Title

The Search for Invariance: Repeated Positive Testing Serves the Goals of Causal Learning

ISBN

9783030355937

Authors

Lapidow, Elizabeth

Walker, Caren M

Publication Date

2020

DOI

10.1007/978-3-030-35594-4_10

Peer reviewed

The Search for Invariance: Repeated Positive Testing Serves the Goals of Causal Learning.

Elizabeth Lapidow & Caren M. Walker
University of California, San Diego

To appear in Childers, J.B., Graham, S.A. & Namy, L. (Eds). (2019). Learning Language and Concepts from Multiple Examples in Infancy and Childhood

What is invariant does not emerge unequivocally except with a flux. The essentials become evident in the context of changing nonessentials. – James Gibson, 1979

Human learners are intuitively exploratory: We acquire new knowledge from the outcomes of our actions. However, in order for exploration to support learning, at least some of these actions must serve to evaluate our *existing* knowledge. Despite this need for informative 'hypothesis testing' in everyday learning, decades of research examining self-directed experimentation suggests that learners rarely choose informative tests. That is, instead of selecting actions to test *whether* their current hypothesis is correct, both children and adults tend to prefer 'positive tests': actions that will produce an effect assuming their current hypothesis *is* correct (see Klayman, 1995; Zimmerman, 2007).

To illustrate, suppose you drop an ice cube on the floor and it shatters. As a learner, you might form an initial hypothesis that 'impact with an unyielding surface causes ice to shatter.' This hypothesis is also a causal explanation for your observation: indicating how one variable (*X*) *makes a difference* to the state of another variable (*Y*). According to traditional interpretations of Popper's (1959) falsificationist approach, testing this hypothesis would require disconfirming its alternatives. That is, assessing whether '*X* is the cause of *Y*,' requires 'negative tests,' or actions to determine whether *Y* occurs in the absence of *X*. Here, since *Y* is 'shattering' and *X* is 'impacting an unyielding surface,' you should drop an ice cube on a *yielding* surface (*not-X*), like rubber or cotton, to determine whether it will shatter.

However, learners rarely choose this kind of disconfirming action during their exploration. Instead, they are much more likely to *repeat* the initial observation: e.g., to pick up another ice-cube and drop it on the same surface, or a similar one. This tendency to generate multiple positive examples is a puzzling characteristic of self-directed learning, since it does not initially appear to be informative. After all, these repeated demonstrations often produce the same evidence, and do not distinguish between the current hypothesis (i.e., impacting an unyielding surface) and potential alternatives (e.g., impacting *any* surface at a particular speed), since they are consistent with both. Why then, do self-directed learners consistently and repeatedly conduct positive tests?

In this chapter, we propose a novel answer to this question: the Search for Invarinace (SI) hypothesis, which suggests that observing muliple, positive examples may facilitate learning by allowing us to assess the *invariance* of our causal theories. That is, by repeatedly activating a hypothesized cause and checking if its anticipated effect occurs, positive tests generate information about the degree to which this relationship holds across time and contexts. Given that the majority of the causal

relationships we encounter are probabilistic and interdependent, determining the degree of invariance is important for utilizing causal knowledge as a basis for action and inference. In order to test whether and when *X* (e.g., impacting an unyielding surface) reliably brings about *Y* (e.g., shattering in ice) it is necessary to *repeat X* (e.g., dropping more ice on similar surfaces), and observe whether *Y* occurs *again.*

The aim of the current chapter is to unpack this claim that the tendency to engage in positive testing is motivated by (and serves) our goals as causal learners. First, we will outline the empircial evidence for the use of a 'positive testing strategy,' during exploratory learning, and review existing theortical accounts that have been proposed to explain it. We will then introduce the Search for Invariance (SI) hypothesis and explain its foundations within theories of causality. After establishing this background, we will aim to apply our novel account to reinterpret some of the primary examples of positive testing in exploratory learning, and address potential objections and misinterpretations (e.g., sufficiency).

Positive Testing Strategy

A variety of learning and reasoning behaviors have been linked to (and confounded with) the Positive Testing Strategy (PTS), so it is important to first establish a working definition of this term. For the purposes of the current discussion, PTS will be treated as a phenomenon of hypothesis *testing*, and defined as the tendency (or preference) to select actions with the highest probability of producing the expected effect, if the current hypothesis were correct.[1] That is, we will focus on cases in which the learner assesses a hypothesis by examining its positive instances—either checking whether the expected effect occurs when the hypothesized conditions are met, or checking whether the conditions of the hypothesis are met when the event occurs (Klayman & Ha, 1987).

We will therefore *not* address accounts that focus primarily on whether young learners are able to generate hypotheses and evaluate their fit to evidence more broadly (e.g., Carey, Evans, Honda, Jay, & Unger, 1989; Kuhn, 1989; Kuhn et al., 1988). This discussion is also not intended to address '*confirmation bias*,' the failure to seek and consider (or even to avoid and distort) conflicting evidence, which is often presented alongside PTS in adult research[2] (for reviews, see Klayman, 1995; Nickerson, 1998). While the ability (and willingness) to reconcile an existing theory with new evidence is critical for exploration to support learning, it simply falls outside our current focus on the generation of evidence through self-directed action.

PTS in Scientific Reasoning

Inhelder and Piaget (1958) were the first to experimentally examine the understanding and use of the principles of experimentation in children. Later researchers

---

[1] As discussed below, there are many accounts of this behavior, not all of which use the term 'PTS.' Additionally, while the term has also been used to describe learners' motivation for conducting positive tests, we will restrict our use of 'PTS' to refer to observable behavior.

[2] The exact nature of the relationship between PTS and confirmation bias differs between accounts. PTS is variously suggested to be (a) an instance of (Nickerson, 1998; Wason, 1962), (b) a source of (see Nickerson, 1998 for review), and (c) a departure from confirmation bias (Klayman, 1995; Klayman & Ha, 1987).

adapted their methodologies to assess and improve scientific reasoning (e.g., Kuhn & Angelev, 1976; Kuhn & Brannock, 1977; Siegler & Liebert, 1975), and a tremendous body of research has grown out of this initial work (see Zimmerman, 2000, 2007; Zimmerman & Klahr, 2018 for reviews). Studies typically present children with multivariate contexts and assess their ability to systematically combine and isolate these variables to reveal causal relationships. In some cases, participants are instructed to determine the variable(s) causally related to an outcome (e.g., which chemicals cause a color reaction when mixed; Kuhn & Phelps, 1982). In others, children are asked to determine whether and how variable(s) make a difference to a certain outcome (e.g., which features of a racecar determine its speed; Schauble, 1990). Another common approach is to indicate a variable of interest and ask participants to test hypotheses about its effect (e.g., the operation performed by a computer input command; Dunbar & Klahr, 1989; Klahr & Dunbar, 1988; Klahr, Fay, & Dunbar, 1993).

The bulk of this research finds the development of experimentation skills to be slow and error-filled (e.g., Dunbar & Klahr, 1989; Inhelder & Piaget, 1958; Klahr & Chen, 2003; Klahr et al., 1993; Kuhn, 1989; Tschirgi, 1980; Valanides, Papageorgiou, & Angeli, 2014). Importantly, many of these errors resemble PTS: Children tend to repeatedly choose actions and experiments that are expected to generate an effect (e.g., the color reaction, the fastest car), if their causal hypothesis were correct. This behavior is often interpreted as driven by children's desire to 'demonstrate the correctness' of their hypotheses (e.g., Dunbar & Klahr, 1989; Inhelder & Piaget, 1958; Klahr et al., 1993; Kuhn & Phelps, 1982). Other researchers have viewed these explorations as evidence of a misplaced focus on an action's *tangible* outcomes, rather than its informative potential (e.g., Kuhn & Phelps, 1982; Schauble, 1990; Schauble, Glaser, Duschl, Schulze, & John, 1995; Siler & Klahr, 2012; Siler, Klahr, & Price, 2013; Tschirgi, 1980; Zimmerman & Glaser, 2001). That is, rather than trying to learn the relations between cause and effect, children seem to select experiments in order to reproduce positive outcomes and avoid negative ones.

Tschirgi (1980) is perhaps the most cited example of PTS in scientific reasoning (see Croker & Buchanan, 2011; Klayman & Ha, 1987 for discussions). In this study, 2nd-, 4th-, and 6th-grade children and adults were asked to choose an experiment to prove that a variable was causally responsible for either a positive or negative outcome. In one scenario, a character bakes a cake using one of two types of each of three ingredients (a flour, a sweetener, and a fat), and the cake comes out well. The character believes that the type of *sweetener* causes this outcome, while the types of *flour* and *fat* do not matter. Participants were then given a choice between three potential experiments and asked to select one to prove the character's hypothesis. They could (1) change the suspected cause and keep the other two variables constant ('VARY'), (2) change the other two variables and keep the suspected cause constant ('HOLD'), or (3) change all three variables ('CHANGE ALL'). According to Tschirgi, the VARY option, which isolates the suspected causal variable, is the only informative test of the character's hypothesis. Interestingly, however, participants of all ages only preferred this option when the outcome of the initial scenario was *negative* (i.e., when the cake came out badly). Otherwise, learners preferred the HOLD option—keeping the variable of interest constant and changing the others. Tschirigi (1980) explains this finding as evidence that children and adults tend to select experiments based on practical, rather than 'logical', concerns.

PTS in Rule Learning

There is also extensive evidence of PTS within adult's hypothesis testing (see McKenzie, 2004 for review). The classic example of this behavior is the 2-4-6 task (Wason, 1960),[3] which asks participants to determine the rule used to generate sequences of three numbers. At the start of the task, the experimenter provides an example of a sequence generated by the rule (e.g., "2, 4, 6"). Participants are then able to experiment by generating novel sequences and requesting feedback from the experimenter about whether they follow the rule. Notably, the predominant strategy is to test cases that *fit* the rule one has in mind: to test positive instances of one's current hypothesis. For example, most adults, given the "2, 4, 6" example, form the hypothesis that the rule is 'ascending even numbers.' In gathering evidence, the majority of participants will test sequences that follow this rule (e.g., "10, 12, 14", "-2, 0, 2", "104, 106, 108", etc.) and treat affirmative feedback from the experimenter as evidence serving to increase the strength of their belief in the accuracy of their hypothesis.

The problem with using a positive testing strategy in this task is that it misses the correct (and more general) rule, 'ascending numbers,' and provides no evidence to suggest that an error has been made. Wason (1960) interpreted this as evidence of participants arriving at their hypotheses through a process of 'enumeration' rather than 'elimination.' In other words, participants assume that confirming evidence alone is enough to justify their conclusions. While some participants do eventually discover the correct rule, most of them only do so after *many* positive tests of incorrect rules. Wason's task, and his conclusion that learners have a general tendency to only consider verification (Wason, 1960; Wason & Johnson-Laird, 1972), have both been used extensively as evidence of adults' biased hypothesis testing and reasoning (e.g., Gorman & Gorman, 1984; Mahoney & DeMonbreun, 1977; Tukey, 1986; Tweney et al., 1980; Wetherick, 1962).

Theories of PTS

Given that PTS is a widespread and well-documented phenomenon, there have been numerous attempts to theoretically account for the behavior. These accounts may be roughly separated into two categories, depending on whether PTS is explained as a means of generating an outcome or as a means of generating evidence.

PTS as a means of generating outcomes.

As mentioned above, some have suggested that learners prefer PTS because their selection of actions is motivated by tangible outcomes, rather than information value (e.g., Kuhn & Phelps, 1982; Tschirgi, 1980). This interpretation is particularly prominent in explanations of the failures observed in early scientific reasoning that suggest that young learners begin with an incorrect intuition about the *function* of experimentation. Specifically, children are described as being preoccupied with 'making events happen' (e.g., making a good cake, producing a color reaction), rather than identifying the causal factors responsible for these outcomes.

Based on this evidence, Schauble, Klopfer, and Raghavan (1991) proposed the 'science vs. engineering models' of self-directed behavior: Individuals can adopt either a 'science goal' (to determine causal relationships) or an 'engineering goal' (to generate or

[3] Wason also created another classic task of hypothesis testing, the Selection Task (1968), which falls outside the scope of the current discussion.

reproduce a particular effect) in their interactions with their environment. According to the authors, children incorrectly approach scientific reasoning tasks using an 'engineering' model, which is concerned with generating outcomes and stops whenever its target (or an acceptable approximation) is achieved. Schauble (1990) distinguishes this account of PTS from the one offered by Klayman and Ha (1987). In particular, the engineering model does *not* claim that learners are seeking falsification, or information of any kind. Instead of trying to determine the causal relationships between variables and outcomes, young explorers manipulate variables in an attempt to bring about desirable outcomes (Schauble, 1990). In other words, according to this theory, learners' interventions are often uninformative because *information is not their goal*. The strongest version of this account implies that children do not differentiate between understanding an event and making it occur. Although later empirical work provides evidence against this claim (Sodian, Zaitchik, & Carey, 1991; see also Lapidow & Walker, 2019), the 'science vs. engineering' explanation of PTS continues to be cited and used as a framework for understanding choice behavior within scientific reasoning (e.g., Siler & Klahr, 2012; Siler et al., 2013).

Other researchers have made suggestions similar to the 'science vs. engineering' hypothesis, but concerning adult choice behavior (e.g., Friedrich, 1993; Schwartz, 1982; Vogel & Annau, 1973). These interpretations suggest we employ PTS because we value maximizing current over long-term gains. One early articulation of this idea comes from Einhorn and Hogarth (1978), who describe a conflict between acting in accordance with the current hypothesis (i.e., maximizing current success), and acquiring information to improve it (i.e., increasing long-term success). As a result, in contexts where the potential cost of false positives is more damaging than that of false negatives, learners may be entirely justified in gathering positive evidence.

These 'outcome-focused' theories of PTS all point to the tension between 'exploitation,' taking an action that is known to have a high likelihood of success, and 'exploration,' taking a less certain (or less rewarding) action in order to improve one's epistemic status. Schauble and colleagues (1990; 1991) suggest that this tension results from a lack of understanding on the part of the learner, while Einhorn and Hogarth (1978) and others point to situational factors that lead the learner to prioritize exploitation. Regardless of the source, this is a real tension in behavior, and has been shown to influence decision-making on a variety of tasks. For example, Heyman and Dweck (1992) report that an individual's response to challenges and failures may be explained by whether they hold a 'performance goal' (to prove one's competence) or a 'learning goal' (to improve one's skills).[4] A recent study has also shown that sensitivity to the distinction between production and investigation influences the actions of causal learners as a rational adaption to context demands (Yoon, MacDonald, Asaba, Gweon, & Frank, 2018). However, regardless of whether it is due to lack of understanding or situational pressures, the tension between productive and informative actions is likely not the *only* factor responsible for PTS.

[4] While this distinction resembles Schauble et al.'s (1991) notion of 'science vs. engineering models,' Heyman and Dweck's (1992) account is agnostic about the immediate goal. You could, therefore, conceivably have either a 'learning' or a 'performance' goal *within* either a 'science' or an 'engineering' model.

PTS as a means of generating evidence.

The accounts reviewed above argue that positive tests reflect learners' desire to bring about tangible effects. However, others accounts have suggested that PTS is an *intentional* (though not always valid) form of hypothesis testing. For example, Wason's (1960; 1972) interpretation of adults' behavior on the 2-4-6 task suggested that learners believe that their positive tests provided valid conclusive evidence. Wason based this analysis on the prescriptions for hypothesis testing laid out by Popper (1959) who argued that instances that *verify* a hypothesis are always ambiguous, since they can occur even if the hypothesis is ultimately incorrect. Learners should therefore aim to collect counter-evidence when testing hypotheses, since falsifying evidence is always *conclusive* (i.e., observing a single black swan can overturn an entire lifetime of positive evidence that 'all swans are white'). Further, this type of observation cannot be countered by positive evidence later on (e.g., no number of white swans observed *after* the single black one will make the statement that 'all swans are white' true). Popper's prescription for scientific hypothesis testing is therefore to make the falsification of alternatives—not the generation of positive evidence—the primary goal of experimentation. By comparing participants' behavior to this standard, Wason concluded that PTS is a logical or cognitive failing of our intuitive hypothesis testing, and many others have echoed this interpretation (e.g., Baron, Beattie, & Hershey, 1988; Devine, Hirt, & Gehrke, 1990; Skov & Sherman, 1986; Wason & Johnson-Laird, 1972).

Indeed, hints of this perspective appear in the account of PTS suggesting that children aim to 'demonstrate the correctness' of their hypotheses during scientific reasoning tasks (e.g., Dunbar & Klahr, 1989; Inhelder & Piaget, 1958; Klahr et al., 1993; Kuhn & Phelps, 1982). For example, Klahr and colleagues (1993) gave 3rd- and 6th-grade children and adults either a plausible or implausible hypothesis for the operation of one input in a simple programming system. If the starting hypothesis was plausible, participants tended to go about generating positive evidence of its validity. When the starting hypothesis was implausible, however, participants were more likely to set up experiments to discriminate between this initial claim and self-proposed alternatives. Tellingly, these participants (and young children in particular) were often 'sidetracked' by generating positive evidence for their rival hypothesis (Klahr et al., 1993).

In other accounts, PTS is not assumed to be driven by an error or illusion of validity, but is instead treated as one of several potentially useful inquiry strategies available to the learner. The classic form of this argument comes from Klayman and Ha's (1987) analysis of the 2-4-6 task. These authors distinguish between the use of falsification as a *goal* versus as a *method* of hypothesis testing. The goal is what Popper (1959) prescribes, but the method is not necessary to achieve it. The two are confounded in Wason's task because most participants begin with a hypothesis that is more specific than the correct one. As a result, testing negative instances of one's hypothesis (falsification as method) is the *only* means of generating disconfirming evidence (falsification as goal). Klayman and Ha argue that this circumstance is neither necessary nor typical in real world learning. More often, correct hypotheses are a 'minority phenomenon,' so most tests of positive instances will result in negative responses (Klayman & Ha, 1987). Given this 'rarity', PTS is argued to be a more efficient means of seeking informative disconfirmation than negative tests, and therefore a reasonable hypothesis testing heuristic.

Klayman and Ha's account has since been broadly adopted to explain PTS in adult learning (see Coenen, Nelson, & Gureckis, 2018), and similar arguments have been successfully applied to other instances of PTS beyond the 2-4-6 task (e.g., Oaksford & Chater, 1994). Navarro and Perfors (2011) have also extended this argument by demonstrating that PTS is a 'near-optimal' learning strategy in contexts where correct hypotheses are 'sparse,' and provide further justification for learners' intuitive assumption of sparsity. Importantly, these accounts present PTS as a default heuristic approach to hypothesis testing. That is, in the absence of more specific information, learners employ PTS because it is cognitively inexpensive and often effective (Klayman & Ha, 1987).

In contrast, other accounts have stressed that learners are sensitive to their learning context, and *selectively* employ PTS. For example, McKenzie and Mikkelsen (2000) show that manipulating participants' beliefs about the rarity of events changes the degree to which they appeal to PTS. Furthermore, recent computational work finds that intervention choice is best captured by a model employing a *mix* of PTS and expected information gain in a way that is sensitive to task demands. Coenen, Rehder, and Gureckis (2015) modeled PTS in causal learning as a preference to intervene on variables with the greatest proportion of downstream effects, treating each consistent outcome as a point of positive evidence for the hypothesis. Their analysis compared this model, and one designed to favor interventions most likely to distinguish between competing hypotheses (i.e., with the highest expected information gain), to participants' chosen interventions. A model that employed *both* strategies best captured the intervention choices of adult causal learners. Results also revealed that strategy selection was sensitive to the learning context: When the task feedback indicated that one strategy was insufficient for distinguishing the true causal structure, or when participants were placed under time pressure, they shifted flexibly between the PTS and information gain models. Thus, the most recent research suggests that PTS is *not* a default heuristic for generating evidence, but an adaptive and efficient hypothesis-testing strategy employed by context-sensitive learners.

Other accounts and overlapping evidence.

While these two broad categories of accounts—positive tests as a means of generating outcomes and positive tests as a means of generating evidence—help to organize much of the existing literature, some prior accounts do not neatly conform to *either* category. For example, a recent 'self-teaching' model of active learning (Yang, Vong, Yu, & Shafto, 2019), proposes that PTS may be a by-product of the way that interventions are chosen. That is, although the ideal learner usually selects actions according to expected information gain, selections that deviate from this accord with PTS. In this context, PTS is not presented as a mistaken focus (e.g., Schauble et al., 1991) or a logical error (e.g., Wason, 1960), but it is also not presented as an adaptive learning tool (e.g., Coenen et al., 2015).

It is also not always possible to distinguish between generating evidence and generating effects within a learner's behavior. For example, in many cases of causal learning, the action that is expected to maximize the probability of the most likely hypothesis is *also* the action that is expected to have the most positive tangible

outcomes.[5] For example, McCormack, Bramely, Frosch, Patrick, and Lagnado (2016) presented children with a three-variable causal system and three competing hypotheses: a common cause (activating component A causes components B and C to activate) or one of two causal chains (activating A causes B to activate, which causes C to activate, or activating A causes C to activate, which causes B to activate). Five- to six-year-olds preferred to *repeatedly* activate component A (the root node in all hypotheses). Although activating A is expected to activate all the other components in the system, it is not clear what motivates this positive intervention. This behavior accords with 'generating effects' theories of PTS, since turning on the root node of a system 'makes the most things happen,' but also with 'generating evidence', theories, since this action also tests the greatest proportion of downstream connections.

A similar study by Meng, Bramley, and Xu (2018) tested 5- to 7-year-olds on a modified version of Coenen et al.'s (2015) task, and also found a preference for this type of intervention. Although children's intervention choice was best captured by a model incorporating both PTS and information gain (as with adults), this mix was heavily skewed towards PTS (i.e., intervening on the node with the greatest proportion of dependent causal links), with the vast majority of children using PTS as their primary intervention strategy. Again, we observe evidence that young causal learners preferentially select positive tests, and again, the source of this preference remains unclear.

The Current Theory: Positive Testing and Causal Learning

Rather than explaining PTS as an error, bias, or byproduct of our inquiry strategies, the current proposal (the Search for Invariance [SI] hypothesis) presents an alternative account that draws on evidence describing our strengths as self-directed *causal* learners. In contrast to the difficulties documented in the scientific and rule learning domains, we excel at exploratory causal learning from an early age.

Considerable evidence shows that children spontaneously and preferentially explore what is most likely to be informative, given their current causal beliefs (see Schulz, 2012 for a review). Research in causal learning typically presents children (3- to 6-year-olds) with evidence about a novel physical device, and then allows them to interact with it during a period of free-play. When the evidence is ambiguous or violates their current theories, children engage in significantly more exploration (Bonawitz, van Schijndel, Friel, & Schulz, 2012; Gweon & Schulz, 2008; Schulz & Bonawitz, 2007; Schulz, Standing, & Bonawitz, 2008). They are also more likely to take potentially informative actions during this exploration (Lapidow & Walker, 2019; Schulz & Bonawitz, 2007; van Schijndel, Visser, van Bers, & Raijmakers, 2015).

For example, Cook, Goodman, and Schulz (2011) found that 4- to 5.5-year-olds select and even spontaneously invent informative interventions in their exploration of an ambiguous causal system. Children were introduced to a toy that played music when beads were placed on top of it. They were taught either that *all* beads caused the toy to activate, or that only *some* did, while the rest were inert. The beads could be snapped together to form two-bead pairs, and a snapped pair of novel beads was given to children during their free-play. This pair caused the toy to activate, but it was impossible to tell

[5] See Bramley, Lagnado, & Speekenbrink (2015) for an in depth treatment of this overlap between expected probability gain and expected utility gain models of intervention.

from this observation alone whether *both* beads in the pair had the power to activate the toy, or only *one*. Faced with this ambiguity, children in the 'some-beads' condition took informative actions: pulling the pair apart and trying the beads on the toy in isolation, in order to disambiguate their causal status. In contrast, children in the 'all-beads' condition, for whom the pair was *not* perceived as ambiguous, did not produce these actions. Furthermore, when given an ambiguous pair that was permanently glued together, several children in the 'some-beads' condition spontaneously turned the pair on its end, conducting informative hypothesis tests to isolate the beads in a way they had never seen demonstrated.

The fact that young children are such voracious and effective self-directed learners in these contexts raises the possibility that causal learners may have different information-seeking goals than those typically assumed in studies of scientific experimentation. This view of self-directed causal learners as 'intuitive scientists' (Brewer & Samarapungavan, 1991; Carey, 1985; Gopnik & Meltzoff, 1997; Karmiloff-Smith, 1988) is the impetus for the SI hypothesis: that PTS is normatively motivated by the concerns of causal learning and an informative means of assessing causal *invariance* across the set of examples tested. In a causal world of predominately probabilistic and interdependent relationships, repeated generation and investigation of positive instances provides critical information about the reliability and consistency of our causal hypotheses. In other words, PTS is useful and informative because it allows learners to examine the degree to which a dependency between variables continues to hold over time and across contexts. In this way, the SI hypothesis suggests that our goals as causal learners might explain our behavior as intuitive scientists.

Causal Invariance and Interventionism

In order to more fully describe the SI hypothesis and establish its relevance for interpreting PTS, we will first situate the concept of invariance within theories of causality.

Numerous theories of causality and causal explanation include a central idea that the function of causal knowledge is to highlight patterns of dependence that will generalize to future contexts (see Hitchcock, 2012 for details). A few notable examples of this include notions of 'sensitivity' (Lewis, 1974), 'robustness' (Redhead, 1987), 'non-contingency' (Kendler, 2005), 'exportability' (Lombrozo & Carey, 2006), 'insensitivity' (Ylikoski & Kuorikoski, 2010), 'portability' (Weslake, 2010), and 'transportability' (Pearl & Bareinboim, 2011). Indeed, this sensitivity to regularities in variable input is critical for knowledge acquisition in other domains as well (e.g., see Wu, Gopnik, Richardson, & Kirkham, 2011).

Although these accounts vary, Sloman's (2005) description of 'invariance' provides a sense of the common ground among them. He argues that, in every domain, aspects that are *invariant* across instances hold the most valuable information for learners. These aspects represent the "stable, consistent, and reliable properties that hold across time and across different instantiations of a system" (Sloman, 2005, p. 15). Knowledge of invariance therefore allows learners to predict, explain, and manipulate events in the world. To clarify, the concept of invariance is not necessarily defined in terms of causality—for example, recognizing the invariant statistical regularities among syllables within streams of continuous speech supports early language learning (e.g., Saffran, Aslin, & Newport, 1996; Saffran, Johnson, Aslin, & Newport, 1999). However,

the causal relations that govern observable events are usually highly systematic and generalizable. Causal knowledge is therefore a key source of invariance, and stable causal principles are often the most reliable basis for inference and interaction available to us (Sloman, 2005).

Much of Solman's account is drawn from the interventionist perspective on causal explanation (e.g., Woodward, 1997, 2003, 2006, 2010), which defines a causal relationship in terms of the invariance between variables following some change. That is, if *X* causes *Y*, then intervening to change the value of *X* would result in a change to the value of *Y*. A causal explanation is therefore a true claim describing how some factors *make a difference* to others. To illustrate this, suppose that two variables, *X* and *Y*, are observed to co-occur. If an intervention changing the value of *X* maintains this correlation (that is, it leads to a corresponding change in the value of *Y*), then the relationship between *X* and *Y* is *invariant* (it continues to hold), under at least some interventions. This relationship is regarded as causal, and can be used as a basis for inference and manipulation. To make this idea more concrete, imagine that *X* is fertilizer and *Y* is plant growth. If the statement, "Fertilizer causes plant growth," is an accurate causal explanation, we would expect changes to the amount of fertilizer (*X*) to lead to *systematic* changes in the growth of the plant (*Y*) (Woodward, 1997).

Note that the interventionist concept of invariance outlined so far defines a causal explanation as representing a kind of *counterfactual dependence*: It indicates how changing the factors included in the explanation would lead to a difference in the phenomenon being explained (Woodward, 1997; 2003). For example: the hypothesis that 'impacting an unyielding surface causes ice to shatter' provides a causal explanation for shattering, and implies the counterfactual that, in the absence of such an impact, shattering would not occur. Causal knowledge not only allows us to reason about how one factor makes a difference to another, but also to exploit those dependencies in our actions. Interventionism views causal learning and reasoning as rooted in our "highly practical interest" in manipulation and control of our environment (Woodward, 2003, p. 10). As a result, the value of identifying causal relationships is inherently tied to the knowledge of the actions this information supports.

In addition to its role in defining causal relationships, invariance is also highlighted as a quality that is expressed by causal relationships. While the former captures the continuity of causal relationships when related variables change, the latter emphasizes the 'stability' of those relationships across contexts and conditions. That is, *X* is a cause of *Y*, if and only if, an intervention on *X* changes the value of *Y* in at least some background circumstances *b*. These 'background circumstances' are aspects of a situation that are not explicitly represented by *X* or *Y*, but that are critical to the meaning and commitments of causal explanations (Blanchard, Vasilyeva, & Lombrozo, 2018). For example, "Fertilizer causes plant growth," is only true if the plant is also getting water, sunlight, and oxygen. These are some of the *background circumstances* of the causal relationship between fertilizer and plant growth; these are both critically important to our understanding of the causal claim and not explicitly stated as part of the relation.

Of course, for any causal relationship, there are changes to both the related variables and the background conditions under which the relationship will *not* hold. The interventionist definition of causality only requires a relationship to hold under at least *some* conditions (e.g., fertilizer *is* a cause of plant growth, because of the dependence

between them, even though this relationship will not hold if the plant is kept in the dark or the fertilizer is infested with harmful bugs). Our notion of invariance includes our understanding of these conditions. Vasilyeva and colleagues (2018; 2018) explain this 'stability' component of invariance as a combination of 'breadth,' or generality (the range of background circumstances in which the generalization holds) and 'guidance,' or accuracy (the support a causal explanation provides for generalization to new circumstances). This dual consideration reveals the importance that causal learners place both on identifying causal relationships that are most reliable *and* knowing the contexts in which they may be relied on. Empirical evidence supporting interventionism's role for invariance in our causal thinking comes primarily from studies looking at highly similar concepts, such as 'explanation generality' (e.g., Friedman, 1974; Gelman, Star, & Flukes, 2002; Johnston, Sheskin, Johnson, & Keil, 2018; Kitcher, 1981; Strevens, 2009; Walker, Lombrozo, Legare, & Gopnik, 2014). Very recently, computational (Morris et al., 2018) and behavioral (Vasilyeva et al., 2018) studies have begun to look explicitly at interventionist invariance and find that it both reflects and influences our causal judgments.

According to interventionism then, the purpose of causal learning lies in acquiring representations of counterfactual dependence between variables that generalize to guide future action and inference. However, learners are almost never in an epistemic position to form exception-less generalizations, which would require specifying *every* relevant contributing and enabling factor involved in causing an event. And, in fact, such an explanation would be severely limited in usefulness as it would be applicable to far fewer new situations than a less complete account. Instead, causal learners acquire causal explanations that are incomplete generalizations, and augment them with evaluations of their invariance over time and across contexts. Given this, it makes sense for causal learners to value information about invariance and to seek to uncover it through the exploration of multiple examples.[6] This is the foundation of the Search for Invariance (SI) hypothesis.

The Search for Invariance (SI) Hypothesis

The SI hypothesis proposes that the positive testing strategy (PTS) may provide a means of assessing this critical characteristic of causal invariance during exploration and hypothesis testing. Since no causal relationship is without exceptions or holds in all circumstances, knowing the extent and contexts in which relationships are invariant is critical for causal knowledge to guide action and reasoning. It is therefore incumbent upon causal learners to determine the invariance of putative causal relationships by asking, for example: Is this dependency reliable? Is it generalizable? And, if so, with what probability and in what contexts? Activating the hypothesized causal variable (or examining cases in which it has already been activated), allows the learner to assess whether the effect behaves as hypothesized in the current context. Repeating these interventions provides evidence for how consistently and extensively this relationship holds. Negative tests—while assessing whether an alternative cause might also bring about the effect in the current context—cannot, by definition, provide this information.

[6] See also Casasola and Park, this volume, and Imai and Childers, this volume, for other instances of how the observation of multiple examples influences how learners generalize their knowledge.

Thus, positive tests are potentially informative for self-directed learners, *regardless* of their expectations about the sparsity or rarity of causal relationships, by providing evidence about their relative invariance.

To further illustrate this claim, the following sections will return to three key examples of PTS to demonstrate how reinterpreting these results in terms of the learner's search for invariance provides a novel and more complete account of these behaviors.

As an alternative to 'engineering desirable outcomes'.

Tschirgi (1980) provides an excellent example of how the SI hypothesis reframes the tendency to conduct positive tests to bring about desirable outcomes. Recall that when asked to select one of three actions to prove that the type of sweetener was the reason that the cake was good, and that the types of fat and flour did not matter, the majority of participants chose to HOLD the suspected causal variable and change the others, *rather* than VARY only the suspected cause (Tschirgi, 1980).

From the perspective of the SI hypothesis, the HOLD option presents a valuable test of the circumstances under which the suspected causal dependency holds. Baking another cake in which the type of sweetener is kept constant, and all the other ingredients are changed, allows the learner to assess whether this relationship is invariant under different background circumstances[7]. Another example of this kind of hypothesis testing in scientific reasoning comes from Zimmerman and Glaser (2001), who asked 6th graders to test the claim that coffee grounds are good for plants. The authors found that the majority of students designed a series of positive tests—that is—they checked the outcome of adding coffee grounds to a variety of different plants, thereby testing the hypothesized effect of this intervention across a variety of background conditions.

On the other hand, it might be argued that assessing invariance is illogical in a situation in which the causal status of the hypothesized relationship has not yet been conclusively established. Zimmerman (2007) makes precisely this objection, explaining that the participants in Tschirgi (1980) failed to first confirm the claim that the sweetener produces good cake in a controlled manner. However, it is not clear that this confirmation is necessary. The hypothesis presented in the task makes two distinct claims: (1) good cake is causally dependent on the sweetener, and (2) good cake is causally independent of the type of flour and fat. Tschirgi sees (1) as the *only* claim participants are being asked to test. However, there is nothing to stop participants from choosing to evaluate (2), which would make HOLD, and not VARY, the disconfirming test. This duality means that HOLD is just as valid a test of the hypothesis stated in the problem text as VARY. Further, HOLD has the additional attraction of also assessing the invariance of the causal relationship between the sweetener and good cake that was singled out by the prompt.[8]

As an alternative to 'seeking confirmatory evidence.'

---

[7] In fact, since the scenarios only contained two values for each variable, the HOLD option tests invariance for *all* possible kinds (though not combinations) of other factors.

[8] This is not the only study in the scientific reasoning literature with such ambiguities. Assumptions about parameters—the number of causal variables, whether their effects are independent or interdependent, probabilistic, or deterministic—are regularly made by experimenters, but not conveyed to participants or considered when evaluating their behavior. Ongoing work in our lab aims to remove these ambiguities to better assess children's intuitive experimentation.

The SI hypothesis also provides an explanation for participants' repeated testing of instances that are consistent with their current hypothesis, as seen in Wason's 2-4-6 task. Again, learners are testing the invariance of their hypothesized rule, but it is not the same quality of invariance as the one tested in the previous example. In that case, the learner's goal was to assess the range of background circumstances in which a causal claim holds; what Vasilyeva and colleagues (2018; 2018) call 'breadth.' In contrast, explorations in the Wason task are more concerned with 'guidance,' the *accuracy* of a causal hypothesis for developing expectations about novel circumstances. By checking multiple sets that are all consistent with their current hypothesis (e.g., 'increasing even numbers') across distinct instances, learners can determine whether the rule invariantly identifies sets with the target property. Further evidence for this interpretation is what Klayman and Ha (1987) call *limit testing*: Within their repeated positive tests, participants in the 2-4-6 task commonly select extreme or unusual instances of their hypothesized rule. For example, a participant considering the hypothesis 'increasing even numbers' might choose to test the set: -2, 0, 2 (Klayman, 1995; Klayman & Ha, 1987).

Again, it might be objected that the current hypothesis has not yet been verified. It is true that evidence of the invariance of 'increasing even numbers' does not rule out the possibility that the rule is actually 'increasing numbers.' However, if any instance of the 'increasing even numbers' rule *ever* fails one of these tests of invariance, the learner will know that *neither* hypothesis is correct. If learners were simply hoping to generate confirmatory evidence or produce affirmative responses from the experimenter, then we would *not* expect them to preferentially test their hypotheses at the regions of highest uncertainty (i.e., at the *limits*). Instead, these investigations serve as a stress test of the invariance of their hypothesis, even at its boundaries.

Beyond evidence of sufficiency.

Another likely objection to the SI hypothesis is to claim that invariance is not meaningfully distinguishable from *sufficiency*. A sufficient cause is adequate, but not required, for bringing about an effect (Klayman & Ha, 1987) and researchers have previously proposed the use of sufficient hypotheses as an explanation of PTS. These accounts tend to emphasize pragmatic motivations (e.g., a desire to achieve or avoid specific outcomes) over epistemic ones. For example, Friedrich (1993) suggests that human cognition, which was shaped by a drive to ensure survival, is better suited to identifying sufficient mechanisms than determining truth. Similarly, Schwartz (1982) explains PTS as a kind of error avoidance, motivated in part by the conditions of reward and reinforcement. In other words, once the learner identifies a sufficient cause, they may feel no compulsion to determine whether that condition is also necessary (Nickerson, 1998).

It is difficult to distinguish the SI hypothesis from this alternative account since the occurrence of PTS does not, in itself, indicate the motivation behind it. For example, in the 2-4-6 task, selecting sets that are sufficient (i.e., have a high probability of producing an affirmative response based on what is currently known) and selecting sets that test the invariance of the hypothesized rule can lead to the same actions. Despite this challenge, we maintain that invariance *cannot* be reduced to sufficiency.

To explain this position, it is important to recognize that causal inference and rule-based inference involve different notions of sufficiency. Rule learning typically assumes that there is only one correct rule, which must be both sufficient and necessary

(Klayman & Ha, 1987). Further, these conditions are defined in terms of propositional logic (see Johnson-Laird & Byrne, 2002). For example, in the 2-4-6 task, any set that follows the correct rule will have the target property (the rule is sufficient), and all sets that have the target property will follow the rule (the rule is necessary). Causation, on the other hand, does *not* follow the rules of standard inference, and causal logic involves assumptions that cannot be captured by the principles of propositional logic (see Sloman & Lagnado, 2015; 2005).

To illustrate this difference, consider these two pairs of hypotheses: (a) *sets of numbers increasing by two have the target property*, and (b) *sets of increasing numbers have the target property*, as opposed to: (c) *sex causes pregnancy*, and, (d) *embryo fertilization causes pregnancy*. Both (a) and (c) are cases in which the antecedents ('*increasing by two*' and '*sex'*) are sufficient, but not necessary for their consequents ('*having the target property*' and '*pregnancy*'). Sets of numbers increasing by two *will* be in the target set, and sex *will* (under certain background conditions) be a cause of pregnancy. However, the truth of the hypotheses (b) and (d) accounts for *why* the antecedents of (a) and (c) bring about their consequents, which makes them unnecessary. Sex is not necessary for pregnancy (which can also occur through in vitro fertilization), and increasing by two is not necessary for a set to have the target property (as all sets of numbers increasing by two are also sets of increasing numbers).

However, there is also an important difference here: (a) and (b) are hypotheses about rules, while (c) and (d) are hypotheses about causes. The domain of rule discovery requires that there be only one correct rule that is both necessary and sufficient for the target property. In contrast, the domain of causal reasoning allows for multiple possible causes to exist simultaneously. As a result, (b) being true means (a) cannot be the correct rule, but (d) being true does not mean that (c) cannot be a correct causal explanation. Put another way, the hypothesis that '*the rule for the target property is numbers increasing by two*' is incorrect, but the hypothesis that '*sex is a cause of pregnancy*' is not. The difference between propositional and causal logic means that the impact of one statement's truth-value on another's differs between the domains of rule and causal learning. Lack of necessity makes a rule false, but it does *not* make a cause false. That said, it *does* make it less invariant.

Judgments of necessity and sufficiency are critical to our reasoning in the causal domain, but in a way that is unique to the domain. Necessity captures our intuition that causal variables ought to *make a difference* to outcomes: We judge dropping an ice-cube on the ground as the cause of it shattering, since, in the absence of the first event, the second would not have occurred. Indeed, a wealth of evidence shows such evaluations are central in our causal judgments (e.g., Gerstenberg, Goodman, Lagnado, & Tenenbaum, 2014; Gerstenberg, Peterson, Goodman, Lagnado, & Tenenbaum, 2017; Icard, Kominsky, & Knobe, 2017; Morris et al., 2018; Wells & Gavanski, 1989).

Causal sufficiency is also intertwined with the notion of necessity. For example, when we credit one event (e.g., dropping an ice-cube) as the difference-making cause of another event (e.g., shattering), we understand that the necessity of the first event for the occurrence of second is *dependent on the context* in which they take place. That is, dropping the ice-cube will only cause it to shatter under certain background conditions (e.g., gravity). We also understand that shattering might occur in other cases, even in the absence of dropping (e.g., if someone hits the ice-cube with a hammer).

Thus, unlike a rule, a single causal variable is never sufficient *or* necessary for bringing about an effect in and of itself (Mackie, 1974). In causal reasoning, these qualities exist only given certain background conditions in which the occurrence of the variable makes a difference to the outcome. Our knowledge of invariance, of the aspects of a situation *must* be in place for the cause to produce its effect, captures this. It thereby requires a consideration of necessity and cannot be reduced to assessment or employment of sufficiency.

As an account of previously ambiguous evidence.

Finally, the SI hypothesis also provides a consistent interpretation of learners' tendency to intervene on the root node in McCormack et al. (2016) and Meng et al. (2018).[9] Recall that according to interventionism, the commitments a causal hypothesis makes about the potential for action and manipulation are central to our reasoning about it. In fact, according to Woodward, the difference between competing causal models is *understood* in terms of the difference in the predicted outcomes of interventions (2006, p. 61). This suggests that a causal learner might not feel the need to distinguish between causal hypotheses that make the same predictions about the majority or more salient interventions on a system.

The problems used in McCormack et al. (2016) and Meng et al. (2018) place participants in precisely this situation. The competing possible causal structures all indicate the same variable as the root node of the system. Children's preference for activating this node may therefore be interpreted as uninformative, since this intervention cannot distinguish between the competing hypotheses. Considered in light of the SI hypothesis, however, this choice need not indicate a weakness in self-directed hypothesis testing. All the hypotheses predict that intervening to activate the root node will lead to activation of the other variables in the system, and so the interventionist difference-making perspective may not meaningfully distinguish between these causal structures. Thus, the primary concern for a causal learner would be to assess the degree to which manipulation of this putative root cause reliably leads to its predicted effects, rather than to disambiguate exactly how it does so, which fits the behavior seen in the task.

Conclusion: Relationship to Truth

The claim that we are 'intuitive scientists' in our exploratory learning is well established as part of the account of human inquiry (Coenen et al., 2018). However, the fact that self-directed learners often choose to conduct repeated positive tests of their hypotheses, rather than (apparently) more informative interventions has historically complicated this claim. Repeated positive testing is characteristic of exploratory behavior across development, yet it would seem to be at odds with the aims of self-directed information seeking. In this chapter, we introduced a novel account of this behavior—the Search for Invariance (SI) hypothesis—which suggests that seeking multiple positive examples may in fact serve the information-seeking goals of causal learning.

To summarize, the SI hypothesis draws on the interventionist framework of causal reasoning, which suggests that causal learners are concerned with the *invariance* of

[9] As a reminder, the 'root node' is the starting point for the causal model of the system. Here, the structure is either a common cause (activating component A causes components B and C to activate) or a causal chain (activating A causes B to activate, which causes C to activate, etc.) meaning component A is the root node in both cases.

candidate hypotheses. In a probabilistic and interdependent causal world, our primary goal is to determine whether (and in what contexts) our current hypothesis provides an accurate basis for inference and intervention—not to disconfirm alternatives. By recognizing the central role of invariance in causal learning, positive testing may be reinterpreted as a rational and necessary information-seeking strategy. The SI hypothesis therefore provides an explanation of PTS that accords with theories that portray self-directed learning as intuitive science. Of course, empirical work is needed to establish the importance of invariance to learning, and to specify *how* learners form estimates of invariance from the multiple examples they generate, what type of examples are needed, and how many. By providing a novel approach to PTS, we hope that the SI hypothesis will serve as a promising theoretical foundation to guide future work.

That said, by aligning PTS with theories of the intuitive experimentation, it is also important to acknowledge that the 'learner-as-scientist' approach typically emphasizes increasing the *accuracy* of current knowledge as the primary goal of self-directed exploration. Indeed, Gopnik and colleagues (e.g., Gopnik, 1998, 2000; Gopnik & Walker, 2013) have variously asserted that the *raison d'etre* of our self-directed causal learning is to form veridical causal models of the world. According to this view, learning as an 'intuitive scientist' ought to be characterized by movement towards more accurate knowledge—and the proposal that causal learners are more concerned with assessing the invariance of a causal explanation than whether it is more accurate than alternatives may seem initially incompatible.

However, recall that the interventionist account of causality is inherently tied to action. As causal learners, we are concerned with refining the accuracy of our causal models, but *only* insofar as this meaningfully improves our ability to predict, explain, and manipulate the world. These goals do not require absolute accuracy; they require *reliability.* The SI hypothesis does not aim to imply that learners are disinterested in evaluating the truth of competing hypotheses, but that the concerns and priorities of causal learning will determine which aspects of these hypotheses are most informative to evaluate.

References

Baron, J., Beattie, J., & Hershey, J. C. (1988). Heuristics and biases in diagnostic reasoning: II. Congruence, information, and certainty. *Organizational Behavior and Human Decision Processes*, *42*(1), 88–110.

Blanchard, T., Vasilyeva, N., & Lombrozo, T. (2018). Stability, breadth and guidance. *Philosophical Studies*, *175*(9), 2263–2283.

Bonawitz, E. B., van Schijndel, T. J. P., Friel, D., & Schulz, L. (2012). Children balance theories and evidence in exploration, explanation, and learning. *Cognitive Psychology*, *64*(4), 215–234. https://doi.org/10.1016/j.cogpsych.2011.12.002

Bramley, N. R., Lagnado, D. A., & Speekenbrink, M. (2015). Conservative forgetful scholars: How people learn causal structure through sequences of interventions. *Journal of Experimental Psychology: Learning Memory and Cognition*. https://doi.org/10.1037/xlm0000061

Brewer, W. F., & Samarapungavan, A. (1991). Children's theories vs. scientific theories: Differences in reasoning or differences in knowledge?

Carey, S. (1985). *Conceptual change in childhood*. *The MIT series in learning development and conceptual change*. Cambridge, MA, US: MIT Press. https://doi.org/10.1016/S0016-6995(85)80176-5

Carey, S., Evans, R., Honda, M., Jay, E., & Unger, C. (1989). 'An experiment is when you try it and see if it works': A study of grade 7 students' understanding of the construction of scientific knowledge. *International Journal of Science Education*, *11*(5), 514–529. https://doi.org/10.1080/0950069890110504

Coenen, A., Nelson, J. D., & Gureckis, T. M. (2018). Asking the right questions about the psychology of human inquiry: Nine open challenges. *Psychonomic Bulletin & Review*, 1–41. https://doi.org/10.3758/s13423-018-1470-5

Coenen, A., Rehder, B., & Gureckis, T. M. (2015). Strategies to intervene on causal systems are adaptively selected. *Cognitive Psychology*, *79*, 102–133.

Cook, C., Goodman, N. D., & Schulz, L. E. (2011). Where science starts: spontaneous experiments in preschoolers' exploratory play. *Cognition*, *120*(3), 341–349. https://doi.org/10.1016/j.cognition.2011.03.003

Croker, S., & Buchanan, H. (2011). Scientific reasoning in a real-world context: The effect of prior belief and outcome on children's hypothesis-testing strategies. *British Journal of Developmental Psychology*, *29*(3), 409–424.

Devine, P. G., Hirt, E. R., & Gehrke, E. M. (1990). Diagnostic and confirmation strategies in trait hypothesis testing. *Journal of Personality and Social Psychology*, *58*(6), 952.

Dunbar, K., & Klahr, D. (1989). Developmental differences in scientific discovery processes. In D. Klahr & K. Kotovsky (Eds.), *Complex information: the impact of Herbert A. Simon.* (pp. 109–143). Hillsdale, N.J.: L. Erlbaum Associates. https://doi.org/10.1109/cipe.2004.1428147

Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgment: Persistence of the illusion of validity. *Psychological Review*, *85*(5), 395.

Friedman, M. (1974). Explanation and scientific understanding. *The Journal of Philosophy*, *71*(1), 5–19.

Friedrich, J. (1993). Primary error detection and minimization (PEDMIN) strategies in social cognition: A reinterpretation of confirmation bias phenomena. *Psychological*

Review, *100*(2), 298.
Gelman, S. A., Star, J. R., & Flukes, J. (2002). Children's use of generics in inductive inferences. *Journal of Cognition and Development*, *3*(2), 179–199.
Gerstenberg, T., Goodman, N., Lagnado, D. A., & Tenenbaum, J. (2014). From counterfactual simulation to causal judgment. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 36).
Gerstenberg, T., Peterson, M. F., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2017). Eye-tracking causality. *Psychological Science*, *28*(12), 1731–1744.
Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston, MA: Houghton Mifflin.
Gopnik, A. (1998). Explanation as Orgasm. *Minds and Machines*, *8*(1), 101–118. https://doi.org/10.1023/A:1008290415597
Gopnik, A. (2000). Explanation as orgasm and the drive for causal knowledge: The function, evolution, and phenomenology of the theory formation system. In *Explanation and cognition.* (pp. 299–323). Cambridge, MA, US: The MIT Press.
Gopnik, A., & Meltzoff, A. N. (1997). *Words, thoughts, and theories*. *Learning, development, and conceptual change*. Cambridge, MA, US: MIT Press. https://doi.org/10.1057/9781137322746
Gopnik, A., & Walker, C. M. (2013). Considering counterfactuals: The relationship between causal learning and pretend play. *American Journal of Play*.
Gorman, M. E., & Gorman, M. E. (1984). A comparison of disconfirmatory, confirmatory and control strategies on Wason's 2–4–6 task. *The Quarterly Journal of Experimental Psychology*, *36*(4), 629–648.
Gweon, H., & Schulz, L. E. (2008). Stretching to learn: Ambiguous evidence and variability in preschoolers exploratory play. *Proceedings of the 30th Annual Meeting of the Cognitive Science Society*.
Heyman, G. D., & Dweck, C. S. (1992). Achievement goals and intrinsic motivation: Their relation and their role in adaptive motivation. *Motivation and Emotion*, *16*(3), 231–247.
Hitchcock, C. (2012). Portable Causal Dependence: A Tale of Consilience. *Philosophy of Science*, *79*(5), 942–951. https://doi.org/10.1086/667899
Icard, T. F., Kominsky, J. F., & Knobe, J. (2017). Normality and actual causal strength. *Cognition*, *161*, 80–93.
Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence : an essay on the construction of formal operational structures*. New York: Basic Books.
Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: a theory of meaning, pragmatics, and inference. *Psychological Review*, *109*(4), 646.
Johnston, A. M., Sheskin, M., Johnson, S. G. B., & Keil, F. C. (2018). Preferences for explanation generality develop early in biology but not physics. *Child Development*, *89*(4), 1110–1119.
Karmiloff-Smith, A. (1988). The child is a theoretician, not inductivist. *Mind and Language*, *3*, 183–195.
Kendler, K. S. (2005). "A gene for…": the nature of gene action in psychiatric disorders. *American Journal of Psychiatry*, *162*(7), 1243–1252.
Kitcher, P. (1981). Explanatory unification. *Philosophy of Science*, *48*(4), 507–531.

Klahr, D., & Chen, Z. (2003). Overcoming the positive-capture strategy in young children: Learning about indeterminacy. *Child Development*, *74*(5), 1275–1296.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, *12*(1), 1–48.

Klahr, D., Fay, A. L., & Dunbar, K. (1993). Heuristics for Scientific Experimentation: A Developmental Study. *Cognitive Psychology*, *25*(1), 111–146. https://doi.org/10.1006/cogp.1993.1003

Klayman, J. (1995). Varieties of confirmation bias. In *Psychology of learning and motivation* (Vol. 32, pp. 385–418). Elsevier.

Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, *94*(2), 211.

Kuhn, D. (1989). Children and Adults as Intuitive Scientists. *Psychological Review*. https://doi.org/10.1037/0033-295X.96.4.674

Kuhn, D., Amsel, E., O'Loughlin, M., Schauble, L., Leadbeater, B., & Yotive, W. (1988). *The development of scientific thinking skills.* San Diego: Academic Press.

Kuhn, D., & Angelev, J. (1976). An experimental study of the development of formal operational thought. *Child Development*, 697–706.

Kuhn, D., & Brannock, J. (1977). Development of the isolation of variables scheme in experimental and" natural experiment" contexts. *Developmental Psychology*, *13*(1), 9.

Kuhn, D., & Phelps, E. (1982). The development of problem-solving strategies. *Advances in Child Development and Behavior*, *17*(C), 1–44. https://doi.org/10.1016/S0065-2407(08)60356-0

Lapidow, E., & Walker, C. M. (2019). Does the intuitive scientist conduct informative experiments?: Children's early ability to select and learn from their own interventions. In *41st Annual Meeting of the Cognitive Science Society*.

Lewis, D. (1974). Causation. *The Journal of Philosophy*, *70*(17), 556–567.

Lombrozo, T., & Carey, S. (2006). Functional explanation and the function of explanation. *Cognition*, *99*(2), 167–204. https://doi.org/10.1016/j.cognition.2004.12.009

Mackie, J. L. (1974). *The cement of the universe: A study of causation*. Oxford: Clarendon Press.

Mahoney, M. J., & DeMonbreun, B. G. (1977). Psychology of the scientist: An analysis of problem-solving bias. *Cognitive Therapy and Research*, *1*(3), 229–238.

McCormack, T., Bramley, N. R., Frosch, C., Patrick, F., & Lagnado, D. A. (2016). Children's use of interventions to learn causal structure. *Journal of Experimental Child Psychology*. https://doi.org/10.1016/j.jecp.2015.06.017

McKenzie, C. R. M. (2004). Hypothesis testing and evaluation. In D. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 200–219). Blackwell Publishing Ltd.

McKenzie, C. R. M., & Mikkelsen, L. A. (2000). The psychological side of Hempel's paradox of confirmation. *Psychonomic Bulletin & Review*, *7*(2), 360–366.

Meng, Y., Bramley, N. R., & Xu, F. (2018). Children's causal interventions combine discrimination and confirmation. In *Proceedings of the 40th Annual Conference of the Cognitive Science Society*.

Morris, A., Phillips, J. S., Icard, T. F., Knobe, J., Gerstenberg, T., & Cushman, F. (2018).

Causal judgments approximate the effectiveness of future interventions.
Navarro, D. J., & Perfors, A. F. (2011). Hypothesis generation, sparse categories, and the positive test strategy. *Psychological Review*, *118*(1), 120.
Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175–220.
Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, *101*(4), 608.
Pearl, J., & Bareinboim, E. (2011). Transportability of causal and statistical relations: A formal approach. In *Twenty-Fifth AAAI Conference on Artificial Intelligence*.
Popper, K. R. (1959). The Logic of Scientific Discovery. *Physics Today*. https://doi.org/10.1063/1.3060577
Redhead, M. (1987). Incompleteness, nonlocality, and realism: a prolegomenon to the philosophy of quantum mechanics.
Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926
Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*. https://doi.org/10.1016/S0010-0277(98)00075-4
Schauble, L. (1990). Belief revision in children: The role of prior knowledge and strategies for generating evidence. *Journal of Experimental Child Psychology*, *49*(1), 31–57. https://doi.org/10.1016/0022-0965(90)90048-D
Schauble, L., Glaser, R., Duschl, R. A., Schulze, S., & John, J. (1995). Students' understanding of the objectives and procedures of experimentation in the science classroom. *The Journal of the Learning Sciences*, *4*(2), 131–166.
Schauble, L., Klopfer, L. E., & Raghavan, K. (1991). Students' transition from an engineering model to a science model of experimentation. *Journal of Research in Science Teaching*, *28*(9), 859–882. https://doi.org/10.1002/tea.3660280910
Schulz, L. E. (2012). The origins of inquiry: Inductive inference and exploration in early childhood. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2012.06.004
Schulz, L. E., & Bonawitz, E. B. (2007). Serious Fun: Preschoolers Engage in More Exploratory Play When Evidence Is Confounded. *Developmental Psychology*. https://doi.org/10.1037/0012-1649.43.4.1045
Schulz, L. E., Standing, H. R., & Bonawitz, E. B. (2008). Word, Thought, and Deed: The Role of Object Categories in Children's Inductive Inferences and Exploratory Play. *Developmental Psychology*. https://doi.org/10.1037/0012-1649.44.5.1266
Schwartz, B. (1982). Reinforcement-induced behavioral stereotypy: How not to teach people to discover rules. *Journal of Experimental Psychology: General*, *111*(1), 23.
Siegler, R. S., & Liebert, R. M. (1975). Acquisition of formal scientific reasoning by 10- and 13-year-olds: Designing a factorial experiment. *Developmental Psychology*, *11*(3), 401.
Siler, S. A., & Klahr, D. (2012). Detecting, Classifying, and Remediating: Children's Explicit and Implicit Misconceptions about Experimental Design. In *Psychology of Science: Implicit and Explicit Processes*. https://doi.org/10.1093/acprof:oso/9780199753628.003.0007
Siler, S. A., Klahr, D., & Price, N. (2013). Investigating the mechanisms of learning from

a constrained preparation for future learning activity. *Instructional Science*, *41*(1), 191–216.
Skov, R. B., & Sherman, S. J. (1986). Information-gathering processes: Diagnosticity, hypothesis-confirmatory strategies, and perceived hypothesis confirmation. *Journal of Experimental Social Psychology*, *22*(2), 93–121.
Sloman, S. A. (2005). *Causal models : how people think about the world and its alternatives*. Oxford; New York: Oxford University Press.
Sloman, S. A., & Lagnado, D. A. (2005). Do We "do"? *Cognitive Science*, *29*(1), 5–39. https://doi.org/10.1207/s15516709cog2901_2
Sloman, S. A., & Lagnado, D. A. (2015). Causality in thought. *Annual Review of Psychology*, *66*, 223–247.
Sodian, B., Zaitchik, D., & Carey, S. (1991). Young children's differentiation of hypothetical beliefs from evidence. *Child Development*, *62*(4), 753–766.
Strevens, M. (2009). *Depth: An Account of Scientific Explanation*.
Tschirgi, J. E. (1980). Sensible Reasoning: A Hypothesis about Hypotheses. *Child Development*, *51*(1), 1–10. https://doi.org/10.2307/1129583
Tukey, D. D. (1986). A philosophical and empirical analysis of subjects' modes of inquiry in Wason's 2–4–6 task. *The Quarterly Journal of Experimental Psychology Section A*, *38*(1), 5–33.
Tweney, R. D., Doherty, M. E., Worner, W. J., Pliske, D. B., Mynatt, C. R., Gross, K. A., & Arkkelin, D. L. (1980). Strategies of rule discovery in an inference task. *Quarterly Journal of Experimental Psychology*, *32*(1), 109–123.
Valanides, N., Papageorgiou, M., & Angeli, C. (2014). Scientific Investigations of Elementary School Children. *Journal of Science Education and Technology*, *23*(1), 26–36. https://doi.org/10.1007/s10956-013-9448-6
van Schijndel, T. J. P., Visser, I., van Bers, B. M. C. W., & Raijmakers, M. E. J. (2015). Preschoolers perform more informative experiments after observing theory-violating evidence. *Journal of Experimental Child Psychology*, *131*, 104–119. https://doi.org/10.1016/j.jecp.2014.11.008
Vasilyeva, N., Blanchard, T., & Lombrozo, T. (2018). Stable causal relationships are better causal relationships. *Cognitive Science*, *42*(4), 1265–1296.
Vogel, R., & Annau, Z. (1973). AN OPERANT DISCRIMINATION TASK ALLOWING VARIABILITY OF REINFORCED RESPONSE PATTERNING 1. *Journal of the Experimental Analysis of Behavior*, *20*(1), 1–6.
Walker, C. M., Lombrozo, T., Legare, C. H., & Gopnik, A. (2014). Explaining prompts children to privilege inductively rich properties. *Cognition*, *133*(2), 343–357. https://doi.org/10.1016/j.cognition.2014.07.008
Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, *12*(3), 129–140.
Wason, P. C. (1962). Reply to wetherick. *Quarterly Journal of Experimental Psychology*, *14*(4), 250.
Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, *20*(3), 273–281.
Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content* (Vol. 86). Harvard University Press.
Wells, G. L., & Gavanski, I. (1989). Mental simulation of causality. *Journal of*

Personality and Social Psychology, *56*(2), 161.
Weslake, B. (2010). Explanatory depth. *Philosophy of Science*, *77*(2), 273–294.
Wetherick, N. E. (1962). Eliminative and enumerative behaviour in a conceptual task. *Quarterly Journal of Experimental Psychology*, *14*(4), 246–249.
Woodward, J. (1997). Explanation, invariance, and intervention. *Philosophy of Science*, *64*, S26–S41.
Woodward, J. (2003). *Making things happen : a theory of causal explanation*. New York: Oxford University Press.
Woodward, J. (2006). Sensitive and insensitive causation. *The Philosophical Review*, *115*(1), 1–50.
Woodward, J. (2010). Causation in biology: stability, specificity, and the choice of levels of explanation. *Biology & Philosophy*, *25*(3), 287–318.
Wu, R., Gopnik, A., Richardson, D. C., & Kirkham, N. Z. (2011). Infants Learn About Objects From Statistics and People. *Developmental Psychology*, *47*(5), 1220–1229. https://doi.org/10.1037/a0024023
Yang, S. C., Vong, W. K., Yu, Y., & Shafto, P. (2019). A unifying computational framework for teaching and active learning. *Topics in Cognitive Science*.
Ylikoski, P., & Kuorikoski, J. (2010). Dissecting explanatory power. *Philosophical Studies*, *148*(2), 201–219.
Yoon, E. J., MacDonald, K., Asaba, M., Gweon, H., & Frank, M. C. (2018). Balancing informational and social goals in active learning. In *Proceedings of the 40th Annual Meeting of the Cognitive Science Society*.
Zimmerman, C. (2000). The Development of Scientific Reasoning Skills. *Developmental Review*, *20*(1), 99–149. https://doi.org/10.1006/drev.1999.0497
Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, *27*(2), 172–223. https://doi.org/10.1016/j.dr.2006.12.001
Zimmerman, C., & Glaser, R. (2001). *Testing Positive Versus Negative Claims: A Preliminary Investigation of the Role of Cover Story on the Assessment of Experimental Design Skills. CSE Technical Report*.
Zimmerman, C., & Klahr, D. (2018). Development of Scientific Thinking. In J. Wixted (Ed.), *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience* (4th ed., pp. 1–25). New York: John Wiley & Sons, Inc. https://doi.org/10.1002/9781119170174.epcn407