their understanding of causal mechanisms as

being much greater than it actually is. This is a fact that they recognize upon reading a subsequent expert explanation. The present task of estimating the number of causes is similar to the assessment of one's mechanical understanding in the IOED paradigm, thus raising the intriguing possibility that participants might show a greater illusion of explanatory depth for psychological than for physical events if participants' biases do not map cleanly onto the actual causal structure. On the other hand, if participants' biases do track some element of true causal structure, then one might expect to find equivalently large illusions of explanatory depth in both domains.²

It is not surprising that people think about their social and physical worlds differently. It is, however, much more remarkable that across a wide range of social and physical events, people have sharply contrasting expectations about the causal structures in which social and physical agents are typically embedded, even as they do not appear to have any explicit awareness of these contrasts.

Author note This research was supported in part by the National Institute of Health's Grant No. R37-HD023922 to F.C.K. It has also benefited from funding from the European Research Council under the European Union's Seventh Framework Programme (Grant No. FP/2007-2013)/ERC Grant Agreement 313610 (PI: Emmanuel Chemla), IDEX Grant Nos. ANR-10-IDEX-0001-02 PSL and ANR-10-LABX-0087 IEC, and a Fyssen Foundation postdoctoral grant.

Appendix: By-item breakdown of stimuli for Experiments 1a–1d

The first number gives the average number of estimated causes, and the second number the average number of estimated effects. All averages are rounded to the nearest whole number.

Experiment 1a

Physical

- A house burns down. (27/42)
- A window breaks. (26/23)
- An airplane explodes. (34/61)
- A car starts. (17/30)
- A fire is ignited. (35/38)

Psychological

- A teacher becomes depressed. (50/29)
- A man decides to leave his family. (40/55)

- A person is surprised. (51/30)
- A politician changes her mind about a policy. (44/45)
- A woman bursts into tears. (59/33)

Experiment 1b

Physical

- A house burns down. (4.55/7.83)
- A window breaks. (2.83/3.72)
- An airplane explodes. (4.5/8.44)
- A car starts. (2.94/4.39)
- A fire is ignited. (3.72/6.61)

Psychological

- A teacher becomes depressed. (5.44/6.56)
- A man decides to leave his family. (5.56/8.16)
- A person is surprised. (4.06/4.06)
- A politician changes her mind about a policy. (4.72/5.78)
- A woman bursts into tears. (4.17/4.44)

Experiment 1c

Physical

- Simple (artifact)** A computer starts. (31/62)
- A light is turned on. (26/40)
- An airplane lands. (28/52)
- A cigarette is lit. (26/66)
- A bottle is opened. (33/33)

Complex (nonartifact) There is a gust of wind. (45/67)

- It begins to rain. (33/72)
- A hurricane forms. (47/75)
- A volcano erupts. (27/66)
- A tidal wave occurs. (38/78)

Psychological

Simple (individual) A person decides to believe in God. (61/57)

- A lady becomes happy. (60/46)
- A professor changes his mind. (42/51)
- A criminal decides to be a better person. (49/45)
- A man suddenly feels lonely. (53/44)

Complex (organization) One country decides to attack another. (66/77)

- A corporation becomes interested in making computers. (51/57)
- A military unit forms a plan to siege a castle. (55/56)

² We thank an anonymous reviewer for this suggestion.

A police department makes up its mind to arrest a senator. (37/51)

A basketball team becomes motivated. (53/42)

Experiment 1d (items from Experiment 4 with asterisk)

Physical

A radio makes a noise. (14/20)

*A car crashes. (32/43)

A wave crashes. (13/16)

*Rain falls. (16/57)

*The earth spins. (32/65)

A cloud floats. (19/10)

*A building collapses. (31/64)

A volleyball gets hit. (6/8)

*A ship sinks. (32/47)

*An earthquake shakes. (25/65)

The rockets separate from the space capsule. (30/29)

The rock hits the tree. (11/9)

*Lightning strikes the ground. (16/25)

A pebble falls down a waterfall. (13/7)

*The sun shines. (22/74)

*The ground opened up. (35/58)

The faucet leaks. (7.3/11)

*A landslide falls. (28/53)

Water falls off the cliff. (16/24)

The ball rolls. (13/12)

Psychological

*The woman cries. (40/17)

A person decides to believe in reason. (29/46)

A man becomes happy. (38/40)

*A child becomes frustrated. (38/28)

A man feels elated. (33/29)

A man became angry. (39/30)

*A nurse felt crestfallen. (27/15)

*A mother becomes exhausted. (43/31)

*He missed his wife. (26/15)

*My nephew has become depressed and angry. (38/29)

A teacher becomes less depressed. (26/26)

He fell in love with his tutor. (17/20)

*The child misses his father. (11/21)

*A parent loses her patience. (41/31)

A politician feels embarrassed. (28/28)

*The boy was frustrated. (37/24)

*I became withdrawn. (39/27)

The volleyball player believes in his strength. (20/16)

The man was depressed. (38/31)

A friend has suicidal thoughts. (40/30)

References

- Ahn, W.-K., & Kalish, C. W. (2000). The role of mechanism beliefs in causal reasoning. In F. C. Keil & R. A. Wilson (Eds.), *Explanation and cognition* (pp. 199–225). Cambridge: MIT Press.
- Ahn, W.-K., & Nosek, B. A. (1998). Heuristics used in reasoning with multiple causes and effects. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the twentieth annual conference of the cognitive science society* (pp. 24–29). Mahwah: Erlbaum.
- Bloom, P. (2004). *Descartes' baby: How the science of child development explains what makes us human*. New York: Basic Books.
- Bloom, P. (2006). My brain made me do it. *Journal of Culture and Cognition*, 6, 209–214.
- Byrne, R. M. J. (1989). Suppressing valid inferences with conditionals. *Cognition*, 31, 61–83.
- Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, 104, 367–405. doi:10.1037/0033-295X.104.2.367
- Cummins, D. D. (2014a). The impact of disablers on predictive inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40, 1638–1655.
- Cummins, D. D. (2014b). Neural correlates of causal power judgments. *Frontiers in Human Neuroscience*, 8, 1014. doi:10.3389/fnhum.2014.01014
- Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory & Cognition*, 19, 274–282. doi:10.3758/BF03211151
- Danks, D. (2007). Causal learning from observations and manipulations. In M. C. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 359–388). New York: Erlbaum.
- Fernbach, P. M., Darlow, A., & Sloman, S. A. (2010). Neglect of alternative causes in predictive but not diagnostic reasoning. *Psychological Science*, 21, 329–336.
- Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naive theory of rational action. *Trends in Cognitive Sciences*, 7, 287–292. doi:10.1016/S1364-6613(03)00128-1
- Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: Two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology*, 37, 620–629.
- Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review*, 116, 661–716. doi:10.1037/a0017201
- Hamlin, J. K., Hallinan, E. V., & Woodward, A. L. (2008). Do as I do: 7-month-old infants selectively reproduce others' goals. *Developmental Science*, 11, 487–494.
- Hepach, R., & Westermann, G. (2013). Infants' sensitivity to the congruence of others' emotions and actions. *Journal of Experimental Child Psychology*, 115, 16–29.
- Jenkins, H. M., & Ward, W. C. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs: General and Applied*, 79, 1–17. doi:10.1037/h0093874
- Juhos, C., Quelhas, A. C., & Byrne, R. M. J. (2015). Reasoning about intentions: Counterexamples to reasons for actions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41, 55–76. doi:10.1037/a0037274
- Kemp, C., Goodman, N. D., & Tenenbaum, J. B. (2007). Learning causal schemata. In D. McNamara & G. Trafton (Eds.), *Proceedings of the 29th annual conference of the cognitive science society* (pp. 389–394). New York: Erlbaum.
- Lagnado, D. A., & Sloman, S. A. (2006). Time as a guide to cause. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 451–460. doi:10.1037/0278-7393.32.3.451
- Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, 25, 265–288.

- Markovits, H. (1986). Familiarity effects in conditional reasoning. *Journal of Educational Psychology, 78*, 492–494.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W. H. Freeman.
- Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's mechanical Turk. *Behavior Research Methods, 44*, 1–23. doi:10.3758/s13428-011-0124-6
- Mills, C., & Keil, F. C. (2004). Knowing the limits of one's understanding: The development of an awareness of an illusion of explanatory depth. *Journal of Experimental Child Psychology, 87*, 1–32.
- Newman, G. E., Keil, F. C., Kuhlmeier, V. A., & Wynn, K. (2010). Early understandings of the link between agents and order. *Proceedings of the National Academy of Sciences, 107*, 17140–17145.
- Nichols, S. (2004). *Sentimental rules: On the natural foundations of moral judgment*. Oxford: Oxford University Press.
- Olineck, K. M., & Poulin-Dubois, D. (2005). Infants' ability to distinguish between intentional and accidental actions and its relation to internal state language. *Infancy, 8*, 91–100.
- Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 208*, 255–258.
- Repacholi, B., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14- and 18-month-olds. *Developmental Psychology, 33*, 12–21.
- Rottman, B. M., & Keil, F. C. (2012). Causal structure learning over time: Observations and interventions. *Cognitive Psychology, 64*, 93–125.
- Rottman, B. M., Kominsky, J. F., & Keil, F. C. (2014). Children use temporal cues to learn causal directionality. *Cognitive Science, 38*, 489–513.
- Rozenblit, L. R., & Keil, F. C. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science, 26*, 521–562.
- Saxe, R., Tenenbaum, J. B., & Carey, S. (2005). Secret agents: Inferences about hidden causes by 10- and 12-month-old infants. *Psychological Science, 16*, 995–1001.
- Shlottmann, A. (1999). Seeing it happen and knowing how it works: How children understand the relation between perceptual causality and underlying mechanism. *Developmental Psychology, 35*, 303–317. doi:10.1037/0012-1649.35.1.303
- Shanks, D. R. (1995). *The psychology of associative learning*. Cambridge: Cambridge University Press.
- Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2004). Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science, 28*, 303–333. doi:10.1207/s15516709cog2803_1
- Spelke, E., Phillips, A., & Woodward, A. (1995). Infants' knowledge of object motion and human action. In D. Sperber, D. Premack, & A. Premack (Eds.), *Causal cognition: A multidisciplinary debate. Symposia of the Fyssen Foundation* (pp. 44–78). Oxford: Oxford University Press, Clarendon Press.
- Steyvers, M., Tenenbaum, J. B., Wagenmakers, E.-J., & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive Science, 27*, 453–489. doi:10.1207/s15516709cog2703_6
- Strevens, M. (2008). *Depth: An account of scientific explanation*. Cambridge: Harvard University Press.
- Strickland, B., & Suben, A. (2012). Experimenter philosophy: The problem of experimenter bias in experimental philosophy. *Review of Philosophy and Psychology, 3*, 457–467.
- Tenenbaum, J. B., & Griffiths, T. L. (2001). Structure learning in human causal induction. In T. K. Leen, T. G. Dietterich, & V. Tresp (Eds.), *Advances in neural information processing systems 13* (pp. 59–65). Cambridge: MIT Press.
- Tenenbaum, J. B., & Niyogi, S. (2003). Learning causal laws. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th annual conference of the cognitive science society* (pp. 1153–1157). Mahwah: Erlbaum.
- Waldmann, M. R., & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General, 121*, 222–236. doi:10.1037/0096-3445.121.2.222
- Walsh, C. R., & Byrne, R. M. J. (2007). The effects of reasons for acting on counterfactual thinking. *Thinking & Reasoning, 13*, 461–483.
- Wolff, P., Ritter, S., & Holmes, K. J. (2014). Causation, force, and the sense of touch. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th annual conference of the cognitive science society* (pp. 1784–1789). Austin: Cognitive Science Society.
- Woodward, A. (1998). Infants selectively encode the goal of an actor's reach. *Cognition, 69*, 1–34.