Review

A Model of Scientific Data Reasoning

Amy M. Masnick [1],[*] and Bradley J. Morris [2]

[1] Psychology Department, Hofstra University, Hempstead, NY 11549, USA
[2] Educational Psychology Department, Kent State University, Kent, OH 44242, USA; bmorri20@kent.edu
[*] Correspondence: amy.m.masnick@hofstra.edu

Abstract: Data reasoning is an essential component of scientific reasoning, as a component of evidence evaluation. In this paper, we outline a model of scientific data reasoning that describes how data sensemaking underlies data reasoning. Data sensemaking, a relatively automatic process rooted in perceptual mechanisms that summarize large quantities of information in the environment, begins early in development, and is refined with experience, knowledge, and improved strategy use. Summarizing data highlights set properties such as central tendency and variability, and these properties are used to draw inferences from data. However, both data sensemaking and data reasoning are subject to cognitive biases or heuristics that can lead to flawed conclusions. The tools of scientific reasoning, including external representations, scientific hypothesis testing, and drawing probabilistic conclusions, can help reduce the likelihood of such flaws and help improve data reasoning. Although data sensemaking and data reasoning are not supplanted by scientific data reasoning, scientific reasoning skills can be leveraged to improve learning about science and reasoning with data.

Keywords: data reasoning; scientific reasoning; statistics education; numerical cognition; cognitive development; number sense

1. Introduction

Data reasoning is a critical skill in scientific reasoning. Although evidence evaluation is a step in many models of scientific reasoning (e.g., [1–3]), there has been much less attention on the interpretation of numerical data itself within this context, which has been investigated largely within the field of statistics [4]. We outline a model of data reasoning that describes how data sensemaking underlies data reasoning (both defined below). We further suggest that scientific data reasoning differs from both informal data reasoning and data sensemaking. We use the phrase scientific data reasoning to refer to a set of skills that help reasoners improve the quality of their data analysis and interpretation, which improves the quality of inferences that can be drawn from data. Although these skills are most commonly used in scientific reasoning contexts, they can be used in any context. Scientific data reasoning includes a set of skills that help to harness data sensemaking and strengthen everyday data reasoning by improving the systematicity of data collected via the scientific method, the quality of analysis via statistical tools, and inferences by reducing cognitive bias and providing tools for evaluating conclusions.

Science refers to both a body of knowledge and the processes that create and evaluate this knowledge [5]. These processes include generating and testing hypotheses, acquiring data, and evaluating theories with new data [6–8]. The cornerstone of the scientific process is the reliance on empirical data that are formally analyzed [5,8]. Research that has focused on understanding the cognitive processes that underlie scientific reasoning includes studies of generating hypotheses [1,5,9,10], making predictions [11], deciding how to measure variables [12–14], and interpreting data in light of theory and prior beliefs [2,3,15,16]. We focus on one area that has gotten less attention, specifically on how people make sense of numerical data. Grolemund and Wickham [4] propose a cognitive interpretation of

data analysis, considering the sensemaking process of data, focused on how we reason with quantitative measurements. We argue that data reasoning is a unique skill, built on intuition and knowledge about math and statistics, and that a full understanding of scientific reasoning requires understanding of data reasoning as a core competency. In turn, this understanding has implications for teaching of these skills, and for future research directions.

We suggest that scientific reasoning is a specific type of information seeking that emerges from an acquired set of cultural tools and knowledge that augment general cognitive mechanisms (e.g., encoding, strategy use) and dispositions (e.g., curiosity [5,7,17,18]). Scientific reasoning is defined in multiple ways. Zimmerman and Klahr [5] describe scientific thinking as a multi-faceted, multi-level process that includes problem solving and reasoning, knowledge seeking, curiosity refined by science education that teaches hypothesis testing and evidence evaluation, and the development of metacognitive skills. Englemann et al. [19] describe scientific reasoning conceptualizations as falling into discussions of the process of scientific discovery, scientific argumentation, and understanding the nature of science. We position ourselves within Englemann et al.'s description of the process of scientific discovery, specifically focused on the evidence evaluation described by Zimmerman and Klahr, while still shaped by teaching, hypothesis testing, and general reasoning.

Data reasoning is situated within this definition in multiple places. One, informal data reasoning occurs frequently in daily life and requires no formal instruction (e.g., [19,20]). This is the process of drawing inferences from everyday data, requiring only general-purpose mechanisms for their operation (e.g., strategy acquisition and selection). Two, cultural tools relevant to data include number systems, mathematical formalisms, and data visualization tools (e.g., [21]), as well as the scientific process that produces (e.g., experimentation, data recording, as noted by many researchers, including [5,22]) and allows for analysis of data (e.g., inferential statistics, [4]). The acquisition of these formal tools improves the accuracy of data reasoning and improves the basic reasoning processes as well. Thus, we argue that understanding the cognitive processes underlying data reasoning is an important part of a fuller understanding of scientific reasoning.

We define data sensemaking as detecting trends or differences in sets of data without the use of formal analytical tools [4,23,24]. We draw evidence from perception [25] and numerical cognition [26] to suggest that data sensemaking relies on the same perceptual mechanisms that allow people to quickly aggregate and summarize complex information from the environment [27]. The perceptual processes and intuitions in data sensemaking work impressively well for detecting strong patterns in data without formal analysis (e.g., [3,28,29]). Detecting features in data provides information from which people can draw inferences and interpret the meaning of these features and patterns.

We define data reasoning as reasoning in which quantitative data are used as evidence for drawing conclusions or making decisions [30]. Data reasoning is a deliberate cognitive process that is limited by our memory [31] and our cognitive biases [4]. We argue that data sensemaking is a precursor to data reasoning, in that first the data are summarized (often via automatic perceptual processes), and then the summaries are used to draw inferences. Everyday data reasoning is common, though such inferences may fall prey to cognitive biases such as the tendency to confirm one's previous beliefs (i.e., confirmation bias, [32]).

Data reasoning occurs informally in many contexts. For example, when comparing the performance of players for a fantasy sports league, people do not need to conduct an experiment or test hypotheses. Instead, they are more likely to pay attention to features of the data, such as means or medians, variation, and trends, to guide their inferences or decisions (e.g., [20,30,33–36]). Researchers offer varied definitions of informal inferential reasoning (what we are calling data reasoning), but Makar and Rubin [30] highlight five key components that help describe the space: (1) informal reasoning makes a claim beyond the data; (2) conclusions are expressed with uncertainty; (3) data are used as evidence (i.e., the conclusions cannot be derived solely from prior beliefs, though the data can be considered

in theoretical contexts); (4) the data are considered in aggregate, and the reasoning stems from assessments of the aggregate data; and (5) context is considered.

In contrast, when engaging in scientific data reasoning, researchers use some of the same techniques as in data reasoning, but add safeguards to limit biases, which lead to the more deliberative scientific reasoning process. For example, scientists aim to ground research questions in theories backed by past evidence (e.g., [1,8,37]), design studies that measure and control variables to limit confounds (e.g., [38]), use external representations to represent a larger quantity of data at once (e.g., [21]), and apply formal statistical analyses to provide quantitative evidence when making inferences (e.g., [4]).

Our definitions of both scientific reasoning and scientific data reasoning are situated at a relatively course grain size. One piece of every model of scientific reasoning involves evidence evaluation, and that evidence evaluation typically includes evaluation and interpretation of quantitative data. Our model of data reasoning is at a somewhat more detailed level, including descriptions of component processes at a finer grain size than general models of scientific reasoning. One example of a description on a fine grain size is the process of summarizing data in sets. We propose that the same processes that underlie summarization in perceptual sets (e.g., dots) operate upon sets of numbers. These processes provide summaries that set up inferences from data. Thus, we provide descriptions at varying grain sizes across the scope of the review.

In this paper, we suggest a model of the cognitive processes underlying how people make sense of and draw inferences from data. We further suggest that, as people learn about science, they acquire tools to improve both of these processes. Our review targets cognitive processes as described from general Cognitive Science framework. We describe scientific reasoning as a process that includes declarative (e.g., scientific facts) and procedural (e.g., conducting unconfounded experiments) elements. We screened the literature to find research that was relevant to such a cognitive process model. We did not perform a literature search by term because the same term can have quite different meanings in different research traditions. Even within the community of scholars who study scientific reasoning, there is little consistency in the terminology that is used to describe it. For example, the special issue uses the term "competencies" [39] to describe what we and our colleagues often refer to as "processes" (e.g., [40]). Another reason for our selections is that we have attempted to bridge multiple literatures that have not regularly communicated, such as cognitive scientists, statisticians, and science educators. As Fischer et al. [6] note, "contemporary knowledge about what constitutes these competencies . . . is scattered over different research disciplines" (p. 29).

In sum, we propose that numerical data reasoning is rooted in intuitions about number sets and becomes more sophisticated as people acquire scientific and statistical reasoning skills. Although there are many types of nonquantitative data used in both everyday and scientific reasoning, in this paper, we focus exclusively on numerical data reasoning, and hereafter use the term data reasoning to refer to this type of reasoning. We argue that numerical data reasoning begins with data sensemaking, a largely intuitive process that summarizes sets of numbers much like summaries of perceptual features (e.g., relative size). Data sensemaking creates approximations of data rapidly. Data reasoning, or drawing conclusions from these data, is derived from these data summaries. Although these processes are fast and accurate given clear patterns or differences, both may be sensitive to cognitive limitations such as confirmation bias. Scientific data reasoning augments these informal processes by adding cultural tools for improving the accuracy of data gathering, representation, analysis, and inferences. We summarize our proposed model in Table 1.

Table 1. Summary of data sensemaking and data reasoning processes.

	Processes	Examples	Key References
Data Sensemaking	Summarization	Product of perceptual and cognitive mechanisms Implicit grouping of numbers that yields summary values (e.g., mean, variance)	[25,27,28,41–43]
Data Reasoning	Detecting Patterns	Detecting covariation between variables	[3,11,44,45]
	Detecting Differences	Noticing differences between sets	[28,41–43,46]
Scientific Data Reasoning	External Representation	External representations	[21,47,48]
	Scientific hypothesis testing	Conducting unconfounded experiments	[38,49–52]
	Probabilistic Conclusions	Evaluating the likelihood of conclusions or inferences	[4,20,53]
Limits to Data sensemaking and reasoning	Heuristics and biases	Confirmation bias, Anchoring effect	[15,32,54,55]
Sources of Change	Strategies	Acquiring better strategies for summarization, reasoning	[41,44,56,57]
	Instruction	Using data sensemaking to support formal reasoning and analysis	[58–60]

2. Data Sensemaking

We begin with a discussion of how data sensemaking occurs. As we will detail below, data sensemaking is the summarization of numerical information, a product of perceptual and cognitive mechanisms that summarize large quantities of information [25]. Numbers have unique properties such as relative magnitudes that are represented in both an approximate and exact fashion. Even young children have some elements of number sense that allows them to detect differences between quantities [61]. When these summarization mechanisms operate on number sets, they yield approximate representations of a set's statistical properties [25,28]. The following sections will outline the evidence for this account of data sensemaking, how it allows for the extraction of central tendency and variability, and how it changes over the course of development.

2.1. Sensemaking of Set Means and Variance

When people see a set of numerical data, they can summarize the data without using calculations, yielding approximate set properties such as means and variance [62]. Decades of evidence demonstrate that the properties of number sets can be summarized quickly [26,63] and accurately [43,64]. Without computation, people can detect and generate approximate means [41,43,65–70], detect relative variance [28,69,71], and increase their confidence in conclusions from larger samples as compared to smaller ones [42,43,68,69,71–74]. In one example of this early work with adults, participants were given a series of index cards each with a two-digit number, and asked to generate "a single value that best represented the series presented" [69] (p. 318). Participants generated a value that deviated from the arithmetic mean by less than 1%. When asked to estimate the mean from a set of numbers, and explicitly instructed not to calculate, participants were surprisingly accurate in generating an approximate mean (within ~3% of actual mean; [65]).

How might this process occur? We suggest that children and adults quickly summarize the properties of number sets similarly to how they summarize other types of complex information in their environments. This research tradition includes work from Gestalt psychologists [75] and recent research on ensemble perception and cognition [25,27]. Number sets may be summarized "automatically" [76] in that this may occur before conscious processing occurs (less than 100 MS; [77]) and even when instructions prohibit such processing [78]. Finally, reaction times are faster with larger sets than smaller sets with no loss in accuracy [76,79], further suggesting that summaries are the result of an automatic process.

Young children can summarize complex perceptual information, even in infancy [80]. However, summarization becomes more precise over the course of development [41,81]. For example, 6-month-olds can distinguish sets of dots with a 2:1 ratio (e.g., 10 from 20; [82]), while 6-year-olds can distinguish sets of dots at a 6:7 ratio [83]. Children as young as six can summarize the average happiness of a set of faces, but their summaries are less accurate than those of adolescents [84]. Given a set of objects (e.g., oranges), 4-year-olds can summarize the average size, though not as accurately as adults [81]. These findings suggest that summarization abilities emerge early in development and become refined over time.

Another numerical set characteristic is variability. The critical role of variability in empirical investigations has been noted for decades (e.g., [85]) and recently, Lehrer et al. [12] argued that variability is one of the fundamental issues it is necessary for students to understand when reasoning effectively about science. Functionally, it is only possible to measure the variability of a set of data, not a single data point; there must be data that vary to measure variability [86]. In considering what sets statistical reasoning apart from mathematical reasoning, Cobb and Moore [87] argue that although mathematical principles underlie many parts of statistical reasoning, it is the presence of variability due to real-world context that makes it statistical.

By first grade, children show an understanding that different variables are likely to differ in their variability when examining a data set [73], suggesting a conceptual understanding of underlying reasons to expect variation. Similarly, lay adults can demonstrate the ability to use variability in comparing data sets when the data are contextualized within a story, suggesting the likelihood of variability or not [71]. These findings indicate an expectation of variation when taking measurements from a heterogeneous sample, suggesting an understanding that variation is common in many contexts.

One component of understanding variability involves understanding the value of repeated measurements; without repeated measurements, there is nothing that can vary. Surveying 11-, 13-, and 15-year-olds indicated many areas of both clarity and confusion about experimental error, as well as the value of repeated measurements [88]. Although most students believed it was necessary to repeat measurements in science experiments, approximately 40% of participants in each age group focused solely on the means, and said the data with different variance levels were equally reliable because of the same average value, ignoring the variance. Detecting variability is also closely related to children's emerging understanding of the sources of this variability. Children understand by about age eight that measurement error is possible, and that repeated measurements, therefore, might not yield precisely the same results [89]. Children, especially 8-year-olds, were still not that likely to refer to measurement error in justifying their reasoning.

Other work has demonstrated that children and adults respond to variability information differently when they expect variability based on the context. For example, children ages 10–14 had an easier time using data to revise a belief from a noncausal one to a causal one [90]; the key complication in reasoning about noncausal links was in understanding measurement error and the value of repeated measurements to improve accuracy of estimations about data sets. In a converging set of findings, children ages 10–11 who expected a pattern of results indicating differences between conditions (such as in whether the length of a string affects the speed of a pendulum) were able to differentiate the small variance of noise from the larger differences between conditions [91]. At the same time, these children struggled more when they expected a difference but there was no true effect, and only small differences due to repeated measures. This point also emphasizes the close connection between data reasoning and scientific reasoning. For example, correctly interpreting data might hinge on recognizing the possibility of measurement error.

Sample size is also linked with reasoning about variability in data. There is a lot of evidence that people are more confident with larger samples of data (e.g., [68,69]). Further, many studies have found an interaction between sample size and variability when both are manipulated within the same study (e.g., [28,68,71,73]). For example, Jacobs and Narloch report that when samples had low variability, participants did not differentiate between

samples of 3 and 30, whereas in high-variability samples, even 7-year-olds responded differently based on sample size, and there were no age differences in the use of sample size. At the same time, there is evidence of failure to use sample size consistently in some contexts (e.g., [42,71]). More recent work has tried to reconcile apparently contradictory work about people's ability to use sample size in data reasoning, arguing that in fact weighted sample size follows a curvilinear function [74]. That is, with small sample sizes, participants are sensitive to sample size differences, but with large sample sizes, participants are no longer as sensitive to such differences and weight the differences much less. This finding suggests that numerical representations can affect broader data reasoning skills. Further, Obrecht found that intuitive number sense was also linked with the use of sample size, suggesting this factor may also play a role [74].

In addition to studies of implicit reasoning about variance, there are also several studies that have demonstrated that when children are asked to collect or are given data and asked to develop their own ways of summarizing the information, they can develop measures of center and variability that make sense to them. Additional design studies have focused on the integration of variation into describing data. For example, in figuring out how to display plant growth over two weeks of measurements, students had to consider how to represent both center (averaging) and variation [13]. Similarly, when children measure data in different contexts (for example, measuring the height of a flagpole with a handmade "height-o-meter" as compared to a standardized measuring tool), they observe a different amount of spread [14]. Another study involved asking 11-year-old children to each measure the perimeter of a table with a 15 cm ruler [92]. As expected, students' measurements varied, and then students worked in pairs to consider how to represent the full set of classroom data. These classroom studies also demonstrated a critical role for discussion as a means of advancing reasoning through relevant concepts to improve understanding.

2.2. Refining Data Sensemaking

What changes throughout development to refine this ability to summarize numerical data to estimate central tendency and variability? One contributing factor is acquiring and using more efficacious strategies (e.g., [93,94]). Children asked to summarize the spatial center in a series of dots used more strategies than adults, suggesting a less efficient process, and many of the strategies children used were not efficacious, resulting in fewer correct responses when compared to adults [41]. This result suggests that children's approaches to attending to and encoding information influence the resulting summaries [27]. Alibali et al. [56] recently proposed considering that the process of developing new strategies may be similar to a diathesis-stress model, in which there is an interaction between a "vulnerability to change . . . and a 'trigger' that actually provokes change (p. 164)". In other words, they suggest that once children have reached a point of being able to encode target problems in a way that makes key features salient, then it is possible that external input, such as feedback from successfully trying a new approach, will lead to the generation of new strategies. As they note, this does not fully explain the process of strategy generation, but it does suggest the importance of considering perceptual encoding as a factor in learning. There may also be value in considering more domain-specific models of change that occur within specific types of problems and across different age groups, for a more nuanced picture of the process [93].

As discussed above, adults are also adept at summarizing numerical information presented in sets [27]. Although people are often capable of summarizing data without conscious awareness, and encoding and drawing conclusions based on those summaries, one facet of learning to reason with data involves understanding what the summary values represent [95]. Students often gather or are given a series of individual data points, and then are asked to summarize the data. To do so effectively, they must recognize that reasoning about sets of data most commonly involves considering the data as an aggregate set, not as individual data points [96,97]. As students transition from informal reasoning about data to more scientific data reasoning, they are often taught formulas, enabling them for

example to compute means, and later standard deviations. However, the ability to apply formulas does not necessarily lead to understanding what the resulting values indicate, and how these summary values are related to the individual members of the set [35,98–100]. Nine-year-old children sometimes reason about a data set by referring only to a subset of the data [68]. Even university students who sometimes use aggregate reasoning are often inconsistent in their reasoning approaches and vary in whether they consider the full set of data or not, based on context [101,102].

Effective instruction makes use of a student's skills and prior knowledge to support their learning [103]. In the case of data reasoning, leveraging intuitions and prior knowledge about data can help students attend to relevant problem features [58], focus on possible strategies [59], and generate and attempt potential solutions that may be helpful in learning [60]. One instructional technique, preparation for learning, introduces students to relevant content before any formal instruction takes place [104]. In one application of this technique, students played a video game (Stats Invaders!) in which players identify the shape of the distributions in which invading aliens appear [58]. Students who played this video game before receiving instruction produced significantly higher scores at posttest than students who received instruction first, likely due to familiarizing students with statistical distributions before instruction began. Further exploration of how to bridge the gap between statistical intuitions and teaching statistical tools is important for clarifying this area. Statistical tools can augment and improve data reasoning, and provide some protection against cognitive biases. For example, statistical tools provide steps of formal analysis that control for sources of bias in informal analysis and allow for generalization beyond the data collected [4].

A different approach, productive failure, provides students with an opportunity to attempt to solve problems, and often fail, before instruction [59,60]. In two experiments comparing productive failure to direct instruction, students saw two instructional phases in one of two orders: (1) a data set with basketball performance and asked to determine the most consistent player and (2) direct instruction on calculating standard deviation [60]. Participants, who first explored the data and then were given direct instruction, outperformed students who were first given direct instruction before exploring. These findings suggest potential for broader applications of this concept.

To summarize, the evidence demonstrates that even young children can quickly summarize data resulting in approximate representations of statistical features such as variability, including the role of sample size. Much like summarization of sets of objects or other complex perceptual information, this process is rapid and occurs without any formal instruction. At the same time, variability is a more complex concept than the average, and children and adults often struggle to use variability information effectively. The following section will begin to explore how this initial data sensemaking underlies reasoning with data. Although children and adults can summarize large amounts of numerical information rapidly, drawing inferences and conclusions based on summary values may be skewed by mental shortcuts, known as heuristics.

2.3. Sensemaking and Reasoning from Associations between Variables

The section above described the initial process of data sensemaking that allows children and adults to summarize data. This process spares limited processing resources and provides information not available in individual numbers within a set. Summary values are one piece of information used to draw inferences. The following sections review research on both data sensemaking and reasoning from data. We combine these sections because detecting patterns in data or comparing data include both summarization and making sense of the patterns or differences that emerge from these summaries, and most of the tasks cited ask participants to reason with the data. Covariation refers to the relation between two or more variables and is one of the foundational principles in statistics and research [105]. Thus, one common application of data sensemaking and reasoning is within the contexts of reasoning about covariation between variables [106]. In the section below, we review

the experimental evidence, from research with children and adults, that illustrates data sensemaking and reasoning with covariation data and how strategy use influences informal data reasoning.

2.4. Making Sense of and Reasoning with Covariation Data

Data sensemaking often occurs when reasoning about covariation data, and drawing inferences from the patterns and relations between variables. Children and adults can detect differences in covariation data when those differences are large [44,45] or when covariation is presented within a constrained context [11]. More nuanced detection occurs as children acquire more sophisticated strategies for making sense of and interpreting covariation [44,57]. Early work in this vein [3] indicated children struggle with using covariation evidence to draw conclusions in line with the data, at least until ages 11 or 12. In many cases, children and even some adults referred to prior beliefs as justification, rather than the covariation evidence provided. For example, even if the data indicated more colds with carrot cake than chocolate cake (or no relationship), some children talked about how chocolate cake had more sugar, and was, therefore, less healthy and would lead to more colds. These findings have been used to suggest that children struggle with understanding covariation evidence, and have difficulty reasoning with this type of data.

However, follow-up work suggested that in fact young children could reason with covariation evidence when the tasks were simplified. For example, when given a less complex task, children by age six demonstrate an understanding of how covariation works. That is, they can use patterns of evidence to draw conclusions [11], particularly when the examples used tested equally plausible hypotheses. Similarly, young children ages 4–6 show evidence of the ability to use covariation evidence in drawing conclusions [107]. This suggests that young children can make sense of covariation data, when the differences are large.

Shaklee and colleagues report a series of studies in which they explored how people interpret covariation data (i.e., use strategies to reason with data) presented in contingency tables, in which there are four cells [29,44,45,108,109]. Participants were asked to consider whether there was a relationship between two dichotomous variables, such as the presence or absence of plant food and plant health or sickness. These studies demonstrated that children struggle to reason about contingency tables using sophisticated strategies, often ignoring some of the data. For example, Shaklee and Mims (1981) found that although strategy sophistication improved with age from fourth graders to college students, it was still only a minority of students even at the college level who used the most sophisticated strategy of conditional probability. Additional studies have found similar difficulties with strategy use in both children [29,110] and adults [57,109,111].

Taken together, the covariation results described above are consistent with data sensemaking that involves rapid summarization of data. In this case, detecting associations between variables would be possible with a mechanism that represents the event itself and represents an aggregate of multiple events of the same type. For example, seeing multiple instances of carrot eaters catching a cold would provide a strong pattern that should be readily detected by tracking cases [112]. However, only tracking this one outcome will lead to incorrect reasoning when there is a larger proportion of cases of carrot eaters who do not catch a cold. Finally, the consistency of the data (i.e., the strength of the correlation between variables) will make identifying the relations easier because more data points will be predictably in line with previous data points.

In many covariation studies, such as those described above, covariation is considered sufficient to demonstrate causation, and a mechanism linking two variables is not necessary. For example, in asking children to draw conclusions about the link between types of food and a cold, children are given no reason to believe one type of food would be better than another, beyond their knowledge of which foods are healthier than others. Similarly, figuring out which objects make a machine light up is determined by covariation evidence and temporal precedence. However, although covariation is one required piece of evidence

for inferring a causal relationship, it is not sufficient on its own. Further, analyzing data independent of theory is not what real-world scientists do, and there is an argument that it is important to consider the data in the context of one's prior knowledge about mechanisms that might link a cause and effect, enabling one to make an inference to the best explanation [37]. In many of these covariation studies, children are expected to ignore prior beliefs, even when prior beliefs suggest the data presented are implausible [2]. When given a potential explanatory mechanism, both children and adults reason based at least in part on these prior beliefs and mechanistic explanations, instead of exclusively on the data [3].

2.5. Sensemaking of and Reasoning with Group Comparisons

Another common inferential goal of scientific data reasoning is to determine if two (or more) groups are different on some outcome measure, and, again, data sensemaking and reasoning play a role. In this case, we are specifically talking about comparing categorical groups on a numerical or scale outcome measure. The origin of the first formal statistical test for group comparisons, Student's *t*-test, was to provide a method to compare samples of ingredients during beermaking [113]. People with training in statistics would typically use a *t*-test in making inferences about differences in an outcome between two groups. However, a small series of studies has demonstrated that children and adults often use the same components that are part of a formal *t*-test (i.e., differences between means, variance, and sample size) in drawing conclusions, even when comparing datasets without any calculation.

When comparing datasets, people generally rely most heavily on differences between means, with less attention to variance or sample size [35,68]. In a more recent study with adults, with more systematic manipulation of the mean difference and variance, larger mean differences and smaller variance in the datasets led to more accurate reasoning, more confidence in answers, and fewer visual fixations on the data [28]. These patterns suggest people summarize the data and compare the summaries quickly and accurately, without explicit computation. Other work has provided converging evidence that the magnitude of inferred (not computed) averages when comparing groups can depend in part on the magnitude of the values sampled [114], and that the ratio of means is a critical factor in reasoning about numerical data, such that accuracy of numerical perception varied in accordance with Weber's law (e.g., [83,115–117]).

A study of similar concepts looked at college students who compared pairs of consumer products in which the mean product ratings, the sample size, and the variance were all systematically manipulated. Participants focused most heavily on product ratings (magnitude of the outcome variable and the difference between means), and gave less weight to sample size. They gave the least weight of all to the sample variance [42].

Additional work has examined how college students compute analyses of variance (ANOVAs) intuitively, in which they are comparing four columns of data [46]. The data varied in their within-group variance and between-group variance, though students only saw raw data and this variance was not summarized. These students, similar to others described above, focused more in between group differences than within-group variance at the beginning of a semester-long statistics course.

The results reviewed in this section suggest that children and adults rely on data sensemaking; they make group comparisons quickly and without evidence for formal calculations, even in those who have received some formal instruction on data. Children and adults focus on differences between groups, as demonstrated by performance related to the statistical properties of the stimuli. This result pattern is consistent with a process of rapid summaries of both sets for comparisons [25,28]. In these models, individual numbers are represented as activation functions on a mental number line [118]. Multiple numbers are summarized by a secondary activation that is heightened with overlap among the individual values. The larger the distance between secondary activations, the faster the detection of difference. However, as with the detection of covariation, while children and adults are able to detect large differences, they become less accurate given smaller

differences (e.g., 9:10 ratio of means [28]) may attend to less diagnostic data features, and this informal detection may be influenced by cognitive biases.

3. Scientific Data Reasoning

Scientific reasoning processes provide tools to increase the validity and reliability of data reasoning while helping people reduce, or even avoid, common reasoning biases (e.g., confirmation bias) [119]. We will briefly describe three such tools that can improve data reasoning: external representations, scientific (i.e., theory-driven) hypothesis testing, and probabilistic conclusions. We recognize these descriptions are not comprehensive, but highlight a few key points about each topic.

3.1. External Representations

External representations refer to representations outside the mind that can be detected and processed by the perceptual and cognitive systems [21]. Examples of external representations related to numerical data are scatterplots, bar graphs, and columns of numbers in a spreadsheet. External representations allow us to record and display much larger amounts of data than can be held with fidelity in human memory [48]. Since internal representations are bound by the constraints of the human cognitive architecture, people can only attend to and process a finite amount of data at any given time [31].

External representations reduce this load by providing a representation of information, thereby allowing limited resources to be focused away from low-level process such as maintaining information in memory, to higher-order processes such as problem solving and reasoning. For example, 2nd graders were more likely to change beliefs in response to a diagram than an explanation [120], and 5th and 8th grade children were more successful in testing links between switch settings to make an electric train run when they kept external records [121]. A similar pattern holds with older participants: novice undergraduate and graduate students were more successful in solving a medical diagnostic problem when they created external representations of the problem (e.g., lists of symptoms or decision trees) than students who did not [47].

In addition to providing a reliable and durable record of data, external representations are accessible to others, and can allow for the discovery of patterns and higher-order features that would be difficult to detect in other formats (e.g., trends in a scatterplot; [21]). For example, scientists often compare their internal (i.e., mental) representations to external representations when reasoning with and interpreting ambiguous data [122].

However, even with external representations, cognitive biases can still influence data reasoning. For example, there is a tendency to underestimate means in bar graphs (though less so in point graphs), even in the presence of outliers [123]. This is likely because, as most models of graph comprehension suggest, people initially, and rapidly, summarize the main features of the graph, which forms the basis for subsequent inferences [124,125]. In sum, external representations provide powerful tools that aid with scientific data reasoning by reducing working memory burdens and making patterns and relations between variables more apparent; however, they are also subject to cognitive biases (e.g., mean underestimation).

3.2. Scientific Hypothesis Testing

Another tool for scientific data reasoning is scientific hypothesis testing. It has been documented that even young children engage in hypothesis testing [22]. The evidence for hypothesis testing and its development provides several seemingly contradictory findings. Developmental research demonstrates that young children have many of the rudiments of scientific reasoning [10,51,126,127]. When kindergartners use an inquiry-guided process, they can develop scientific questions and hypotheses more effectively than similar children not given such guidance [128]. There is also evidence that young children were more likely to seek information when evidence was inconsistent with their prior beliefs than when evidence was consistent with their beliefs [49]. Children as young as five spontaneously performed contrastive tests (i.e., compared two different experimental setups), in which

they tested whether a machine lit up with or without a targeted variable [51]. These findings collectively suggest more robust scientific reasoning ability in children than assumed in early developmental research [129].

At the same time, evidence from several studies suggest that children and adults often fail to conduct unconfounded hypothesis tests, as would occur in scientific experimentation [40,51,130]. Children often conduct confounded experiments before they have received instruction on this topic [40] and sometimes struggle to construct unconfounded hypotheses in unconstrained contexts such as discovery learning [50]. Adults sometimes do not perform contrastive tests and sometimes fail to identify causal variables [131]. There is research that demonstrates a tendency for children [132] and adults [54] to seek to confirm beliefs. Additionally, preschoolers sometimes do not seek disconfirming evidence after hearing misleading testimonial evidence [133]. This pattern of results might arise from either lacking knowledge about scientific hypothesis testing or not implementing this knowledge correctly. Further, even when seeking evidence, children and adults sometimes misinterpret or misperceive data such that new data conform with their prior beliefs, despite the misconceptions of those prior beliefs [15].

This pattern of evidence likely suggests a developmental and educational trajectory in which children's curiosity drives them to understand the world by seeking information [49,51,134,135]. Children quickly acquire impressive skills for information seeking [49] and evidence evaluation [10]. However, these skills are limited by children's emerging understanding of scientific experimentation [50], implementing this knowledge in novel contexts [38,136], and cognitive biases (e.g., confirmation bias [15,119]). The acquisition and use of scientific reasoning skills improves how people evaluate and understand the data about which they are reasoning, which improves the quality of the conclusions drawn from data. In short, the acquisition and application of scientific hypothesis testing can help to protect reasoners from errors that may reduce the accuracy of their data [52].

3.3. Probabilistic Conclusions

Data reasoning leads to conclusions, but these conclusions are always probabilistic rather than deterministic (e.g., [4,53]). Science education, including scientific data reasoning, often presents scientific conclusions as definitive [20]. Including acknowledgement of the uncertainty inherent in scientific data is an important, but often overlooked, area of science education (e.g., [137–139]). When children work with real-world data, with its variability and uncertainty, they often come to understand the nature of science more effectively [140]. Young children often appear to have a bias towards deterministic conclusions, preferring to select a single outcome when multiple outcomes are possible [141,142]. This tendency to look for a single conclusion is robust but is reduced with age [142] and can be reduced after multiple training experiences [9].

At the same time, Denison and Xu [143] argue that the majority of empirical evidence into infant (and primate) reasoning under certainty suggests that they use probabilistic reasoning in drawing conclusions. Young children have some intuitions about probability that help them make sense of situations such as the likelihood of selecting a specific object from a target set. In one recent experiment, 6- and 7-year-old children were shown a set of white and red balls with specific ratios of difference between red and white balls (e.g., 1.10–9.90; [144]). In line with previous results reported above, children's accuracy in selecting the most likely ball was closely associated with the ratio of difference.

These intuitions of data sensemaking can influence reasoning from data, but the understanding that inferences from data must be probabilistic is necessary for effective scientific reasoning, despite its challenges (e.g., [12,53]). Even when adding inferential statistics to the toolkit, there can still be a wide range of approaches taken by experts in the field [145]. That variation is one of the many challenges in thinking through scientific data reasoning, and a factor that makes teaching these concepts especially difficult. How do we leverage intuitions about data to promote scientific data reasoning?

4. Heuristics in Data Reasoning

A key limitation of all data reasoning is that humans are subject to cognitive biases, and when reasoning with data, we can fall prey to them. Tversky and Kahneman's classic work on heuristics and biases [32] suggests several ways in which shortcuts we often take to reason about data can lead us astray. For example, people are more likely to think things that come to mind easily are more common than those that do not come to mind as quickly, a phenomenon known as the availability heuristic. People also often estimate magnitude by anchoring to an initial value. When seeing data, the anchor then affects conclusions, and adjustment for the anchor is often insufficient. Additionally, people often test hypotheses with a confirmation bias, looking for evidence to support their initial beliefs rather than seeking and evaluating evidence independent of hypotheses (e.g., [15,54,55]).

Although mental shortcuts can lead to suboptimal conclusions, under some conditions, shortcuts may lead to better conclusions and decisions than deliberative reasoning, a phenomenon termed adaptive heuristics [146]. For example, when selecting the best mutual fund for retirement investments, a simple, adaptive heuristic in which one allocates equally to all options, outperformed data-driven models that far exceeded human processing limits [147]. In this case, the use of a simple strategy was highly efficient and could easily be implemented within limited cognitive capacity. Adaptive heuristics are useful when thinking about reasoning with data because we often have to make sense of large amounts of information (e.g., data) and formal data calculations require significant time, energy, and working memory capacity [148].

However, reliance on heuristics alone might result in suboptimal conclusions, as described above. Recent evidence demonstrates that training in the scientific process leads to reduced susceptibility to cognitive biases [149]. It is important to note that heuristics are not supplanted by scientific reasoning. Heuristics continue to operate even for experts and may compete for cognitive resources [150]. Experts might use heuristics in a more controlled and deliberate fashion than novices [151]. In addition, reliance on prior knowledge about mechanisms, and assessing data in light of that knowledge, often makes sense in a scientific context (e.g., [2,37,152]). For example, if an initial analysis provides evidence against a well-established pattern of evidence, it is often reasonable to check the data and analysis or even replicate a study before abandoning one's hypothesis. Additionally, consideration of the plausibility of the proposed mechanisms for an effect play a role. In the following section, we will discuss how intuitive data reasoning strategies (e.g., heuristics) play a role in data reasoning and how these processes can be leveraged through instruction to help students learn scientific data reasoning.

5. Future Directions

Many basic research questions remain in this realm of data sensemaking, informal reasoning with data, and scientific data reasoning. For example, although the evidence presented above suggests rapid summarization of data sets (e.g., [26,29,68]), more research is needed to determine the extent to which summarizing data is made on the basis of the same mechanisms underlying ensemble perception and cognition. Further, as we have discussed, data reasoning occurs in a wide ranges of contexts, including scientific reasoning, science education, decision-making, and other fields. We have focused on scientific reasoning and a little bit of the science education literature. Further exploration of differences in data reasoning across disciplines, with and without the supports of external representations, scientific hypothesis testing, and probabilistic conclusions, would also help in understanding the process of data reasoning more thoroughly.

We have suggested that reasoning with data begins with data sensemaking, a rapid summarization process that reduces processing burdens while providing information about the statistical properties of number sets. This process appears to improve through development and education, resulting in more accurate summaries. We suggested that one important factor underlying these improvements is the acquisition and use of more effective strategies, which are developed with experience and education. Data reasoning

is drawing inferences from the data, along with prior knowledge and other relevant information. Much like data summaries, data reasoning is often "accurate enough" for everyday contexts. One limitation to accuracy for summaries and inferences is reasoning biases, such as the confirmation bias or the tendency to seek data consistent with prior expectations. We propose that the acquisition of scientific data reasoning provides tools that improve the fidelity of the data itself, the conditions through which data are acquired, representation of data (e.g., figures), and the types of conclusions drawn from data. An important future direction is to evaluate this model experimentally.

Our proposed model needs direct testing of its components, though focus on the concepts at a fine grain size provides researchers with opportunities to evaluate elements of the model or the model itself. Our predictions about relatively automatic summarization of data sets can be evaluated directly. Educators can implement parts of our model individually or in concert. For example, lessons on data interpretation can encourage reliance on data summarization, with instruction guiding students to describe patterns and compare data sets in consistent ways. Below, we highlight several specific suggestions for future directions.

One complex topic in need of much further exploration is the interplay between prior knowledge and data reasoning. Although there is a fair amount of work about integrating theory and evidence (e.g., [2,3,90,152,153]), there is less work on how prior beliefs interact with different types of numerical data (e.g., [13,14,91]). There is evidence that prior knowledge increases attention to diagnostic features [154] and helps reasoners solve problems more effectively [155]. However, this attention to diagnostic features has not to our knowledge been tested with data reasoning.

In addition, there are many educational applications of data reasoning, and specifically scientific data reasoning. Future research aimed at effective application of these concepts in the classroom can be beneficial both to understanding scientific data reasoning, and to developing best practices in education. As discussed above, classroom studies in which students develop their own measures of description and inferences from data have been shown to facilitate a more comprehensive understanding of concepts such as the aggregation of data and variability, building on initial intuitions (e.g., [13,14,35,156]). Considered within the framing outlined above by Alibali et al. [93], the classroom conversations could be considered a potential trigger for provoking changes in strategies used to approach these problems, and in turn, increase learning. This process can work in informal data reasoning or scientific data reasoning contexts. Follow-up studies could directly examine strategy acquisition and be used to develop a more comprehensive understanding of how strategies aid in learning about data reasoning.

There is a lot of work demonstrating the efficacy of classroom interventions or curricular approaches in improving people's ability to reason statistically [96]. A meta-analysis of scientific reasoning interventions, targeting a wider range of topics than just data reasoning, indicated there was a small effect in classroom interventions across ages [19]. Similarly, there are many demonstrations of efficacy of specific tools aiding in data reasoning within lab contexts (e.g., [56,120,157,158]. However, scaling up these interventions into more effective curricula at all levels (including teacher training) remains a challenge.

Teaching materials are also important in facilitating (or unintentionally hindering) student learning. Textbooks play an important role in student learning, and limitations in textbook content can affect student learning. Children can acquire misconceptions through misaligned instructional materials. One source of misconception can be examples in textbooks. A notable example from mathematics is children's misconception of the equal sign, in which children interpret the equal sign as a signal for executing an operation rather than balancing both sides of an equation [159]. An analysis of math textbooks demonstrated that most practice problems had the same structure (e.g., 3 + 5 = ?) that is consistent with this misconception [160]. Another study of middle school science textbooks showed they typically include limited guidance in appropriate use of data [161]. In fact, the majority of data reasoning activities in science texts provided little guidance on how to analyze or draw inferences from data formally. Thus, one step that can help improve student learning

of data concepts involves improved integration of descriptions and applied exercises in textbooks used in science classes. This research demonstrates the importance of using instructional materials that do not promote biases or misconceptions [160,162].

Finally, one last suggested future direction is investigating the role of intuitions and potential misconceptions, both about science and about data, in scientific data reasoning. One difficulty in science education is that many scientific phenomena are challenging to understand, and in many cases intuitions conflict with scientific consensus, such as in understanding of physical principles of heat and motion as well as biological principles of inheritance and evolution [163]. Thus, although intuitions about data can be useful in data reasoning, intuitions about conceptual content sometimes lead people to incorrect beliefs and misconceptions. Indeed, there is some evidence it can be at the observation stage where incorrect prior beliefs interfere with accurate perception of physical phenomena and the gathering of potentially-informative data [15].

6. Conclusions

We proposed a model of data reasoning and its relation to scientific reasoning. Specifically, we suggest that data reasoning begins developmentally with data sensemaking, a relatively automatic process rooted in perceptual mechanisms that summarize large quantities of information in the environment. As these summarization mechanisms operate on number sets, they yield approximate representations of statistical properties, such as central tendency and variation. This information is then available for drawing inferences from data. However, both data sensemaking and informal data reasoning may lead to erroneous conclusions due to cognitive biases or heuristics. The acquisition of scientific data reasoning helps to reduce these biases by providing tools and procedures that improve data reasoning. These tools include external representations, scientific hypothesis testing, and drawing probabilistic conclusions. Although data sensemaking and informal data reasoning are not supplanted by scientific data reasoning, these skills can be leveraged to improve learning of science and reasoning with data.

References

1. Klahr, D.; Dunbar, K. Dual space search during scientific seasoning. *Cogn. Sci.* **1988**, *12*, 1–48. [CrossRef]
2. Koslowski, B. *Theory and Evidence: The Development of Scientific Reasoning*; MIT Press: Cambridge, MA, USA, 1996; ISBN 978-0-262-11209-4.
3. Kuhn, D.; Amsel, E.; O'Loughlin, M.; Schauble, L.; Leadbeater, B.; Yotive, W. *The Development of Scientific Thinking Skills*; Academic Press: San Diego, CA, USA, 1988; ISBN 978-0-12-428430-2.
4. Grolemund, G.; Wickham, H. A cognitive interpretation of data analysis. *Int. Stat. Rev.* **2014**, *82*, 184–204. [CrossRef]
5. Zimmerman, C.; Klahr, D. Development of scientific thinking. In *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*; Wixted, J.T., Ed.; John Wiley & Sons, Inc.: New York, NY, USA, 2018; pp. 1–25, ISBN 978-1-119-17016-7.
6. Fischer, F.; Kollar, I.; Ufer, S.; Sodian, B.; Hussmann, H.; Pekrun, R.; Neuhaus, B.; Dorner, B.; Pankofer, S.; Fischer, M.; et al. Scientific reasoning and argumentation: Advancing an interdisciplinary research agenda in education. *Frontline Learn. Res.* **2014**, *2*, 28–45.
7. Morris, B.J.; Croker, S.; Masnick, A.M.; Zimmerman, C. The emergence of scientific reasoning. In *Current Topics in Children's Learning and Cognition*; Kloos, H., Ed.; InTech: Rijeka, Croatia, 2012; ISBN 978-953-51-0855-9.
8. Zimmerman, C. The development of scientific reasoning skills. *Dev. Rev.* **2000**, *20*, 99–149. [CrossRef]
9. Klahr, D.; Chen, Z. Overcoming the positive-capture strategy in young children: Learning about indeterminacy. *Child Dev.* **2003**, *74*, 1275–1296. [CrossRef]
10. Sodian, B.; Zaitchik, D.; Carey, S. Young children's differentiation of hypothetical beliefs from evidence. *Child Dev.* **1991**, *62*, 753–766. [CrossRef]

11. Ruffman, T.; Perner, J.; Olson, D.R.; Doherty, M. Reflecting on scientific thinking: Children's understanding of the hypothesis-evidence relation. *Child Dev.* **1993**, *64*, 1617–1636. [CrossRef]

12. Lehrer, R.; Schauble, L.; Wisittanawat, P. Getting a grip on variability. *Bull. Math. Biol.* **2020**, *82*, 106. [CrossRef]

13. Lehrer, R.; Schauble, L. Modeling natural variation through distribution. *Am. Educ. Res. J.* **2004**, *41*, 635–679. [CrossRef]

14. Petrosino, A.J.; Lehrer, R.; Schauble, L. Structuring error and experimental variation as distribution in the fourth grade. *Math. Think. Learn.* **2003**, *5*, 131–156. [CrossRef]

15. Chinn, C.A.; Malhotra, B.A. Children's responses to anomalous scientific data: How is conceptual change impeded? *J. Educ. Psychol.* **2002**, *94*, 327–343. [CrossRef]

16. Koslowski, B.; Okagaki, L.; Lorenz, C.; Umbach, D. When covariation is not enough: The role of causal mechanism, sampling method, and sample size in causal reasoning. *Child Dev.* **1989**, *60*, 1316–1327. [CrossRef]

17. Kuhn, D. What is scientific thinking and how does it develop? In *The Wiley-Blackwell Handbook of Childhood Cognitive Development*, 2nd ed.; Wiley-Blackwell: Malden, MA, USA, 2011; pp. 497–523, ISBN 978-1-4051-9116-6.

18. van Lieshout, L.L.; de Lange, F.P.; Cools, R. Why so curious? Quantifying mechanisms of information seeking. *Curr. Opin. Behav. Sci.* **2020**, *35*, 112–117. [CrossRef]

19. Engelmann, K.; Neuhaus, B.J.; Fischer, F. Fostering scientific reasoning in education—Meta-analytic evidence from intervention studies. *Educ. Res. Eval.* **2016**, *22*, 333–349. [CrossRef]

20. Shah, P.; Michal, A.; Ibrahim, A.; Rhodes, R.; Rodriguez, F. What makes everyday scientific reasoning so challenging? In *Psychology of Learning and Motivation*; Elsevier: Amsterdam, The Netherlands, 2017; Volume 66, pp. 251–299, ISBN 978-0-12-812118-4.

21. Zhang, J. The nature of external representations in problem solving. *Cogn. Sci.* **1997**, *21*, 179–217. [CrossRef]

22. Koerber, S.; Mayer, D.; Osterhaus, C.; Schwippert, K.; Sodian, B. The development of scientific thinking in elementary school: A comprehensive inventory. *Child Dev.* **2015**, *86*, 327–336. [CrossRef]

23. Di Blas, N.; Mazuran, M.; Paolini, P.; Quintarelli, E.; Tanca, L. Exploratory computing: A comprehensive approach to data sensemaking. *Int. J. Data Sci. Anal.* **2017**, *3*, 61–77. [CrossRef]

24. Koesten, L.; Gregory, K.; Groth, P.; Simperl, E. Talking datasets—Understanding data sensemaking behaviours. *Int. J. Hum.-Comput. Stud.* **2021**, *146*, 102562. [CrossRef]

25. Alvarez, G.A. Representing multiple objects as an ensemble enhances visual cognition. *Trends Cogn. Sci.* **2011**, *15*, 122–131. [CrossRef] [PubMed]

26. Brezis, N.; Bronfman, Z.Z.; Usher, M. Adaptive spontaneous transitions between two mechanisms of numerical averaging. *Sci. Rep.* **2015**, *5*, 10415. [CrossRef]

27. Whitney, D.; Yamanashi Leib, A. Ensemble perception. *Annu. Rev. Psychol.* **2018**, *69*, 105–129. [CrossRef] [PubMed]

28. Morris, B.J.; Masnick, A.M. Comparing data sets: Implicit summaries of the statistical properties of number sets. *Cogn. Sci.* **2015**, *39*, 156–170. [CrossRef] [PubMed]

29. Shaklee, H.; Paszek, D. Covariation judgment: Systematic rule use in middle childhood. *Child Dev.* **1985**, *56*, 1229–1240. [CrossRef]

30. Makar, K.; Rubin, A. Learning about statistical inference. In *International Handbook of Research in Statistics Education*; Ben-Zvi, D., Makar, K., Garfield, J., Eds.; Springer International Handbooks of Education; Springer International Publishing: Cham, Switzerland, 2018; pp. 261–294, ISBN 978-3-319-66195-7.

31. Cowan, N. Mental objects in working memory: Development of basic capacity or of cognitive completion? *Adv. Child Dev. Behav.* **2017**, *52*, 81–104.

32. Tversky, A.; Kahneman, D. Judgment under uncertainty: Heuristics and biases. *Science* **1974**, *185*, 1124–1131. [CrossRef]

33. Makar, K. Developing young children's emergent inferential practices in statistics. *Math. Think. Learn.* **2016**, *18*, 1–24. [CrossRef]

34. Makar, K.; Bakker, A.; Ben-Zvi, D. The reasoning behind informal statistical inference. *Math. Think. Learn.* **2011**, *13*, 152–173. [CrossRef]

35. Watson, J.M.; Moritz, J.B. The beginning of statistical inference: Comparing two data sets. *Educ. Stud. Math.* **1999**, *37*, 145–168. [CrossRef]

36. Zieffler, A.; Garfield, J.; Delmas, R.; Reading, C. A framework to support research on informal inferential reasoning. *Stat. Educ. Res. J.* **2008**, *7*, 40–58.

37. Koslowski, B. Inference to the best explanation (IBE) and the causal and scientific reasoning of nonscientists. In *Psychology of Science: Implicit and Explicit Processes*; Proctor, R.W., Capaldi, E.J., Eds.; OUP: New York, NY, USA, 2012; ISBN 978-0-19-975362-8.

38. Chen, Z.; Klahr, D. All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Dev.* **1999**, *70*, 1098–1120. [CrossRef]

39. Krüger, D.; Hartmann, S.; Nordmeier, V.; Upmeier zu Belzen, A. Measuring scientific reasoning competencies. In *Student Learning in German Higher Education: Innovative Measurement Approaches and Research Results*; Zlatkin-Troitschanskaia, O., Pant, H.A., Toepper, M., Lautenbach, C., Eds.; Springer Fachmedien: Wiesbaden, Germany, 2020; pp. 261–280, ISBN 978-3-658-27886-1.

40. Klahr, D. *Exploring Science: The Cognition and Development of Discovery Processes*; MIT Press: Cambridge, MA, USA, 2002; ISBN 978-0-262-61176-3.

41. Jones, P.R.; Dekker, T.M. The development of perceptual averaging: Learning what to do, not just how to do it. *Dev. Sci.* **2018**, *21*, e12584. [CrossRef] [PubMed]

42. Obrecht, N.A.; Chapman, G.B.; Gelman, R. Intuitive t tests: Lay use of statistical information. *Psychon. Bull. Rev.* **2007**, *14*, 1147–1152. [CrossRef] [PubMed]

43. Peterson, C.R.; Beach, L.R. Man as an intuitive statistician. *Psychol. Bull.* **1967**, *68*, 29–46. [CrossRef]
44. Shaklee, H.; Mims, M. Development of rule use in judgments of covariation between events. *Child Dev.* **1981**, *52*, 317–325. [CrossRef]
45. Shaklee, H.; Holt, P.; Elek, S.; Hall, L. Covariation judgment: Improving rule use among children, adolescents, and adults. *Child Dev.* **1988**, *59*, 755–768. [CrossRef]
46. Trumpower, D.L. Formative use of intuitive Analysis of Variance. *Math. Think. Learn.* **2013**, *15*, 291–313. [CrossRef]
47. Martin, L.; Schwartz, D.L. Prospective adaptation in the use of external representations. *Cogn. Instr.* **2009**, *27*, 370–400. [CrossRef]
48. Zhang, J.; Wang, H. An exploration of the relations between external representations and working memory. *PLoS ONE* **2009**, *4*, e6513. [CrossRef]
49. Bonawitz, E.B.; van Schijndel, T.J.P.; Friel, D.; Schulz, L. Children balance theories and evidence in exploration, explanation, and learning. *Cogn. Psychol.* **2012**, *64*, 215–234. [CrossRef]
50. Klahr, D.; Nigam, M. The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychol. Sci.* **2004**, *15*, 661–667. [CrossRef]
51. Köksal-Tuncer, Ö.; Sodian, B. The development of scientific reasoning: Hypothesis testing and argumentation from evidence in young children. *Cogn. Dev.* **2018**, *48*, 135–145. [CrossRef]
52. Sandoval, W.A.; Sodian, B.; Koerber, S.; Wong, J. Developing children's early competencies to engage with science. *Educ. Psychol.* **2014**, *49*, 139–152. [CrossRef]
53. Wild, C.J.; Utts, J.M.; Horton, N.J. What is statistics? In *International Handbook of Research in Statistics Education*; Ben-Zvi, D., Makar, K., Garfield, J., Eds.; Springer International Handbooks of Education; Springer International Publishing: Cham, Switzerland, 2018; pp. 5–36, ISBN 978-3-319-66195-7.
54. Klayman, J.; Ha, Y. Confirmation, disconfirmation, and information in hypothesis testing. *Psychol. Rev.* **1987**, *94*, 211–228. [CrossRef]
55. Mynatt, C.R.; Doherty, M.E.; Tweney, R.D. Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Q. J. Exp. Psychol.* **1977**, *29*, 85–95. [CrossRef]
56. Alibali, M.W.; Crooks, N.M.; McNeil, N.M. Perceptual support promotes strategy generation: Evidence from equation solving. *Br. J. Dev. Psychol.* **2018**, *36*, 153–168. [CrossRef]
57. Batanero, C.; Estepa, A.; Godino, J.D.; Green, D.R. Intuitive strategies and preconceptions about association in contingency tables. *J. Res. Math. Educ.* **1996**, *27*, 151–169. [CrossRef]
58. Arena, D.A.; Schwartz, D.L. Experience and explanation: Using videogames to prepare students for formal instruction in statistics. *J. Sci. Educ. Technol.* **2014**, *23*, 538–548. [CrossRef]
59. Kapur, M. Productive failure. *Cogn. Instr.* **2008**, *26*, 379–424. [CrossRef]
60. Kapur, M. Productive failure in learning math. *Cogn. Sci.* **2014**, *38*, 1008–1022. [CrossRef]
61. Dehaene, S. *The Number Sense: How the Mind Creates Mathematics*; Rev and Updated Edition; Oxford University Press: New York, NY, USA, 2011; ISBN 978-0-19-975387-1.
62. Hyde, D.C. Two systems of non-symbolic numerical cognition. *Front. Hum. Neurosci.* **2011**, *5*, 150. [CrossRef]
63. Rosenbaum, D.; de Gardelle, V.; Usher, M. Ensemble perception: Extracting the average of perceptual versus numerical stimuli. *Atten. Percept. Psychophys.* **2021**, *83*, 956–969. [CrossRef] [PubMed]
64. Malmi, R.A.; Samson, D.J. Intuitive averaging of categorized numerical stimuli. *J. Verbal Learn. Verbal Behav.* **1983**, *2*, 547–559. [CrossRef]
65. Beach, L.R.; Swenson, R.G. Intuitive estimation of means. *Psychon. Sci.* **1966**, *5*, 161–162. [CrossRef]
66. Grebstein, L.C. Relative accuracy of actuarial prediction, experienced clinicians, and graduate students in a clinical judgment task. *J. Consult. Psychol.* **1963**, *27*, 127–132. [CrossRef]
67. Irwin, F.W.; Smith, W.A.S. Value, cost, and information as determiners of decision. *J. Exp. Psychol.* **1957**, *54*, 229–232. [CrossRef] [PubMed]
68. Masnick, A.M.; Morris, B.J. Investigating the development of data evaluation: The role of data characteristics. *Child Dev.* **2008**, *79*, 1032–1048. [CrossRef]
69. Spencer, J. Estimating averages. *Ergonomics* **1961**, *4*, 317–328. [CrossRef]
70. Spencer, J. A further study of estimating averages. *Ergonomics* **1963**, *6*, 255–265. [CrossRef]
71. Obrecht, N.A.; Chapman, G.B.; Suárez, M.T. Laypeople do use sample variance: The effect of embedding data in a variance-implying story. *Think. Reason.* **2010**, *16*, 26–44. [CrossRef]
72. Irwin, F.W.; Smith, W.A.S.; Mayfield, J.F. Tests of two theories of decision in an "expanded judgment" situation. *J. Exp. Psychol.* **1956**, *51*, 261–268. [CrossRef]
73. Jacobs, J.E.; Narloch, R.H. Children's use of sample size and variability to make social inferences. *J. Appl. Dev. Psychol.* **2001**, *22*, 311–331. [CrossRef]
74. Obrecht, N.A. Sample size weighting follows a curvilinear function. *J. Exp. Psychol. Learn. Mem. Cogn.* **2019**, *45*, 614–626. [CrossRef]
75. Wagemans, J.; Elder, J.H.; Kubovy, M.; Palmer, S.E.; Peterson, M.A.; Singh, M.; von der Heydt, R. A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychol. Bull.* **2012**, *138*, 1172–1217. [CrossRef] [PubMed]

76. Lee, K.R.; Sobel, K.V.; York, A.K.; Puri, A.M. Dissociating parallel and serial processing of numerical value. *J. Numer. Cogn.* **2018**, *4*, 360–379. [CrossRef]

77. Van Opstal, F.; de Lange, F.P.; Dehaene, S. Rapid parallel semantic processing of numbers without awareness. *Cognition* **2011**, *120*, 136–147. [CrossRef]

78. Dehaene, S.; Akhavein, R. Attention, automaticity, and levels of representation in number processing. *J. Exp. Psychol. Learn. Mem. Cogn.* **1995**, *21*, 314–326. [CrossRef]

79. Lee, K.R.; Dague, T.D.; Sobel, K.V.; Paternoster, N.J.; Puri, A.M. Set size and ensemble perception of numerical value. *Atten. Percept. Psychophys.* **2021**, *83*, 1169–1178. [CrossRef] [PubMed]

80. Zosh, J.M.; Halberda, J.; Feigenson, L. Memory for multiple visual ensembles in infancy. *J. Exp. Psychol. Gen.* **2011**, *140*, 141–158. [CrossRef]

81. Sweeny, T.D.; Wurnitsch, N.; Gopnik, A.; Whitney, D. Ensemble perception of size in 4–5-year-old children. *Dev. Sci.* **2015**, *18*, 556–568. [CrossRef] [PubMed]

82. Xu, F.; Spelke, E.S. Large number discrimination in 6-month-old infants. *Cognition* **2000**, *74*, B1–B11. [CrossRef]

83. Halberda, J.; Feigenson, L. Developmental change in the acuity of the "number sense": The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Dev. Psychol.* **2008**, *44*, 1457–1465. [CrossRef]

84. Rhodes, G.; Neumann, M.; Ewing, L.; Bank, S.; Read, A.; Engfors, L.M.; Emiechel, R.; Palermo, R. Ensemble coding of faces occurs in children and develops dissociably from coding of individual faces. *Dev. Sci.* **2018**, *21*, e12540. [CrossRef] [PubMed]

85. Wild, C.J.; Pfannkuch, M. Statistical thinking in empirical enquiry. *Int. Stat. Rev.* **1999**, *67*, 223–248. [CrossRef]

86. Lee, V.R.; Wilkerson, M.H. Data use by middle and secondary students in the digital age: A status report and future prospects. In *Commissioned Paper for the National Academies of Sciences, Engineering, and Medicine, Board on Science Education, Committee on Science Investigations and Engineering Design for Grades 6–12*; National Academy of Sciences Engineering, and Medicine: Washington, DC, USA, 2018; p. 43.

87. Cobb, G.W.; Moore, D.S. Mathematics, statistics, and teaching. *Am. Math. Mon.* **1997**, *104*, 801–823. [CrossRef]

88. Lubben, F.; Millar, R. Children's ideas about the reliability of experimental data. *Int. J. Sci. Educ.* **1996**, *18*, 955–968. [CrossRef]

89. Masnick, A.M.; Klahr, D. Error matters: An initial exploration of elementary school children's understanding of experimental error. *J. Cogn. Dev.* **2003**, *4*, 67–98. [CrossRef]

90. Kanari, Z.; Millar, R. Reasoning from data: How students collect and interpret data in science investigations. *J. Res. Sci. Teach.* **2004**, *41*, 748–769. [CrossRef]

91. Masnick, A.M.; Klahr, D.; Knowles, E.R. Data-driven belief revision in children and adults. *J. Cogn. Dev.* **2017**, *18*, 87–109. [CrossRef]

92. Lehrer, R. Modeling signal-noise processes supports student construction of a hierarchical image of sample. *SERJ* **2017**, *16*, 64–85. [CrossRef]

93. Alibali, M.W.; Brown, S.A.; Menendez, D. Understanding strategy change: Contextual, individual, and metacognitive factors. *Adv. Child Dev. Behav.* **2019**, *56*, 227–256. [PubMed]

94. Siegler, R.; Jenkins, E.A. *How Children Discover New Strategies*; Psychology Press: New York, NY, USA, 1989; ISBN 978-1-315-80774-4.

95. Hancock, C.; Kaput, J.J.; Goldsmith, L.T. Authentic inquiry with data: Critical barriers to classroom implementation. *Educ. Psychol.* **1992**, *27*, 337–364. [CrossRef]

96. Garfield, J.; Ben-Zvi, D. How students learn statistics revisited: A current review of research on teaching and learning statistics. *Int. Stat. Rev.* **2007**, *75*, 372–396. [CrossRef]

97. Konold, C.; Higgins, T.; Russell, S.J.; Khalil, K. Data seen through different lenses. *Educ. Stud. Math.* **2015**, *88*, 305–325. [CrossRef]

98. Konold, C.; Pollatsek, A. Data analysis as the search for signals in noisy processes. *J. Res. Math. Educ.* **2002**, *33*, 259. [CrossRef]

99. Mokros, J.; Russell, S.J. Children's concepts of average and representativeness. *J. Res. Math. Educ.* **1995**, *26*, 20–39. [CrossRef]

100. Pollatsek, A.; Lima, S.; Well, A.D. Concept or computation: Students' understanding of the mean. *Educ. Stud. Math.* **1981**, *12*, 191–204. [CrossRef]

101. Buffler, A.; Allie, S.; Lubben, F. The development of first year physics students' ideas about measurement in terms of point and set paradigms. *Int. J. Sci. Educ.* **2001**, *23*, 1137–1156. [CrossRef]

102. Lubben, F.; Campbell, B.; Buffler, A.; Allie, S. Point and set reasoning in practical science measurement by entering university freshmen. *Sci. Educ.* **2001**, *85*, 311–327. [CrossRef]

103. Willingham, D.T. *Why Don't Students Like School? A Cognitive Scientist Answers Questions about How the Mind Works and What It Means for the Classroom*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2021; ISBN 978-1-119-71580-1.

104. Schwartz, D.L.; Martin, T. Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cogn. Instr.* **2004**, *22*, 129–184. [CrossRef]

105. Kerlinger, F.N.; Lee, H.B. *Foundations of Behavioral Research*; Harcourt College Publishers: Fort Worth, TX, USA, 2000; ISBN 978-0-15-507897-0.

106. Biehler, R.; Frischemeier, D.; Reading, C.; Shaughnessy, J.M. Reasoning about data. In *International Handbook of Research in Statistics Education*; Ben-Zvi, D., Makar, K., Garfield, J., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 139–192.

107. Koerber, S.; Sodian, B.; Thoermer, C.; Nett, U. Scientific reasoning in young children: Preschoolers' ability to evaluate covariation evidence. *Swiss J. Psychol.* **2005**, *64*, 141–152. [CrossRef]

108. Shaklee, H.; Mims, M. Sources of error in judging event covariations: Effects of memory demands. *J. Exp. Psychol. Learn. Mem. Cogn.* **1982**, *8*, 208–224. [CrossRef]
109. Shaklee, H.; Tucker, D. A rule analysis of judgments of covariation between events. *Mem. Cogn.* **1980**, *8*, 459–467. [CrossRef]
110. Obersteiner, A.; Bernhard, M.; Reiss, K. Primary school children's strategies in solving contingency table problems: The role of intuition and inhibition. *ZDM Math. Educ.* **2015**, *47*, 825–836. [CrossRef]
111. Osterhaus, C.; Magee, J.; Saffran, A.; Alibali, M.W. Supporting successful interpretations of covariation data: Beneficial effects of variable symmetry and problem context. *Q. J. Exp. Psychol.* **2019**, *72*, 994–1004. [CrossRef]
112. Jung, Y.; Walther, D.B.; Finn, A.S. Children automatically abstract categorical regularities during statistical learning. *Dev. Sci.* **2021**, *24*, e13072. [CrossRef]
113. Student. The probable error of a mean. *Biometrika* **1908**, *6*, 1–25. [CrossRef]
114. Spitzer, B.; Waschke, L.; Summerfield, C. Selective overweighting of larger magnitudes during noisy numerical comparison. *Nat. Hum. Behav.* **2017**, *1*, 0145. [CrossRef] [PubMed]
115. Dehaene, S.; Dehaene-Lambertz, G.; Cohen, L. Abstract representations of numbers in the animal and human brain. *Trends Neurosci.* **1998**, *21*, 355–361. [CrossRef]
116. Eckert, J.; Call, J.; Hermes, J.; Herrmann, E.; Rakoczy, H. Intuitive statistical inferences in chimpanzees and humans follow Weber's Law. *Cognition* **2018**, *180*, 99–107. [CrossRef] [PubMed]
117. Libertus, M.E.; Brannon, E.M. Behavioral and neural basis of number sense in infancy. *Curr. Dir. Psychol. Sci.* **2009**, *18*, 346–351. [CrossRef]
118. Dehaene, S. Origins of mathematical intuitions: The case of arithmetic. *Ann. N. Y. Acad. Sci.* **2009**, *1156*, 232–259. [CrossRef] [PubMed]
119. Wickens, C.D.; Helton, W.S.; Hollands, J.G.; Banbury, S. *Engineering Psychology and Human Performance*, 5th ed.; Routledge: New York, NY, USA, 2021; ISBN 978-1-00-317761-6.
120. Koerber, S.; Osterhaus, C.; Sodian, B. Diagrams support revision of prior belief in primary-school children. *Frontline Learn. Res.* **2017**, *5*, 76–84. [CrossRef]
121. Siegler, R.S.; Liebert, R.M. Acquisition of formal scientific reasoning by 10- and 13-year-olds: Designing a factorial experiment. *Dev. Psychol.* **1975**, *11*, 401–402. [CrossRef]
122. Trickett, S.B.; Trafton, J.G.; Schunn, C.D. How do scientists respond to anomalies? Different strategies used in basic and applied science. *Top. Cogn. Sci.* **2009**, *1*, 711–729. [CrossRef] [PubMed]
123. Godau, C.; Vogelgesang, T.; Gaschler, R. Perception of bar graphs—A biased impression? *Comput. Hum. Behav.* **2016**, *59*, 67–73. [CrossRef]
124. Pinker, S. A theory of graph comprehension. In *Artificial Intelligence and the Future of Testing*; Freedle, R.O., Ed.; Psychology Press: New York, NY, USA, 1990; pp. 73–126, ISBN 978-0-8058-0117-0.
125. Zacks, J.M.; Tversky, B. Event structure in perception and conception. *Psychol. Bull.* **2001**, *127*, 3–21. [CrossRef]
126. Köksal, Ö.; Sodian, B.; Legare, C.H. Young children's metacognitive awareness of confounded evidence. *J. Exp. Child Psychol.* **2021**, *205*, 105080. [CrossRef]
127. Tschirgi, J.E. Sensible reasoning: A hypothesis about hypotheses. *Child Dev.* **1980**, *51*, 1–10. [CrossRef]
128. Samarapungavan, A.; Mantzicopoulos, P.; Patrick, H. Learning science through inquiry in kindergarten. *Sci. Educ.* **2008**, *92*, 868–908. [CrossRef]
129. Inhelder, B.; Piaget, J. *The Growth of Logical Thinking from Childhood to Adolescence: An Essay on the Construction of Formal Operational Structures*; Basic Books: New York, NY, USA, 1958; ISBN 978-0-415-21002-7.
130. Crowell, A.; Kuhn, D. Developing dialogic argumentation skills: A 3-year intervention study. *J. Cogn. Dev.* **2014**, *15*, 363–381. [CrossRef]
131. Kuhn, D.; David, D., Jr. Metacognition: A bridge between cognitive psychology and educational practice. *Theory Pract.* **2004**, *43*, 268–273. [CrossRef]
132. Garcia-Mila, M.; Andersen, C. Cognitive foundations of learning argumentation. In *Argumentation in Science Education: Perspectives from Classroom-Based Research*; Erduran, S., Jiménez-Aleixandre, M.P., Eds.; Science & Technology Education Library; Springer: Dordrecht, The Netherlands, 2007; pp. 29–45, ISBN 978-1-4020-6670-2.
133. Hermansen, T.K.; Ronfard, S.; Harris, P.L.; Zambrana, I.M. Preschool children rarely seek empirical data that could help them complete a task when observation and testimony conflict. *Child Dev.* **2021**, *92*, 2546–2562. [CrossRef]
134. Jirout, J.; Klahr, D. Children's scientific curiosity: In search of an operational definition of an elusive concept. *Dev. Rev.* **2012**, *32*, 125–160. [CrossRef]
135. Jirout, J.; Zimmerman, C. Development of science process skills in the early childhood years. In *Research in Early Childhood Science Education*; Cabe Trundle, K., Saçkes, M., Eds.; Springer: Dordrecht, The Netherlands, 2015; pp. 143–165, ISBN 978-94-017-9504-3.
136. Klahr, D.; Chen, Z.; Toth, E.E. Cognitive development and science education: Ships that pass in the night or beacons of mutual illumination? In *Cognition and Instruction: Twenty-Five Years of Progress*; Lawrence Erlbaum Associates Publishers: Mahwah, NJ, USA, 2001; pp. 75–119, ISBN 978-0-8058-3823-7.
137. McComas, W.F.; Olson, J.K. The Nature of Science in international science education standards documents. In *The Nature of Science in Science Education: Rationales and Strategies*; McComas, W.F., Ed.; Science & Technology Education Library; Springer: Dordrecht, The Netherlands, 2002; pp. 41–52, ISBN 978-0-306-47215-2.

138. Osborne, J.; Collins, S.; Ratcliffe, M.; Millar, R.; Duschl, R. What "ideas-about-science" should be taught in school science? A Delphi study of the expert community. *J. Res. Sci. Teach.* **2003**, *40*, 692–720. [CrossRef]

139. Priemer, B.; Hellwig, J. Learning about measurement uncertainties in secondary education: A model of the subject matter. *Int. J. Sci. Math. Educ.* **2018**, *16*, 45–68. [CrossRef]

140. Duschl, R.A.; Grandy, R. Two views about explicitly teaching Nature of Science. *Sci. Educ.* **2013**, *22*, 2109–2139. [CrossRef]

141. Fay, A.L.; Klahr, D. Knowing about guessing and guessing about knowing: Preschoolers' understanding of indeterminacy. *Child Dev.* **1996**, *67*, 689–716. [CrossRef]

142. Metz, K.E. Emergent understanding and attribution of randomness: Comparative analysis of the reasoning of primary grade children and undergraduates. *Cogn. Instr.* **1998**, *16*, 265–285. [CrossRef]

143. Denison, S.; Xu, F. Infant statisticians: The origins of reasoning under uncertainty. *Perspect. Psychol. Sci.* **2019**, *14*, 499–509. [CrossRef]

144. O'Grady, S.; Xu, F. The development of nonsymbolic probability judgments in children. *Child Dev.* **2020**, *91*, 784–798. [CrossRef] [PubMed]

145. Silberzahn, R.; Uhlmann, E.L.; Martin, D.P.; Anselmi, P.; Aust, F.; Awtrey, E.; Bahník, Š.; Bai, F.; Bannard, C.; Bonnier, E.; et al. Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Adv. Methods Pract. Psychol. Sci.* **2018**, *1*, 337–356. [CrossRef]

146. Gigerenzer, G. Why heuristics work. *Perspect. Psychol. Sci.* **2008**, *3*, 20–29. [CrossRef]

147. DeMiguel, V.; Garlappi, L.; Nogales, F.J.; Uppal, R. A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Manag. Sci.* **2009**, *55*, 798–812. [CrossRef]

148. Gigerenzer, G.; Gaissmaier, W. Heuristic decision making. *Ann. Rev. Psychol.* **2011**, *62*, 451–482. [CrossRef] [PubMed]

149. Čavojová, V.; Šrol, J.; Jurkovič, M. Why should we try to think like scientists? Scientific reasoning and susceptibility to epistemically suspect beliefs and cognitive biases. *Appl. Cogn. Psychol.* **2020**, *34*, 85–95. [CrossRef]

150. Beilock, S.L.; DeCaro, M.S. From poor performance to success under stress: Working memory, strategy selection, and mathematical problem solving under pressure. *J. Exp. Psychol. Learn. Mem. Cogn.* **2007**, *33*, 983–998. [CrossRef] [PubMed]

151. Raab, M.; Gigerenzer, G. The power of simplicity: A fast-and-frugal heuristics approach to performance science. *Front. Psychol.* **2015**, *6*, 1672. [CrossRef]

152. Ahn, W.; Kalish, C.W.; Medin, D.L.; Gelman, S.A. The role of covariation versus mechanism information in causal attribution. *Cognition* **1995**, *54*, 299–352. [CrossRef]

153. Amsel, E.; Brock, S. The development of evidence evaluation skills. *Cogn. Dev.* **1996**, *11*, 523–550. [CrossRef]

154. Blanco, N.J.; Sloutsky, V.M. Adaptive flexibility in category learning? Young children exhibit smaller costs of selective attention than adults. *Dev. Psychol.* **2019**, *55*, 2060–2076. [CrossRef] [PubMed]

155. Nokes, T.J.; Schunn, C.D.; Chi, M. Problem solving and human expertise. In *International Encyclopedia of Education*; Elsevier Ltd.: Amsterdam, The Netherlands, 2010; pp. 265–272, ISBN 978-0-08-044894-7.

156. Konold, C.; Robinson, A.; Khalil, K.; Pollatsek, A.; Well, A.; Wing, R.; Mayr, S. Students' use of modal clumps to summarize data. In Proceedings of the Sixth International Conference on Teaching Statistics, Cape Town, South Africa, 7–12 July 2002; p. 6.

157. Saffran, A.; Barchfeld, P.; Alibali, M.W.; Reiss, K.; Sodian, B. Children's interpretations of covariation data: Explanations reveal understanding of relevant comparisons. *Learn. Instr.* **2019**, *59*, 13–20. [CrossRef]

158. Saffran, A.; Barchfeld, P.; Sodian, B.; Alibali, M.W. Children's and adults' interpretation of covariation data: Does symmetry of variables matter? *Dev. Psychol.* **2016**, *52*, 1530–1544. [CrossRef]

159. McNeil, N.M.; Alibali, M.W. Knowledge change as a function of mathematics experience: All contexts are not created equal. *J. Cogn. Dev.* **2005**, *6*, 285–306. [CrossRef]

160. McNeil, N.M.; Grandau, L.; Knuth, E.J.; Alibali, M.W.; Stephens, A.C.; Hattikudur, S.; Krill, D.E. Middle-school students' understanding of the equal sign: The books they read can't help. *Cogn. Instr.* **2006**, *24*, 367–385. [CrossRef]

161. Morris, B.J.; Masnick, A.M.; Baker, K.; Junglen, A. An analysis of data activities and instructional supports in middle school science textbooks. *Int. J. Sci. Educ.* **2015**, *37*, 2708–2720. [CrossRef]

162. Siegler, R.S.; Im, S.; Schiller, L.K.; Tian, J.; Braithwaite, D.W. The sleep of reason produces monsters: How and when biased input shapes mathematics learning. *Ann. Rev. Dev. Psychol.* **2020**, *2*, 413–435. [CrossRef]

163. Shtulman, A.; Walker, C. Developing an understanding of science. *Ann. Rev. Dev. Psychol.* **2020**, *2*, 111–132. [CrossRef]