Received: 9 May 2021 | Revised: 8 September 2021 | Accepted: 13 September 2021
DOI: 10.1002/wcs.1580

PERSPECTIVE

WIREs COGNITIVE SCIENCE WILEY

Rethinking the "gap": Self-directed learning in cognitive development and scientific reasoning

Elizabeth Lapidow | **Caren M. Walker**

Department of Psychology, University of California, San Diego, La Jolla, California, USA

Correspondence
Caren M. Walker, Department of Psychology, University of California, San Diego, La Jolla, CA, USA.
Email: carenwalker@ucsd.edu

Funding information
This work was supported by funding from the Hellman Fellowship, Jacobs Foundation, and National Science Foundation (CAREER award #2047581) to C. M. Walker, and from the National Defence Science and Engineering Graduate Fellowship to E. Lapidow.

Abstract
To improve upon their current knowledge, learners must be able to generate informative data and accurately evaluate this evidence. However, there is substantial disagreement regarding self-directed learners' competence in these behaviors. Researchers in cognitive development have suggested that learners are "intuitive scientists," generating informative actions and rationally coordinating their current observations and prior beliefs from an early age. Conversely, researchers in scientific reasoning report that learners struggle with experimentation and often fail to reach appropriate conclusions from evidence, even as adults. According to the prevailing narrative, these inconsistent findings must be "bridged" to explain the gap between learners' successes and failures. Here, we advocate for an alternative approach. First, we review the research on scientific reasoning and find that there may be less evidence for learners' failures than is typically assumed. Second, we offer a novel interpretation that aims to account for both literatures: we suggest that self-directed learners may be best understood as competent causal reasoners. That is, many seemingly uninformative or irrational behaviors are consistent with the goals of causal learning. This account not only resolves the apparent contradictions in the existing research, but also offers a way forward towards a more accurate and integrated understanding of self-directed learning.

This article is categorized under:
Psychology > Development and Aging
Psychology > Learning
Psychology > Reasoning and Decision Making

KEYWORDS
causal learning, cognitive development, scientific reasoning

1 | INTRODUCTION

For much of the last 20 years, researchers in psychology and education have discussed the importance of "bridging the gap" between apparently incompatible findings in cognitive development and scientific reasoning. Both literatures are primarily concerned with characterizing and understanding the inquiry and inference skills of self-directed learners (see Kuhn & Pearsall, 2000; Schulz, 2012). *Inquiry* refers to the ability to generate informative evidence, either in the course of unconstrained exploration (Chen & Klahr, 1999; Cook et al., 2011; Wason, 1960) or by choosing between discrete options (Lapidow & Walker, 2020b; Tschirgi, 1980). *Inference* is the ability to evaluate the informativeness of this evidence and draw appropriate conclusions (Koerber et al., 2005; Ruffman et al., 1993).

Researchers in cognitive development often describe self-directed learning as a kind of "intuitive science," in which learners generate informative actions to test and update their current beliefs (see Gopnik & Wellman, 2012; Schulz, 2012). Children and adults are characterized as intrinsically motivated to search for explanations that will support future learning and explore where their existing knowledge is uncertain or incomplete (Liquin & Lombrozo, 2020; Schulz & Bonawitz, 2007). Children make accurate causal inferences from patterns of data as early as 2–3 years (Gopnik et al., 2001; Schulz & Gopnik, 2004; Sobel & Kirkham, 2006) and can identify hypothesis tests that generate informative evidence by early grade school (Bullock & Ziegler, 1999).

In contrast, researchers examining the development of formal scientific reasoning skills report that self-directed learners often *fail* to conduct informative experiments or arrive at valid conclusions (see Zimmerman, 2007; Zimmerman & Klahr, 2018). Children and adults are described as consistently misunderstanding the purpose of inquiry, seeking instead to generate confirmatory, positive evidence (Klahr et al., 1993; Wason, 1960), and engineer desirable, tangible outcomes (Schauble et al., 1991). Not only do school-aged children struggle to make appropriate inferences from their observations (Chen & Klahr, 1999; Klahr et al., 1993) and fail to select informative hypothesis tests (Tschirgi, 1980), these errors appear to persist well into adulthood (Kuhn, 2007; Tschirgi, 1980).

Faced with this apparent contradiction, researchers in both traditions have attempted to "bridge the gap" between these conflicting characterizations. Here, we advocate for an alternative approach. First, we examine classic studies documenting young learners' scientific reasoning errors, revealing a tension between these and more recent findings within the same literature. We then review, in broad strokes, specific proposals to "bridge the gap." Specifically, researchers in scientific reasoning have suggested that early success is due to children's unconscious coordination of theory and evidence, which occurs without a genuine understanding of experimental logic (e.g., Kuhn, 2002). On the other hand, researchers in cognitive development argue that even young learners understand this logic, but lack the coordination and control to systematically employ this understanding (e.g., Bullock & Ziegler, 1999). Both approaches attempt to explain away findings generated by the other side as failing to provide evidence for *true* competence. We argue that neither approach provides a complete picture of the documented behavior, nor can they explain why certain systematic failures persist into adulthood.

Finally, we advance a novel account that interprets self-directed learning in terms of the particular goals and considerations of causal reasoning. In doing so, we illustrate how documented errors of inquiry and inference might be reconceived as both rational and informative. Thus, rather than approaching the available evidence from cognitive development and scientific reasoning as incompatible and in need of reconciliation, this account interprets the findings from both literatures as internally consistent evidence of competence in self-directed learning.

2 | EXAMINING THE GAP: EXISTING APPROACHES AND EVIDENCE

In this section, we highlight two major shortcomings of the prevailing "gap" narrative as an approach to understanding self-directed learning. First, in revisiting classic studies of scientific reasoning, we find that the failures documented are neither as extensive nor as extreme as often assumed, and argue that the field's commitment to classic interpretations may be impeding change to the prevailing narrative. Second, we examine proposals from both traditions, and argue that interpreting learners' behaviors in terms of success or failure may undermine the explanatory value of the existing theories as accounts of self-directed learning. Instead, we suggest that experimenters' assumptions about what "counts" as correct inquiry and inference behavior may have influenced their interpretation of participants' performance in prior work.

2.1 | The tension within scientific reasoning research

As described above, research into the development of scientific reasoning maintains that self-directed learners are not intuitively competent scientific thinkers. Beginning in the 1970s (e.g., Kuhn & Angelev, 1976; Kuhn & Brannock, 1977; Siegler & Liebert, 1975), this side of the "gap" consists of a tremendous body of research on assessing and improving early experimentation skills (see Zimmerman, 2000, 2007; Zimmerman & Klahr, 2018 for reviews). Examined closely, however, this research does not present a uniformly consistent account.

2.1.1 | Classic evidence of children's scientific errors

To begin, findings from early research in scientific reasoning remain widely cited and influential. However, the conclusions drawn from this work are not always well supported by their empirical evidence. For instance, a common example of learners' scientific errors is the tendency—particularly in childhood—to choose inquiries based on their tangible outcomes, rather than their information value (e.g., Croker & Buchanan, 2011; Kuhn & Phelps, 1982; Schauble, 1990; Schauble et al., 1995; Siler et al., 2013; Siler & Klahr, 2012; Zimmerman & Glaser, 2001). This claim largely originates with Tschirgi's (1980) assessment of logical hypothesis testing in children (2nd-, 4th-, and 6th-graders) and adults. The task presented scenarios in which the combination of three variables results in either a positive or negative outcome. For example, a character bakes a cake using one of two types of each of three ingredients (flour, sweetener, and shortening) for which he can choose between one of two types for each (e.g., honey or sugar for the sweetener). The cake comes out well and the character is described as believing that one variable (type of sweetener) caused this outcome, and that the other two variables (types of flour and shortening) were non-causal. Specifically, participants were told, "John thought that the reason the cake was good was the honey. He thought that the type of shortening (butter or margarine) or the type of flour really didn't matter. What should he do to prove this point?" They were then given a choice between three potential experiments: (1) changing the suspected cause and keeping the other two variables constant (VARY), (2) changing the other two variables and keeping the suspected cause constant (HOLD), or (3) changing all three variables.

Tschirgi (1980) treats VARY as the only option that *disconfirms* the suspected cause, and thus the only informative choice (see Section 3.2). The fact that both children and adults prefer this option *only* when the outcome is negative (i.e., a bad cake), and prefer HOLD when the outcome is positive, is interpreted as evidence that self-directed inquiry does not follow scientific logic. Critically, however, the task design does not necessarily justify this conclusion. First, the hypothesis presented to participants 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 shortening. Traditionally, (1) is treated as the only claim that participants are asked to test. However, nothing in the study's logic or instructions prevents participants from evaluating (2), for which HOLD, *not* VARY, is a disconfirming test. Thus, contrary to the author's interpretation, there is no single "correct" answer that offers a conclusive test of the hypothesis.

Despite this fact, subsequent research has expanded on Tschirgi's (1980) conclusion both empirically (e.g., Varma et al., 2018; Zimmerman & Glaser, 2001) and theoretically (Schauble, 1990; Schauble et al., 1991). For example, a recent study by Croker and Buchanan (2011) used a similar design to examine whether children's prior beliefs about the content of the problem (e.g., dental health) would influence their selection of hypothesis tests. This study not only fails to correct the issue described above, but it also brings a second issue from Tschirgi's (1980) original design to the foreground: the VARY option is confounded by the introduction of a new causal variable with an unknown effect. Specifically, the character believes that drink choice (milk or cola) causes the outcome (healthy or unhealthy teeth). Rather than removing the suspected variable, the VARY option replaces one variable with another, while holding the other two variables (brushing vs. not brushing teeth, going vs. not going to the dentist) constant. This replacement renders the VARY option inconclusive (e.g., in the case where milk is replaced with cola, there is no way to know if the result is due to the *absence* of a cause of healthy teeth or the *introduction* of a cause of unhealthy teeth). In fact, since the manipulations for the suspected non-causal variables are removals, rather than replacements, the "incorrect" HOLD option is actually *more* informative than the VARY option, since it offers an unconfounded test of one part of the hypothesis (i.e., milk leads to healthy teeth irrespective of brushing and dentist visits).

Continuing this reexamination, we find other instances of the potential mismatch between experimental intention and execution in classic scientific reasoning tasks. For example, Siegler and Liebert (1975) asked 10- and 13-year-olds to determine what arrangement of four binary switches causes a toy train to move. None of the 10-year-olds and only 10% of the 13-year-olds succeeded on this task without training. Critically, however, the criteria used to define systematic experimentation (i.e., generating an exhaustive set of combinations without repetition or error in the minimum number of actions) far exceed what is necessary for informative intervention. Similarly, Kuhn and Phelps (1982), asked 9–11-year-olds to identify the cause of a color-change reaction between chemicals. The experimenter first demonstrates that a combination of chemicals A, B, and C produces the reaction, and a combination of C, D, and E does not. Afterwards, children are given access to all five chemicals to explore. Here, "genuine, efficient, and sufficient experimentation" is excessively defined as combining individual chemicals with the activation agent. Not only is this unintuitive, since the experimenter's own demonstration combined multiple chemicals at once, it is also unnecessary, given that children already know that some of the chemicals (C, D, and E) are causally inert. Therefore, comparing the result of A, C, and D to the result of B, C, and D is not less efficient or informative than comparing A and B.

Of course, this early research represents only a small portion of the large body of work that makes up the field of scientific reasoning. However, the conclusions drawn from these early studies continue to be influential in shaping the prevailing perspective. Siegler and Liebert (1975) and Kuhn and Phelps (1982), for example, are commonly cited as evidence of experimentation failures in grade-schoolers (Brock & Taber, 2017; Lehrer & Schauble, 2012, 2015) and recent reviews of scientific reasoning highlight Tschirgi's (1980) task as a primary example of learners' errors in experimentation (Toplak et al., 2013; Zimmerman & Klahr, 2018). This continued emphasis on the early evidence for scientific errors may result in a less accurate overall assessment of the competence of self-directed inquiry and inference.

2.1.2 | Recent evidence of early scientific competence

In contrast to the failures reported in the classic work, studies of scientific reasoning conducted over the last two decades have found evidence for competence in elementary and grade-school-aged learners. Children grasp the goals of hypothesis testing (Chen & Klahr, 1999), understand the hypothesis-evidence relation (Koerber et al., 2005), reason appropriately about relevant and irrelevant characteristics of evidence (Masnick & Morris, 2008), and can even articulate an explicit understanding of the nature of science (Sodian et al., 2002). In all of these cases, the errors observed are modest. Even the youngest children tested appear to already grasp the fundamental concepts and skills of scientific inquiry and inference, and this is followed by steady, incremental progress towards consistently correct performance with increasing age.

What accounts for this difference? Despite never explicitly acknowledging the empirical problems in the classic work, modern scientific reasoning research provides far more intuitive and well-executed tests of children's abilities. For example, Chen and Klahr (1999) examined 2nd, 3rd, and 4th graders' self-directed experimentation of multivariate physics phenomena (i.e., springs, slopes, and sinking) before and after instruction in using the control of variables strategy. In each problem, children were given a number of materials to select and combine, and asked to design experiments to determine whether or not a specific variable made a difference to the outcome. The children tested were as young or younger than those in earlier research (e.g., Kuhn & Phelps, 1982; Tschirgi, 1980), and presented with a similar type of problem as in the classic work (determining the causal status of variables in a multivariate context). It is therefore striking that 15% of children performed perfectly prior to receiving any instruction. Furthermore, children almost always designed appropriately contrastive experiments (tests that differed on the target variable), and the proportion of unconfounded experiments (tests differing on the target variable while holding all others constant) was significantly above chance for all ages. In contrast with prior claims, this suggests that even the youngest children understood that the purpose of their experiments was to determine the effect of a target variable on the outcome. Although Chen and Klahr (1999) acknowledge that children's performance was better than expected, their focus is primarily on children's *improvement* following explicit instruction. As a result, the authors frame their findings as evidence for a lack of spontaneous success. This study (and others like it) has therefore largely reinforced the narrative that self-directed learners are both error-prone and biased.

2.2 | Explaining the gap: execution versus understanding

Despite growing evidence for children's competence, the prevailing narrative—that learners' failures on scientific reasoning tasks are *incompatible* with the successes documented by cognitive development research—remains (see Osterhaus et al., 2021; Shtulman & Walker, 2020 for recent examples; however, see Koslowski, 1996 for a notable exception). Maintaining this narrative has required researchers on both sides to redraw the demarcations between what is and is not evidence of successful inquiry and inference behavior.

Theories in the tradition of scientific reasoning, advanced by Kuhn and others, argue that children's early theory revision is an implicit process, lacking the conscious awareness, intent, and control required for adult-like scientific reasoning (Kuhn, 2002, 2012): Young children are described as thinking *with* their theories rather than *about* them. Their competence in cognitive development tasks is attributed to the unconscious coordination of theory and evidence, which supports early learning, but is not understood by the learner. It is not until much later, these accounts argue, that learners develop the metacognitive competencies required to recognize the relationship between their hypotheses and the evidence they observe, or to understand *why* they are changing their beliefs (Kuhn, 2002; Zimmerman & Klahr, 2018, but see Köksal et al., 2021).

In contrast, researchers in the tradition of cognitive development claim that children grasp the principles of scientifically rational inquiry and inference before they are able to employ them. For example, Bullock and Ziegler (1999) argue that learners first develop an understanding of what makes a good experiment, and only later acquire the general cognitive skills (e.g., working memory) needed to apply this understanding in order to systematically generate such tests. Other accounts of cognitive development offer similar explanations, maintaining that children understand and appreciate the principles tested in scientific reasoning, but struggle to demonstrate this capacity due to extraneous task demands. Recent examples of this argument variously posit the use of background knowledge (Weisberg et al., 2020) (see Box 1), opaque presentations of information (Köksal-Tuncer & Sodian, 2018), and explicit verbal responses (Lapidow & Walker, 2020b) as responsible for underestimating young learners' competence.

There is validity to both of these arguments. Belief revision in early learning *does* precede the development of metacognition, and the gradual development of executive function and verbal fluency limits what competencies can be measured in children's behavior. However, neither account is an entirely sufficient or satisfying explanation of self-directed learners' behavior. An obvious objection to both approaches is that appealing to cognitive immaturity cannot explain the persistence of systematic errors into adulthood. Indeed, there is ample evidence that adults—despite having fully developed executive abilities, metacognition, and verbal fluency—consistently err in scientific reasoning tasks of inquiry (e.g., Kuhn et al., 1995; Wason, 1960) and inference (e.g., Kuhn, 2007).

In what follows, we offer a novel alternative proposal that can explain self-directed learners' behavior as internally consistent. Since this account emphasizes learners' fundamental competence, it is more clearly aligned with the perspective of the "intuitive scientist" than the "error-prone scientific reasoner." However, unlike existing accounts, which seek to explain away what learners actually *do*—we aim to explain how the available evidence might fit together. We argue that abandoning the "gap" narrative may ultimately lead to a more accurate understanding of both children's and adults' self-directed inquiry and inference. Notably, our approach is largely consistent with Koslowski's (1996), who previously advocated for reexamining the scientific principles against which learners are evaluated. We expand upon this prior work by proposing a novel means for evaluation in terms of the goals of causal reasoning.

3 | CLOSING THE GAP: A NEW ACCOUNT

Here, we offer an account of self-directed learning that aims to avoid the limitations of existing explanations of learners' systematic errors. We propose that inquiry and inference behaviors are intuitively suited to support the goals of causal

BOX 1 The debate surrounding background knowledge

The effect of background knowledge on self-directed inquiry and inference is a point of theoretical and empirical debate between researchers in scientific reasoning and cognitive development. Both sides agree that learners' prior knowledge of the subject matter influences their learning process. Cognitive development researchers characterize learners' prior beliefs as critical support in constructing and revising hypotheses (e.g., Schulz, 2012). In contrast, scientific reasoning researchers describe learners as irrationally biased towards hypotheses that seem plausible or consistent with what they already know (e.g., Zimmerman & Klahr, 2018).

There is also disagreement about whether tasks eliminating or evoking prior knowledge are more accurate assessments of learners' competence. Cognitive researchers argue that the presence of prior knowledge may underestimate early ability. As a result, these tasks typically employ novel artificial stimuli ("knowledge-lean") to control for the possibility that real-world content will add unnecessary complexity, obscuring early competence. Researchers in scientific reasoning argue that this simplification overestimates ability, and that tasks embedded in real-world contexts about which participants have pre-existing beliefs ("knowledge-rich") are necessary for external validity.

Empirical evidence is not uniformly consistent with either claim: Children show competent inquiry and inference in causal learning tasks that challenge their prior knowledge (e.g., Bonawitz et al., 2012; Schulz, Bonawitz, & Griffiths, 2007) and also struggle with novel and decontextualized scientific reasoning tasks (e.g., Kuhn & Phelps, 1982). Furthermore, growing evidence from Köksal-Tuncer, Sodian, and colleagues (2018, 2021) suggests it may be possible to estimate formal scientific reasoning (such as metacognitive understanding of the relation between belief and evidence) using knowledge-lean designs.

learning, and that seemingly uninformative or irrational behaviors are appropriate when considered in light of the unique challenges and concerns of causal learners. We then illustrate this, using examples from adult behavior, since, as noted above, adults' errors in scientific reasoning present a significant challenge for current "gap" accounts.

3.1 | The interventionist theory of causal learning

Our proposal is grounded in the interventionist theory of causal explanation (Woodward, 1997) and the framework it provides for cognitive psychology (e.g., Campbell, 2007; Lombrozo, 2010) and causal learning (e.g., Schulz, Kushnir, & Gopnik, 2007; Sommerville, 2007). According to this account, causal knowledge consists of representations of the *counterfactual dependencies* between variables (i.e., knowing that *X* causes *Y*, means knowing that changing the value of *X* would result in a change in the value of *Y*). For example, if the hypothesis "coffee causes increased wakefulness" is true, then we would expect changes to the presence of coffee (*X*) to lead to consistent changes in the degree of wakefulness (*Y*) (Woodward, 1997, 2003). Causal representations not only allow us to reason about how one factor makes a difference to another, but also about the consequences of potential *actions* on those factors. These representations are also *generalizable*, holding over time and across superficial variations (Sloman, 2005). Thus, the value and purpose of our causal knowledge is that it provides learners with a basis for predicting, explaining, and manipulating events in new situations.[1]

The goal of causal learning then is to acquire representations of counterfactual dependencies that can be generalized to support action and inference. However, the nature of the causal world makes this a daunting task. Dependencies between events usually rely on the interaction of a variety of enabling and contributing factors (e.g., for coffee to cause increased wakefulness, the beverage needs to be caffeinated, consumed by someone not chronically sleep deprived, not contain excessive sugar, sleeping pills, etc.). Furthermore, learners are rarely in an epistemic position to accurately specify *all* of the relevant factors, and even when they are, such specific representations are ultimately less valuable, since they apply to fewer contexts. As a result, many of the dependencies between events that we care about must be represented in ways that do not hold with absolute consistency (e.g., "caffeine causes increased wakefulness" strikes us as superior to both "coffee causes increased wakefulness" and "caffeine that doesn't contain sleeping pills causes increased wakefulness").

To overcome these challenges, our causal representations include an understanding of the relative *invariance* of the dependency—that is, how reliably and broadly it can be generalized across contexts (Blanchard et al., 2018; Lapidow & Walker, 2020a; Woodward, 1997, 2006). Invariance, or "'stability," combines the notions of *breadth*, that is, the range of background circumstances in which a causal generalization holds, and *guidance*, that is, the accuracy of the generalization within those circumstances. Breadth favors more general representations, which can extend to more cases. Guidance favors more specific representations, correctly excluding cases in which the dependency will not hold (see work by Vasilyeva et al. (2018) and Woodward (2006) for discussion of these distinctions). Consideration of these dual components of invariance results in causal representations that strike a balance between applicability and accuracy, thus providing the most reliable basis for subsequent inference and action.

3.2 | Error in inquiry: positive hypothesis tests

A widely documented error in self-directed inquiry is learners' tendency to conduct *positive tests*—that is, tests expected to produce an effect if the current hypothesis is correct (Klayman, 1995). To illustrate, suppose you drop a glass on the floor, and it shatters. From this, you form the causal hypothesis that impact with a hard surface (*X*) causes glass to shatter (*Y*). Traditional falsificationist approaches dictate that the scientifically correct next step is to conduct disconfirming tests of this hypothesis (Popper, 1959): examining instances of *not-X* (e.g., dropping a glass on a *soft* surface) to determine whether *Y* (shattering) occurs. However, decades of research suggest that both children (Zimmerman, 2007) and adults (McKenzie, 2004) tend to favor positive tests (e.g., dropping a second glass on a similar surface), resulting in consistently incorrect performance in studies of hypothesis testing (Lapidow & Walker, 2020a).

A classic example of positive testing is the 2–4–6 task, first introduced by Wason (1960), and often used as evidence for the lack of rationality in adults' scientific inquiry (e.g., Gorman & Gorman, 1984; Mahoney & DeMonbreun, 1977; Tukey, 1986; Tweney et al., 1980; Wetherick, 1962). In this task, the experimenter first presents a sequence of numbers (e.g., "2, 4, 6") generated according to an unknown rule, which participants are tasked with determining. The participants can experiment by generating sequences and presenting them to the experimenter, who indicates whether or not it follows the rule. Most adult participants conduct only positive tests—generating sequences that follow the suspected rule (e.g., most hypothesize that the correct rule is "ascending even numbers," and test sequences like "10, 12, 14," "−2, 0, 2," "104, 106, 108," etc.)—treating the affirmative feedback as increasing evidence for their hypothesis. Although falsificationist principles require learners to rule out alternative hypotheses (e.g., "ascending numbers"), adults appear to operate under a misconception that verification alone provides conclusive evidence (Wason, 1960; Wason & Johnson-Laird, 1972).

Considered from a causal perspective, however, positive testing is uniquely informative because it is the *only* way to assess invariance. To return to our previous example, in order to determine whether and when *X* (e.g., impacting a hard surface) reliably brings about *Y* (e.g., shattering), it is necessary to *repeat X* (e.g., dropping glass on similarly hard surfaces) and observe whether *Y* occurs *again*. Just as positive tests cannot differentiate between alternative hypotheses, negative tests—which only serve to activate alternatives (*not-X*)—cannot, by definition, provide information about invariance about the current causal hypothesis (see Lapidow & Walker, 2020a, for a complete discussion of the "Search for Invariance" hypothesis). Similarly, by checking multiple instances in the 2–4–6 task that are consistent with their current hypothesis, learners can determine whether the rule invariantly identifies sets with the target property. This interpretation also explains adults' tendency to conduct *limit tests*: selecting extreme or unusual positive instances of their hypothesized rule (e.g., testing −2, 0, 2 as an instance of the rule "increasing even numbers") (Klayman, 1995; Klayman & Ha, 1987).

Thus, rather than stemming from a misconception about what constitutes conclusive evidence, positive tests may query the boundaries of a hypothesis, providing the learner with a critical understanding of the contexts in which it does and does not hold. Thus, self-directed learners' tendency to conduct positive tests can be understood as a rational and informative approach to generating evidence. This account can also be applied to explain various instances of positive testing errors in children's inquiry behavior (see Box 2).

3.3 | Error in inference: controlled comparison

The proposed causal framework also accounts for common errors of self-directed learners in their evaluation of evidence. The Fundraising Problem (Kuhn, 2007) is one such error, and also a typical example of the "gap" between cognitive development and scientific reasoning accounts. In this task, adults were asked to evaluate evidence about the effect of several factors on ticket sales at fundraisers (see Figure 1). The authors aimed to test whether learners could

BOX 2 Positive testing in children

Positive testing in children has been variously explained as stemming from their mistaken focus on generating positive evidence (i.e., experimenting to "demonstrate the correctness" of one's hypothesis, see Dunbar & Klahr, 1989) and positive outcomes (i.e., experimenting to generate tangible effects, see Section 2.1). For example, McCormack et al. (2016) presented 5–6-year-olds with a three component causal system and three competing hypotheses: a common cause (A activates B and C) and two causal chains (A activates B, which activates C; A activates C, which activates B). During exploration, the most common action was to activate component A, which children did repeatedly. This action generates the greatest number of tangible outcomes and provides positive evidence of all three hypotheses, but cannot distinguish among them.

Considered from the perspective of interventionist causal learning, however, this preference is not irrational. Recall that the value of causal knowledge lies in the fact that causal hypotheses are counterfactual (i.e., they indicate differences in effects given differences in causes) and generalizable, providing a guide for predicting and manipulating events. In McCormack et al. (2016) the competing hypotheses all indicate the *same* component (A) as the system's root cause, meaning, they make the same predictions for the majority of possible actions on the system. Thus, for a self-directed causal learner, assessing the reliability of this putative cause for producing its predicted effects can have as much information value as disambiguating among the possible mechanisms involved.

Club members are organizing fundraising parties for their cause. They've tried different combinations of features at the parties to see which sell the most tickets. Here's what they found:	
First Party: Door prizes; Comedian; Costumes	**SALES:** MEDIUM
Second Party: Door prizes; Auction; Costumes	**SALES:** HIGH
Third Party: Door prizes; Auction; Comedian; Costumes	**SALES:** HIGH

FIGURE 1 Information provided to participants in Kuhn (2007). Participants were asked: "Based on the evidence, does [feature] help ticket sales?" for each of the four features

Event:	Outcome:
Monkey sniffs at a vase with Flower A and Flower B.	Monkey sneezes.
Monkey sniffs at a vase with Flower B and Flower C.	Monkey sneezes.
Monkey sniffs at a vase with Flower A and Flower C.	No sneeze.

FIGURE 2 Description of the demonstration shown to participants in Schulz and Gopnik (2004). Following this demonstration, the participants were shown all three flowers and asked to give the experimenter the flower that makes monkey sneeze

recognize the *controlled comparison* (i.e., instances equated on all but one variable) about the effect of "auctions," provided by comparing Party 1 and Party 3. Adults' responses did not demonstrate comprehension of the controlled comparison data. Instead, the vast majority (83%) claimed that at least two factors had a positive causal influence on ticket sales, with nearly half (45%) claiming that three or all four did so.

In contrast, research from cognitive development suggests that preschool-aged children already understand the principles of controlled comparison. Schulz and Gopnik (2004) showed 3–5-year-olds a demonstration in which a character smells different combinations of three flowers, some of which make them sneeze (see Figure 2). The majority of children (79%) correctly identified Flower B as the causally effective variable. Existing "gap" accounts argue that, unlike children, adults fail to appreciate the controlled comparison in the Fundraising Problem, either because of a difference in task demands (e.g., children are asked to identify a single difference-making variable, while adults are asked about each factor in turn) or because adults' explicit coordination of theory and evidence is hampered by their prior beliefs (e.g., that door prizes lead to high ticket sales, see Kuhn, 2012).

Considered from a causal perspective, however, it is not apparent that the inference adults fail to make is a good, well-justified one. That is, while the evidence presented in the Fundraising Problem (see Figure 1) is consistent with the hypothesis that "auctions cause high ticket sales," it does not rule out the alternative hypotheses that auctions only produce this effect in the presence of door prizes, costumes, or both. These variables may be necessary background circumstances (like the lack of sleep deprivation for the dependency between caffeine and increased wakefulness), or high ticket sales may actually be caused by the interaction of all three variables. The existence of these alternatives *matter* to the causal learner—as the predictions, explanations, and actions supported by each possibility are *different*. It would be irrational, therefore, to dismiss these alternatives without sufficient evidence. This explains why children in Schulz and Gopnik (2004) succeeded at recognizing and utilizing controlled comparisons, while adults in Kuhn (2007) apparently failed: Children were presented with exhaustive evidence of all possible combinations of variables, but adults were not. Thus, the adults' conclusion that *multiple* variables were causal may indicate their sensitivity to the importance of context and the potential for combined effects.

4 | CONCLUSION

In this article, we advocate for an alternative to the "bridge the gap" narrative that has dominated discussions of self-directed inquiry and inference. First, we have shown that the general assumption of a "gap" is somewhat unfounded. Revisiting classic literature reveals that experimenters' early assumptions about what is required for scientifically rational behavior are often poorly matched to task designs, leading to an underestimation of learners' competence that has continued to define the prevailing narrative, despite more recent research. We then review the current theoretical attempts to "bridge the gap" between scientific reasoning and cognitive development and argue that these proposals have failed to satisfactorily capture the scope of self-directed learning behavior. Instead, researchers have attempted to *explain away* the evidence generated by the other side: that either the errors of scientific reasoning or the successes of cognitive development are not evidence of learners' true competence. Finally, we offer a coherent explanation of inquiry and inference that goes beyond previous calls to "bridge the gap" and instead aims to capture the actual behavior of learners documented in both literatures. Specifically, we demonstrate that errors observed in scientific reasoning tasks (e.g., positive testing and evidence evaluation) can be interpreted as rational approaches to generating information and interpreting evidence in the context of causal reasoning. Importantly, previous theories of the "gap" could not account for these errors, as they persist well into adulthood.

Two important points about this novel alternative account should be highlighted. First, there is still a need for empirical evidence to demonstrate that "errors" of self-directed inquiry and inference are driven by the goals and expectations of causal reasoning. Research is currently underway in our lab to examine whether self-directed learners are indeed sensitive to the potential invariance information offered by positive tests during their exploration of causal relationships. Second, the fact that learners do not *always* produce these errors suggests that self-directed learners are capable of pursuing other learning goals when directed. The difference between classic and more recent scientific reasoning research also supports this suggestion; task presentation seems to play an important role in determining participants' success in contexts that are unintuitive for causal learners.

Rather than aiming to *bridge* the gap between the intuitive scientist and the error-prone scientific reasoner, this proposal aims to *close* it. This account explains the apparent contradictions in the existing research by demonstrating that the goals of causal learning are *not* the goals typically assumed to be at play in studies of scientific reasoning. As noted above, we are not the first to make such an argument. Koslowski (1996) originally proposed that learners' apparently unscientific behavior may result from an incomplete or incorrect view of scientific thinking. To this, we add the novel proposal that self-directed learners may be most appropriately characterized as rationally seeking generalizable causal knowledge. This perspective, which draws on the philosophy of causality and explanation, offers a path towards a more coherent understanding of the intuitive learning process and formal scientific thinking. For example, this approach can be applied to explain the persistence of inaccurate intuitive theories in some domains of scientific content knowledge (see Shtulman, 2017; Shtulman & Walker, 2020). That is, these misconceptions occur in domains where learners hold an intuitive belief that supports predictions and inferences just as well as the true scientific explanation (see Box 2). Armed with the understanding that self-directed learners are intuitively guided by the goals of causal reasoning, we can arrive at a more complete and accurate account of inquiry and inference.

ACKNOWLEDGMENTS
The authors would like to thank Gail Heyman and Craig McKenzie for their insights, discussions, and feedback on these ideas.

CONFLICT OF INTEREST
The authors have declared no conflicts of interest for this article.

DATA AVAILABILITY STATEMENT
Data sharing is not applicable to this article as no new data were created or analyzed in this study.

AUTHOR CONTRIBUTIONS
Elizabeth Lapidow: Conceptualization (equal); investigation (lead); writing – original draft (lead). **Caren Walker:** Conceptualization (equal); project administration (lead); resources (lead); supervision (lead); writing – original draft (supporting); writing – review and editing (lead).

ORCID

Elizabeth Lapidow https://orcid.org/0000-0003-1290-6507
Caren M. Walker https://orcid.org/0000-0001-6532-8611

ENDNOTE

[1] The importance of this property of causal knowledge is also widely accepted outside of strictly interventionist accounts of causality and causal explanation. For example, Strevens (2011) theory of explanation emphasizes the importance of "robust" difference-making to explanatory practice, which is shaped by a concern for prediction and control. Further examples of similar concepts includes "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).

RELATED WIREs ARTICLES

The development of causal reasoning
Domains and naïve theories
Causal learning in children
Bayesian models of child development

FURTHER READING

Gopnik, A., & Schulz, L. E. (Eds.). (2007). *Causal learning: Psychology, philosophy, and computation*. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195176803.001.0001

Kuhn, D. (2011). What is scientific thinking and how does it develop?. U. Goswami, *The Wiley-Blackwell handbook of childhood cognitive development, second edition*. (497–523). Wiley-Blackwell. https://doi.org/10.1002/9781444325485.ch19

REFERENCES

Blanchard, T., Vasilyeva, N., & Lombrozo, T. (2018). Stability, breadth and guidance. *Philosophical Studies*, *175*(9), 2263–2283. https://doi.org/10.1007/s11098-017-0958-6

Bonawitz, E. B., van Schijndel, T. J. P., Friel, D., & Schulz, L. E. (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

Brock, R., & Taber, K. S. (2017). The application of the microgenetic method to studies of learning in science education: Characteristics of published studies, methodological issues and recommendations for future research. *Studies in Science Education*, *53*(1), 45–73. https://doi.org/10.1080/03057267.2016.1262046

Bullock, M., & Ziegler, A. (1999). Scientific reasoning: Developmental and individual differences. In F. E. Weinert & W. Schneider, *Individual development from 3 to 12: Findings from the Munich longitudinal study* (pp. 38–54). New York, NY: Cambridge University Press.

Campbell, J. (2007). An interventionist approach to causation in psychology. In A. Gopnik & L. E. Schulz, *Causal learning: Psychology, philosophy, and computation* (pp. 58–66). New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195176803.003.0005

Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development*, *70*(5), 1098–1120. https://doi.org/10.1111/1467-8624.00081

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. https://doi.org/10.1348/026151010X496906

Dunbar, K., & Klahr, D. (1989). Developmental differences in scientific discovery processes. D. Klahr & K. Kotovsky, *Complex Information Processing: The Impact of Herbert A. Simon*. (109–143). New York: Lawrence Erlbaum Associates, Inc. https://doi.org/10.4324/9780203761618

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*(5), 620–629. https://doi.org/10.1037/0012-1649.37.5.620

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, *138*(6), 1085–1108. https://doi.org/10.1037/a0028044

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. https://doi.org/10.1080/14640748408402183

Kendler, K. S. (2005). "A gene for…": The nature of gene action in psychiatric disorders. *American Journal of Psychiatry*, *162*(7), 1243–1252. https://doi.org/10.1176/appi.ajp.162.7.1243

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). https://doi.org/10.1016/S0079-7421(08)60315-1

Klayman, J., & Ha, Y. W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, *94*(2), 211–228. https://doi.org/10.1037/0033-295X.94.2.211

Koerber, S., Sodian, B., Thoermer, C., & Nett, U. (2005). Scientific reasoning in young children: Preschoolers' ability to evaluate covariation evidence. *Swiss Journal of Psychology*, *64*(3), 141–152. https://doi.org/10.1024/1421-0185.64.3.141

Köksal, Ö., Sodian, B., & Legare, C. H. (2021). Young children's metacognitive awareness of confounded evidence. *Journal of Experimental Child Psychology*, *205*, 105080. https://doi.org/10.1016/j.jecp.2020.105080

Köksal-Tuncer, Ö., & Sodian, B. (2018). The development of scientific reasoning: Hypothesis testing and argumentation from evidence in young children. *Cognitive Development*, *48*, 135–145. https://doi.org/10.1016/j.cogdev.2018.06.011

Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning*. MIT Press. https://doi.org/10.7551/mitpress/6758.001.0001

Kuhn, D. (2002). What is scientific thinking and how does it develop? In *Blackwell handbook of childhood cognitive development* (pp. 371–393). Wiley-Blackwell. https://doi.org/10.1002/9780470996652.ch17

Kuhn, D. (2007). Jumping to conclusions: Can people be counted on to make sound judgments?. *Scientific American Mind*, *18*(1), 44–51.

Kuhn, D., & Angelev, J. (1976). An experimental study of the development of formal operational thought. *Child Development*, *47*(3), 697–706. https://doi.org/10.2307/1128184

Kuhn, D. (2012). The development of causal reasoning. *WIREs Cognitive Science*, *3*(3), 327–335. https://doi.org/10.1002/wcs.1160

Kuhn, D., & Brannock, J. (1977). Development of the isolation of variables scheme in experimental and "natural experiment" contexts. *Developmental Psychology*, *13*(1), 9–14. https://doi.org/10.1037/0012-1649.13.1.9

Kuhn, D., Garcia-Mila, M., Zohar, A., & Andersen, C. (1995). Strategies of knowledge acquisition. In *Monographs of the society for research in child development*. *60*(4), v–128. https://doi.org/10.2307/1166059

Kuhn, D., & Pearsall, S. (2000). Developmental origins of scientific thinking. *Journal of Cognition and Development*, *1*(1), 113–129. https://doi.org/10.1207/S15327647JCD0101N_11

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. (2020a). The search for invariance: Repeated positive testing serves the goals of causal learning. In J. B. Childers, *Language and concept acquisition from infancy through childhood* (pp. 197–219). Springer. https://doi.org/10.1007/978-3-030-35594-4_10

Lapidow, E., & Walker, C. M. (2020b). Informative experimentation in intuitive science: Children select and learn from their own causal interventions. *Cognition*, *201*, 104315. https://doi.org/10.1016/j.cognition.2020.104315

Lehrer, R., & Schauble, L. (2012). Seeding evolutionary thinking by engaging children in modeling its foundations. *Science Education*, *96*(4), 701–724.

Lehrer, R., & Schauble, L. (2015). The development of scientific thinking. In L. S. Liben, U. Müller & R. M. Lerner, *Handbook of child psychology and developmental science: Cognitive processes* (pp. 671–714). John Wiley and Sons, Inc. https://doi.org/10.1002/9781118963418.childpsy216

Lewis, D. (1974). Causation. *The Journal of Philosophy*, *70*(17), 556–567. https://doi.org/10.2307/2025310

Liquin, E. G., & Lombrozo, T. (2020). A functional approach to explanation-seeking curiosity. *Cognitive Psychology*, *119*, 101276. https://doi.org/10.1016/j.cogpsych.2020.101276

Lombrozo, T. (2010). Causal–explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions. *Cognitive Psychology*, *61*(4), 303–332. https://doi.org/10.1016/j.cogpsych.2010.05.002

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

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. https://doi.org/10.1007/BF01186796

Masnick, A. M., & Morris, B. J. (2008). Investigating the development of data evaluation: The role of data characteristics. *Child Development*, *79*(4), 1032–1048. https://doi.org/10.1111/j.1467-8624.2008.01174.x

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*, *141*, 1–22. 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.

Osterhaus, C., Brandone, A. C., Vosniadou, S., & Nicolopoulou, A. (2021). Editorial: The emergence and development of scientific thinking during the early years: Basic processes and supportive contexts. *Frontiers in Psychology*, *12*, 629384. https://doi.org/10.3389/fpsyg.2021.629384

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*, *12*, 53–54. https://doi.org/10.1063/1.3060577

Redhead, M. (1987). *Incompleteness, nonlocality, and realism: A prolegomenon to the philosophy of quantum mechanics*. Clarendon Press.

Ruffman, T., Perner, J., Olson, D. R., & Doherty, M. (1993). Reflecting on scientific thinking: Children's understanding of the hypothesis-evidence relation. *Child Development*, *64*(6), 1617–1636. https://doi.org/10.2307/1131459

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. https://doi.org/10.1207/s15327809jls0402_1

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*, *16*, 382–389. 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*, *34*, 1045–1050. https://doi.org/10.1037/0012-1649.43.4.1045

Schulz, L. E., Bonawitz, E. B., & Griffiths, T. L. (2007). Can being scared cause tummy aches? Naive theories, ambiguous evidence, and preschoolers' causal inferences. *Developmental Psychology*, *43*(5), 1124–1139. https://doi.org/10.1037/0012-1649.43.5.1124

Schulz, L. E., & Gopnik, A. (2004). Causal learning across domains. *Developmental Psychology*, *40*(2), 162–176. https://doi.org/10.1037/0012-1649.40.2.162

Schulz, L. E., Kushnir, T., & Gopnik, A. (2007). Learning from doing: Intervention and causal inference. In A. Gopnik & L. E. Schulz, *Causal learning: Psychology, philosophy, and computation* (pp. 67–85). New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195176803.003.0006

Shtulman, A. (2017). *Scienceblind: Why our intuitive theories about the world are so often wrong*. Basic Books.

Shtulman, A., & Walker, C. (2020). Developing an understanding of science. *Annual Review of Developmental Psychology*, *2*(1), 111–132. https://doi.org/10.1146/annurev-devpsych-060320-092346

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), 401–402. https://doi.org/10.1037/h0076579

Siler, S. A., & Klahr, D. (2012). Detecting, classifying, and remediating: Children's explicit and implicit misconceptions about experimental design. R. W. Proctor & E. J. Capaldi, *Psychology of Science: Implicit and Explicit Processes*. (137–180). Oxford University Press. 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. https://doi.org/10.1007/s11251-012-9224-7

Sloman, S. A. (2005). *Causal models: How people think about the world and its alternatives*. Oxford University Press.

Sobel, D. M., & Kirkham, N. Z. (2006). Blickets and babies: The development of causal reasoning in toddlers and infants. *Developmental Psychology*, *42*(6), 1103–1115. https://doi.org/10.1037/0012-1649.42.6.1103

Sodian, B., Thoermer, C., Kircher, E., Grygier, P., & Günther, J. (2002). Vermittlung von Wissenschaftsverständnis in der Grundschule. In M. Prenzel & J. Doll, *Bildungsqualität von Schule: Schulische Und Außerschulische Bedingungen Mathematischer, Naturwissenschaftlicher Und Überfachlicher Kompetenzen* (pp. 192–206). Beltz.

Sommerville, J. A. (2007). Detecting causal structure: The role of interventions in infants' understanding of psychological and physical causal relations. A. Gopnik & L. E. Schulz, *Causal Learning: Psychology, Philosophy and Computation*, 48–57, New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195176803.003.0004

Strevens, M. (2011). *Depth: An account of scientific explanation*. Harvard University Press.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Assessing the development of rationality. In H. Markovits, *The developmental psychology of reasoning and decision-making* (pp. 7–35). Psychological Press. https://doi.org/10.4324/9781315856568

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. https://doi.org/10.1080/14640748608401583

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. https://doi.org/10.1080/00335558008248237

Varma, K., Van Boekel, M., & Varma, S. (2018). Middle school Students' approaches to reasoning about disconfirming evidence. *Journal of Educational and Developmental Psychology*, *8*(1), 1–28. https://doi.org/10.5539/jedp.v8n1p28

Vasilyeva, N., Blanchard, T., & Lombrozo, T. (2018). Stable causal relationships are better causal relationships. *Cognitive Science*, *42*(4), 1265–1296. https://doi.org/10.1111/cogs.12605

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, *12*(3), 129–140. https://doi.org/10.1080/17470216008416717

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content* (Vol. 86). Harvard University Press.

Weisberg, D. S., Choi, E., & Sobel, D. M. (2020). Of blickets, butterflies, and baby dinosaurs: Children's diagnostic reasoning across domains. *Frontiers in Psychology*, *11*, 2210. https://doi.org/10.3389/fpsyg.2020.02210

Weslake, B. (2010). Explanatory depth. *Philosophy of Science*, *77*(2), 273–294. https://doi.org/10.1086/651316

Wetherick, N. E. (1962). Eliminative and enumerative behaviour in a conceptual task. *Quarterly Journal of Experimental Psychology*, *14*(4), 246–249. https://doi.org/10.1080/17470216208416542

Woodward, J. (1997). Explanation, invariance, and intervention. *Philosophy of Science*, *64*, S26–S41. https://doi.org/10.1086/392584

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. New York: Oxford University Press. https://doi.org/10.1093/0195155270.001.0001

Woodward, J. (2006). Sensitive and insensitive causation. *The Philosophical Review*, *115*(1), 1–50. https://doi.org/10.1215/00318108-2005-001

Ylikoski, P., & Kuorikoski, J. (2010). Dissecting explanatory power. *Philosophical Studies*, *148*(2), 201–219. https://doi.org/10.1007/s11098-008-9324-z

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). John Wiley & Sons, Inc. https://doi.org/10.1002/9781119170174.epcn407

How to cite this article: Lapidow, E., & Walker, C. M. (2021). Rethinking the "gap": Self-directed learning in cognitive development and scientific reasoning. *Wiley Interdisciplinary Reviews: Cognitive Science*, e1580. https://doi.org/10.1002/wcs.1580