Running head: SCIENTIFIC REASONING IN A REAL WORLD CONTEXT

Scientific reasoning in a real world context: The effect of prior belief and outcome on children’s hypothesis testing strategies

Steve Croker, Heather Buchanan

1 Illinois State University

2 University of Nottingham

(To appear in the British Journal of Developmental Psychology)

*Requests for reprints should be addressed to Steve Croker, Department of Psychology, Illinois State University, Campus Box 4620, Normal, IL, 61790-4620, USA (e-mail: s.croker@ilstu.edu).
Acknowledgements

We thank Faye Yuill for data collection and Corinne Zimmerman for comments on an earlier draft of the paper. We are grateful to the parents, teachers, and children of the participating schools. This research was supported by a University of Nottingham Research Enhancement Award to Heather Buchanan.
Abstract

The impact of event outcome and prior belief on scientific reasoning was investigated within a real-world oral health context. Participants ($N = 144$; ranging from 3 to 11 years) were given hypothesis-testing tasks and asked to explain their answers. Participants were presented with information that was either consistent or inconsistent with their own beliefs. Each task consisted of scenarios in which the outcome was either good or bad oral health. When the information was belief-consistent and the outcome was good, or when the information was belief-inconsistent and the outcome was bad, children were more likely to choose scientifically appropriate tests of the stated hypothesis (i.e., manipulate only one variable). Evidence-based explanations were associated with scientifically appropriate choices in the good-outcome, belief-inconsistent scenario and the belief-consistent, bad outcome scenario. Participants’ performance on these tasks is explained by considering the plausibility of causal variables. A control of variables strategy was used to test hypotheses in cases in which the evidence was consistent with participants’ beliefs and knowledge of causal mechanisms. In contrast, when the evidence was inconsistent with participants’ beliefs, children chose to manipulate behaviours likely to lead to a positive health outcome. These findings demonstrate that context and prior knowledge interact to play an important role in children’s scientific reasoning.
Scientific Reasoning in a Real World Context: The Effect of Prior Belief and Outcome on Children’s Hypothesis Testing Strategies.

Scientific thinking and reasoning skills underpin achievement in science education and the development of these skills is fundamental to becoming a scientifically literate adult. Research into children’s scientific thinking and reasoning is multifaceted and includes the investigation of the formation and revision of theories and the principles of scientific inquiry (see Zimmerman, 2007, for a comprehensive review). The ability to select the correct test of a hypothesis is an integral part of the scientific reasoning process, which is conceived of as including hypothesis generation, evidence evaluation and experimental design skills (Klahr & Dunbar, 1988).

Hypothesis testing involves searching through a space of experiments in order to select a test that will discriminate among competing hypotheses. In order to achieve this, predictions about experimental results must be made (Klahr, Fay & Dunbar, 1993). Prior knowledge can often guide the selection of predictions and the strategy for selecting a hypothesis. This is evidenced by the finding that participants in scientific reasoning studies often select or design experiments that will generate data that will potentially confirm, rather than disconfirm, their hypotheses (Klayman & Ha, 1987). An important first step in the scientific reasoning chain is to select a test of a hypothesis which will generate valid data, regardless of whether the data confirm or disconfirm the hypothesis. This entails using the Control of Variables Strategy (CVS), a domain-general processing strategy in which single variables are manipulated and a distinction between confounded and unconfounded experiments can be made. Use of the CVS gives us the ability to draw valid inferences from experiments by ensuring that the values of multiple variables are not manipulated at the same time (Chen & Klahr, 1999).
Findings from research on children’s ability to design or select unconfounded experiments tend to reveal a pattern of poor performance prior to adolescence (e.g., Kuhn, Amsel, & O’Loughlin, 1988; Kuhn, Garcia-Mila, Zohar & Anderson, 1995; Schauble, 1996). In these studies, children were unable to design experiments which would allow valid inferences to be made and had difficulty evaluating evidence with respect to hypotheses. Klahr et al. (1993) agree that children have limited domain-general reasoning skills compared to adults, but argue that when hypotheses are plausible and there are few alternatives to choose from, children are able to succeed in scientific reasoning tasks. A number of studies demonstrate that domain-specific knowledge is also a relevant factor (e.g., Carey, 1985; Chi & Koeske, 1983; Penner & Klahr, 1996) and recent approaches to scientific reasoning take into account both domain-general strategies and domain-specific concepts, suggesting that the two bootstrap one another (Lehrer & Schauble, 2006; Schauble, 1996; Zimmerman, 2007).

Many studies have demonstrated children’s successful performance on scientific reasoning tasks with simpler task demands (e.g. Ruffman, Perner, Olson, & Doherty, 1993; Samarakungavan, 1992; Sodian, Zaitchik, & Carey, 1991). However, few of these simple tasks offer contexts that are familiar or meaningful to children. Given the relationship between domain-specific knowledge and domain-general strategies, context should have important implications for performance on reasoning tasks. Tschirgi (1980) examined hypothesis testing in familiar contexts, providing participants with scenarios such as baking a cake and making a paper airplane. In these scenarios the outcome could either be positive or negative (e.g., a great cake or a terrible cake). Tschirgi tested the prediction that outcome is important in hypothesis testing performance. If one variable is hypothesised to be the cause of a bad outcome, people will change that variable alone. However, if the outcome is positive, there is a desire to reproduce the
effect by holding the variable hypothesised to cause the good outcome constant. Tschirgi found children were more likely to use a CVS or VOTAT (vary-one-thing-at-a-time) strategy when presented with a bad outcome and a HOTAT (hold-one-thing-at-a-time) strategy when presented with a good outcome. If there are three variables that can be manipulated, one is held constant, and two are manipulated, then the experiment is confounded as any change in outcome may be due to either one of the changes or an interaction between the two. Younger children (6-7 year olds) were also more likely to ‘change all’ in a bad outcome situation in attempt to make a change for the better. Similarly, Zimmerman and Glaser (2001) provided 11-12 year-olds with hypotheses about whether certain liquids were good or bad for plant health. When a negative outcome was presented, participants tended to propose experiments using a VOTAT/CVS strategy, whereas a positive outcome led to attempts to preserve the outcome by manipulating other variables (e.g., type of plant). The apparent aim to produce a positive outcome, reported in these studies, may reflect a distinction between an engineering approach and a hypothesis testing approach. When adopting an engineering approach, the aim is to engineer a desired outcome rather than to test the causal status of individual variables (Schauble, Klopfer & Raghavan, 1991).

Prior belief is another factor which has been shown to have an effect on scientific reasoning. Successful scientific reasoning often involves suspending prior beliefs when the evidence encountered contradicts these beliefs (Greenhoot, Semb, Colombo, & Schreiber, 2004). In evidence evaluation tasks in which participants are asked to interpret patterns of covariation data, Kuhn et al. (1988) and Amsel and Brock (1996) found that children’s answers were biased by their prior beliefs. In addition, Kuhn et al. found that even when the covariation data was correctly interpreted, participants justified their answers in terms of new beliefs about the causal
mechanisms which would lead a variable to have an effect. Kuhn (1989) interpreted this as an inability to consider evidence independently of a theory or belief. Koerber, Sodian, Thoermer, and Nett (2005) also demonstrated an effect of prior belief on evidence evaluation. They found superior performance on belief-neutral tasks compared to tasks in which the participants had prior beliefs which conflicted with the evidence presented. Klaczynski (2000; Klaczynski & Narasimbam, 1998) offers a dual-process account of scientific reasoning, according to which evidence congruent with prior beliefs is evaluated heuristically and accepted whereas evidence incongruent with prior beliefs is rejected. The latter can occur in two ways. Rejection can be heuristic, based on the implausibility of the evidence, or a more cognitively expensive analytic process can be used to reject the evidence in a principled way.

These effects of context suggest that domain-general reasoning processes, such as the control of variables present in the VOTAT strategy, can be mediated by domain-specific knowledge regarding the plausibility of a hypothesis. What is not clear is exactly how domain-specific knowledge facilitates or inhibits the proficient use of hypothesis testing. In sum, previous research has demonstrated that both outcome and prior belief have an effect on children’s scientific reasoning and that responses are often based on domain-specific knowledge. However, studies examining the effect of prior beliefs and whether responses are justified in terms of these beliefs have tended to focus on just one aspect of scientific reasoning: evidence evaluation. There is little research examining how hypothesis testing strategies are affected by prior belief.

The aims of the present study were to further explore the factors that influence developmental trends in scientific reasoning strategies on a hypothesis-testing task. We examined the responses given in a contextualized scientific reasoning task in which children
have strong prior beliefs. Oral health was selected for two key reasons: (a) It is a domain in which behaviours have a direct causal link with outcomes (e.g., regular consumption of sugar results in tooth decay), and (b) from an early age, children have beliefs regarding this link and the behaviours which mediate it (Croker & Buchanan, 2008). Specifically, we predict that whether the evidence is consistent or inconsistent with prior beliefs will have an effect on strategy choice and, in line with Tschirgi’s findings, that a good or bad outcome will also affect strategy choice.

In order to test these predictions, an experimental task was devised in which pictures of oral health behaviours (brushing teeth, visiting the dentist and drinking either cola or milk) were presented alongside pictures of healthy (good outcome) or unhealthy (bad outcome) teeth. The task involved choosing the set of behaviours that should be carried out in order to test the hypothesis. Although previous research has examined the effect of outcome on forced-choice hypothesis testing tasks (e.g., Tschirgi, 1980), these studies have not asked children why they made the choices they did. In a series of evidence evaluation tasks, Kuhn et al. (1988) asked participants to justify their answers; these were coded as being either belief-based or evidence-based. Using similar coding categories, we examined the explanations that children gave for their answers in order to assess whether their hypothesis-testing strategies were driven by their domain-specific knowledge (belief-based) or whether they reasoned independently of the task content (evidence-based).

Method

Pilot of materials

In order to ensure that the healthy and unhealthy teeth pictures we used in the study were perceived to be healthy or unhealthy, we asked adults (n = 38) and children (n = 28; aged 7-8) to
To ensure that we presented drinks perceived to be good or bad for teeth, participants rated pictures of eight drinks (coffee, orange juice, strawberry milkshake, water, tea, cola, milk and hot chocolate). Participants were asked to score how good or bad each drink is for teeth on a 5-point likert scale (1 = very bad, 5 = very good). On the basis of these ratings two pictures of teeth were selected to illustrate healthy and unhealthy teeth. Cola was rated as the most unhealthy drink for teeth ($M = 1.34$, $SD = 0.64$); water ($M = 4.58$, $SD = 0.66$) and milk ($M = 4.34$, $SD = 0.99$) were both rated as healthy. Milk was chosen for the subsequent study as it contains calcium, thus drinking it may be perceived as a more positive oral health behaviour.

**Participants**

Participants were recruited from two primary schools in the East Midlands of England, UK. A total of 144 children aged between 3 and 11 took part in the study, 32 children from foundation and reception classes ($M = 56$ months, $SD = 7.84$, range = 40 – 66 months), 25 year one and year two children ($M = 78$ months, $SD = 6.86$, range = 69 – 89 months), 39 year three and year four children ($M = 100$ months, $SD = 6.37$, range = 90 – 115 months) and 48 year five and year six children ($M = 130$ months, $SD = 6.49$, range = 116 – 139 months). We will henceforth refer to these four groups as 4-, 6-, 8- and 10-year olds.

**Materials**

Materials consisted of four story problems illustrated as sets of pictures printed on sheets of A4 paper. In each story, a character was presented and their oral health practices were described. Each set of pictures showed three variables (what type of drink was consumed: milk or cola, whether they brushed their teeth or not and whether they visited the dentist or not). An
outcome (healthy or unhealthy teeth) was displayed on a separate sheet of paper. There were four hypothesis-testing tasks in all: Two in which the evidence presented was consistent with participants’ beliefs (belief-consistent condition) and two in which the evidence presented was inconsistent with their beliefs (belief-inconsistent condition). For both sets of tasks, one featured a positive outcome (healthy teeth) and the other featured a negative outcome (unhealthy teeth). Table 1 contains a list of the information presented in the four stories. In each story, the character was described as having a hypothesis about the cause of their good or bad oral health. These hypotheses always concerned the type of drink they consumed; the character always thought that whether they brushed their teeth and whether they visited the dentist were non-causal. A set of three pictures was presented and participants were asked to pick the one that illustrated which behaviours should be changed and which should stay the same in order to test this hypothesis. Figure 1 illustrates the belief-consistent, good-outcome condition. In this case, the character’s hypothesis is that the reason she has healthy teeth is because she drinks milk. Participants could choose from a set of pictures showing a pattern of variables in which only the hypothesized variable (the type of drink) was manipulated (vary-one-thing-at-a-time; VOTAT), a set showing the hypothesized variable as held with the other two changed (hold-one-thing-at-a-time; HOTAT), and a set showing all variables changed (change-all, CA). Children’s choices of pictures were noted by a research assistant and the whole session (tasks, interviews and explanations) was recorded using a digital voice recorder. All the data were transcribed verbatim.

(Table 1 about here)

(Figure 1 about here)
Design and Procedure

Half the participants were presented with stories in which the evidence presented was consistent with their own beliefs (belief-consistent) and the other half were presented with stories in which the evidence was inconsistent with their beliefs (belief-inconsistent). Each participant completed two tasks, one with a good outcome and one with a bad outcome. In addition, participants were interviewed to determine their oral health knowledge, beliefs and practice. They were asked what people can do to make sure their teeth are healthy, whether there are things that people do that might make their teeth unhealthy, what drinks are good or bad for teeth, whether they had been to the dentist, what the dentist does, whether they brushed their teeth, and why they brushed their teeth.

The interview ensured that the participants’ beliefs were consistent with the evidence presented in the experimental tasks (e.g., that cola is bad for teeth). Interviews were counterbalanced such that half the children were asked these questions before they carried out the hypothesis-testing tasks and half were asked these questions afterwards.

It has been noted that having both behavioural and explanation data can help clarify children’s reasoning (Atance & Meltzoff, 2005; Wellman & Liu, 2007) and Legare, Wellman, and Gelman (2009) found that children’s explanations demonstrate a more sophisticated understanding than children’s predictions in the domain of naïve biology. Therefore, following each hypothesis-testing task, participants were asked to give explanations for their answers in order to provide insight into the reasoning behind their answers. As soon as they had picked one of the three possible answers, they were asked “Why do you think he/she should do that?” Explanations were coded as being either evidence-based or theory-based. Explanations in which
participants mentioned proof (e.g., they explicitly state that the reason the protagonist should choose a particular course of action is in order to ‘prove their point’ or ‘will prove that it is/isn’t the drink’) or looking for an effect (e.g., changing drink to show that teeth may become less healthy) were coded as evidence-based explanations. If participants explained that they had chosen an answer because that was the correct behaviour (e.g., you should brush your teeth because it is good for you; drinking cola is bad for teeth) then it was coded as theory-based. The coding scheme used and examples of explanations are shown in the Appendix. All explanations were coded by one rater (SC); data from 30 participants were also coded by a second rater (HB). Inter-rater reliability was good for both the good \((K = 0.78)\) and bad \((K = 0.79)\) outcome scenarios. All children were debriefed at the end of the study, and we ensured that they understood which practices were healthy and unhealthy.

Results

Interviews

One of the participants did not answer any of the interview questions. The remaining 143 participants demonstrated a good basic knowledge of oral health behaviours and outcomes. When asked which drinks were bad for teeth, 90% mentioned sugary and/or fizzy drinks, with 59% specifically mentioning cola. The majority (91%) stated they had visited a dentist, and 85% were able to offer an explanation of why people see a dentist. Explanations included checking teeth, health reasons, and treatment. All but 5% claimed they brushed their teeth, and 85% were able to explain why people brush their teeth, referring to reasons such as health, appearance and cleanliness of teeth. The order of the brief oral health interview (before or after the hypothesis-testing tasks) did not affect responses for either the bad outcome, \(\chi^2(2) = .05, p = .98\), or good outcome, \(\chi^2(2) = 2.01, p = .37\), scenarios.
Hypothesis Testing

In the hypothesis-testing tasks, use of the Control of Variables Strategy is indicated by a VOTAT answer. There was a higher rate of CVS use for the good outcome tasks in the belief-consistent condition (43% VOTAT answers) compared to the belief-inconsistent condition (14% VOTAT answers). The pattern was reversed for the bad outcome tasks, with a VOTAT rate of only 7% in the belief-consistent condition and 67% in the belief-inconsistent condition. Tables 2 and 3 show how many participants gave VOTAT answers in the good outcome and bad outcome tasks respectively. Six children did not give an answer in the good outcome task and six children did not give an answer in the bad outcome task.

A three-way loglinear analysis of condition, age and VOTAT answers for the good outcome task produced a final model that retained all effects. The likelihood ratio of this model was $\chi^2(0) = 0, p = 1$, indicating that the highest-order interaction (condition x age x VOTAT answers) was significant, $\chi^2(3) = 11.04, p = .01$, with participants more likely to give VOTAT answers in the belief-consistent condition than the belief-inconsistent condition. Although there was no main effect of age on VOTAT answers, $\chi^2(1) = 1.21, p = .75$, separate tests revealed that the 4-year-olds ($p = .005$, Fisher’s exact test), 6-year-olds ($p = .04$, Fisher’s exact test), and 8-year-olds, $\chi^2(1) = 9.63, p = .003$, were more likely to give VOTAT answers in the belief-consistent condition than in the belief-inconsistent condition, whereas there was no effect of condition for the 10-year-olds, $\chi^2(1) = 0.03, p = 1$.

A three-way loglinear analysis for the bad outcome task produced a final model that retained the condition x VOTAT answers interaction. The likelihood ratio of this model was $\chi^2(9) = 10.79, p = .29$. The condition x VOTAT answers interaction was significant, $\chi^2(1) = 60.05, p < .001$, with participants more likely to give VOTAT answers in the belief-inconsistent
condition than the belief-consistent condition. There was no main effect of age on VOTAT answers, $\chi^2(1) = 2.61, p = .46$.

(Table 2 about here)

(Table 3 about here)

When the percentage of children providing VOTAT responses across the four conditions of the experimental manipulation is considered, an overall pattern is evident. Figure 2 shows the percentage of VOTAT answers across different conditions, collapsed across age. Participants were most likely to select the VOTAT choice when the information presented in the task was consistent with their beliefs and there was a good outcome, or when the information was inconsistent with their beliefs and a bad outcome was presented.

(Figure 2 about here)

Explanations

The explanations given by participants were coded as either evidence-based or theory-based. Although not all participants made a choice in the hypothesis-testing tasks, they were still asked to explain their answers. For the good outcome task, 32% of participants gave evidence-based answers and 37% gave theory-based answers in the belief-consistent condition; 23% of participants gave evidence-based answers and 58% gave theory-based answers in the belief-inconsistent condition. For the bad outcome task, 15% of participants gave evidence-based answers and 72% gave theory-based answers in the belief-consistent condition; 26% of participants gave evidence-based answers and 38% gave theory-based answers in the belief-inconsistent condition. Tables 4 and 5 show how many participants gave each type of explanation in each condition, regardless of their responses on the task. In the good outcome scenario, there was a main effect of condition, $\chi^2(1) = 3.77, p = .05$, with more participants
giving a theory-based explanation in the counter-intuitive condition. The effect was reversed in the bad outcome scenario with more participants giving a theory-based explanation in the intuitive condition, $\chi^2(1) = 6.90, p = .01$. There were effects of age group on explanation for both the good outcome task, $\chi^2(1) = 35.99, p < .001$, and the bad outcome task, $\chi^2(1) = 29.70, p < .001$, with older children giving more evidence-based answers than younger children.

(Table 4 about here)

(Table 5 about here)

Tables 6 and 7 show both the types of answers given and the types of explanation given for the good and bad outcome conditions respectively. The data for participants who did not give an explanation for their answer are not included. Participants who gave VOTAT answers in the belief-consistent/good outcome scenario were more likely to give theory-based explanations than evidence-based explanations, $\chi^2(1) = 14.89, p < .001$, whereas the participants who gave VOTAT answers in the belief-inconsistent/good outcome scenario were more likely to give evidence-based explanations, $\chi^2(1) = 18.69, p < .001$. The situation was reversed in the bad outcome scenario. Participants who gave VOTAT answers in the belief-consistent condition were more likely to give evidence-based explanations, $\chi^2(1) = 25.22, p < .001$, and those who gave VOTAT answers in the belief-inconsistent condition were more likely to give theory-based explanations, $\chi^2(1) = 7.41, p = .007$. There were alternatives to VOTAT answers that participants could choose, one in which one variable is held while the other two are changed, and one in which all three variables are changed. The frequencies with which all three response types were made are shown in Tables 4 and 5. It is clear that in some cases, there was a strong preference for one of the non-VOTAT reasoning strategies. In the belief-consistent/good outcome scenario, the HOTAT response is the most frequent and co-occurs with evidence-based answers, whereas in
both the belief-inconsistent/good outcome and belief-consistent/bad outcome scenarios, the CA response is the most frequent and co-occurs with theory-based answers.

(Table 6 about here)

(Table 7 about here)

Discussion

We set out to examine the responses given in a contextualized scientific reasoning task in which children have strong prior beliefs. The results show that the plausibility of the evidence with respect to prior knowledge affects the strategies children use in hypothesis testing. We also replicated and extended the effect of outcome (good vs. bad). When the information presented was consistent with prior belief and the outcome was good (e.g., positive oral health behaviours lead to healthy teeth), or when the information was inconsistent with prior belief and the outcome was bad (e.g., positive oral health behaviours lead to unhealthy teeth) children were able to choose scientifically appropriate tests of the stated hypothesis (i.e., manipulate one variable). In the opposing cases they chose inappropriate tests such as changing all the variables. Theory-based explanations were associated with non-VOTAT answers, and evidence-based explanations were associated with VOTAT answers, when the evidence presented was belief-inconsistent and the outcome was good, and when the evidence presented was belief-consistent but the outcome was bad. In the scenarios in which participants were presented with negative oral health behaviours they tended to select answers in which the protagonist changes to positive oral health behaviours and the majority of the explanations given were theory based, regardless of plausibility of the hypothesis or the outcome. This suggests that participants were relying on domain-specific knowledge and not trying to disconfirm the hypothesis or consider the evidence. There were no main effects of age on whether participants chose the logically appropriate
VOTAT response or not, but there were effects of age on explanations for both the good and bad outcome scenarios with older children giving more evidence-based answers than younger children.

Prior research falls into two categories: simple and complex tasks. When given a choice between two tests when only one variable is manipulated, and when prior belief is not a factor, children as young as 6 years old can choose between a conclusive and inconclusive test of a hypothesis (Sodian et al., 1991). However, in multivariable studies (e.g., four or more variables) in which participants were required to design their own experiments, or where a large number of tests are possible (Kuhn, Amsel, & O'Loughlin, 1988; Kuhn, Garcia-Mila, Zohar & Anderson, 1995; Schauble, 1996), children tend to produce confounded tests. Our study used a methodology with a level of complexity between these two extremes: three variables and a choice of three tests. Given this mid-level problem space, even some 4- and 5-year-olds are capable of choosing a non-confounded test. However, the scenario which resulted in the highest level of task success was a bad outcome and evidence inconsistent with prior beliefs and poorest performance was in the bad outcome /belief-consistent condition. This suggests a strong interaction between hypothesis-testing strategy selection, domain-specific knowledge, and outcome. Although the effect of outcome is a robust finding (e.g., Tschirgi, 1980; Zimmerman & Glaser, 2001) and is not, therefore surprising, both the choice of scientifically appropriate tests and the impact of prior belief are important with respect to our understanding of children’s abilities and their susceptibility to the influence of knowledge.

One explanation is that this strong effect of belief is indicative of an immature understanding of the relationship between theory and evidence. Kuhn (1989) suggests that children are theory bound in that they ignore discrepant evidence or attend to it selectively. They
may also adjust the evidence to fit in with their theories. This is because they do not differentiate between theory and evidence in a metacognitive way, particularly in cases where they have prior beliefs (or, at least, theories that they favour) and when there are multiple causes. She argues that children need a plausible causal link between a factor and its results before they can accept the data, whereas evidence should be acknowledged and then a theory should be constructed which takes the pattern of evidence into account.

An alternative view is presented by Koslowski (1996). She argues that within the scientific reasoning literature there is a reliance on the covariation aspect of scientific thinking at the expense of other principles (such as theories and mechanisms) and that if these principles are taken into account, people are better at scientific reasoning. She notes that there are lots of examples of covariation/correlation in the world, but we only take seriously those in which there may be a causal mechanism. This consideration allows us to decide which correlations may indicate causality and which do not. Considering causal mechanisms requires a knowledge base (e.g., theories describing the mechanism). Thus, in order to assess scientific reasoning, we need to ascertain whether participants use their knowledge in a methodical way. Koslowski argues that many of the responses judged to be scientifically flawed by Kuhn et al. (1988) are, in fact, reasonable as theories and mechanisms provide a plausible way for mediation between cause and effect. When participants give theory-based answers, they are explaining why they give weight to some patterns of covariation and not to others. Koslowski suggests that where there are flaws in reasoning, it is a result of information-processing limitations rather than an inability to coordinate theory and evidence. Thus, this view of scientific reasoning conceptualizes the inclusion of a consideration of plausible mechanisms as a useful way of excluding artifactual
causes and deciding whether a theory requires outright rejection or modification in the face of disconfirming evidence (Koslowski & Masnick, 2002).

Our findings offer support for Koslowski’s (1996) argument that a consideration of theories and causal mechanisms is an important part of the scientific reasoning process. The pattern of results for children’s explanations of their choices suggest that they use their knowledge of factors leading to good or bad oral health when presented with a good outcome and evidence which is inconsistent with their beliefs, or when presented with a bad outcome and evidence consistent with their beliefs. In both cases this was linked to scientifically inappropriate choices which did not test the hypothesis presented, but did represent the behaviours that, in reality, are likely to lead to a positive oral health outcome. In the remaining two conditions (good/belief-consistent and bad/belief-inconsistent) theory-based answers coincided with selection of the VOTAT strategy. In the former case, the evidence confirmed the children’s beliefs. However, in the latter case, the test chosen by the majority of participants whose explanations were based on their knowledge of oral health meant changing from drinking milk to drinking cola. Although this does seem counterintuitive, an examination of the explanations given reveals that the VOTAT answer was chosen as the least damaging choice. In this condition, the participants were shown a scenario in which a character carried out three positive health behaviours. The three choices available involved changing one (VOTAT), two (HOTAT) or all 3 (CA) of these behaviours. Many of the explanations given for the VOTAT choice were that drinking cola would be acceptable as long as the story character continued to brush his teeth and visit the dentist. So, although the VOTAT option is the logically appropriate choice, participants chose it for reasons based entirely on their understanding of oral health and
knowledge of causal mechanisms that hold true in everyday life, not because they were employing the Control of Variables Strategy.

These findings are also consistent with Klaczynski’s (2000) dual process model. According to this account, evidence that is consistent with prior beliefs is processed heuristically whereas evidence that is inconsistent with prior beliefs can be processed either heuristically or analytically. In the belief-consistent condition, the good and bad outcomes lead easily to the HOTAT and CA choices respectively, both of which were common responses. In the belief-inconsistent condition, the most common response patterns were to select the CA answer for the good outcome and the VOTAT answer for the bad outcome. The former can be explained within Klaczynski’s model as a heuristic rejection of the evidence as implausible; the latter may be evidence for analytic processing of the evidence. Across all four conditions, evidence-based answers occurred most frequently with the VOTAT and HOTAT responses, which suggests that these answers may be evidence for analytic reasoning whereas theory-based responses (regardless of strategy choice) is indicative of heuristic reasoning.

A further theoretical approach that may explain the pattern of data presented here is Bayesian inference (e.g., Schulz, Bonawitz & Griffiths, 2007; Schulz, Goodman, Tenenbaum & Jenkins, 2008). Schulz and colleagues provide an account of causal learning in which evidence interacts with prior beliefs bidirectionally; prior beliefs influence the interpretation of evidence, and evidence leads to belief revision. On this view, strongly held domain-specific prior beliefs can have primacy over domain-general reasoning with covariation evidence. However, in order to provide evidence consistent with this account, participants would have to be presented with multiple instances of covariation evidence that is belief-inconsistent. The Bayesian approach
would be supported if, subsequent to increased strength of evidence that runs counter to prior beliefs, the probability of choosing answers consistent with domain-general reasoning increased.

Another interesting finding from the analysis of explanations is that in the good-outcome/belief consistent condition, evidence-based explanations most frequently occurred with HOTAT responses. Although this is not the logically correct response, it is a sensible one in light of the context. This response can either be interpreted as an effort to engineer a desirable outcome, or as evidence for a confirmation bias in that participants retained the factor (drinking milk) hypothesised to account for the good outcome. However, the HOTAT choice can also be interpreted as a positive test strategy which is a legitimate way of generating evidence to determine the truth status of a hypothesis (Klayman & Ha, 1987). In order to determine the extent to which children employ an engineering approach, future research could employ a modified version of this task in which a scientist proposes confounded tests of the hypotheses. Schultz and Bonawitz (2007) have found that children as young as 4 years old are able to discriminate between confounded and unconfounded evidence. Children who use a scientific approach to the problem should, therefore, be able to identify poor tests of a hypothesis. Alternatively, participants could be induced into adopting either the scientific or engineering approach by asking them to pretend that they are either scientists who are trying to understand the effect of different variables on oral health, or dentists who are trying to ensure that their child patients have healthy teeth.

The interesting findings from this study are mostly a result of the knowledge-rich scenario. It would, therefore, be interesting to contrast the task presented here with an isomorphic task in which there are no prior beliefs in order to determine the extent to which responses and explanations are affected by belief. Similarly, research using other domains can
be used to replicate and extend these findings. Furthermore, as we have used a novel combination of good/bad outcome and belief-consistent/belief-inconsistent evidence, it is not known how adults would perform under the same conditions. Choosing a VOTAT answer in this task requires participants to suggest that the protagonist consumes drinks that could have a detrimental effect on her oral health. It may be the case that adults would also be unwilling to advocate this pattern of behaviour as a test of a hypothesis. Future research using adult participants would enable a baseline measure of performance to be established, against which a developmental trajectory could be plotted.

In conclusion, this study used a novel combination of good vs. bad outcome and a context that is meaningful to children to advance our knowledge of children’s scientific reasoning. Our data replicate the robust effect of outcome and demonstrate a strong effect of prior belief. These findings show that an analysis of explanations can provide an insight into how children think about hypotheses and evidence and offer support for the importance of the consideration of casual mechanisms as an integral part of the scientific reasoning process.
References


Schulz, L., Bonawitz, E.B., & Griffiths, T.L. (2007) Can being scared give you a tummy ache?


Appendix: Coding scheme used for explanations and examples of responses

**Theory-based explanations**

1) Can have fizzy drinks/cola if you brush your teeth and/or visit the dentist

Example: “because visiting the dentist and brushing his teeth is still really good to do… so it doesn’t matter if you drink coke much”

2) Brushing teeth/visiting the dentist/drinking milk is a good thing to do

Example: “brushing the teeth is really healthy, visiting the dentist is really healthy and because drink milk is really good for your teeth”

3) Cola is bad for you / has sugar

Example: “milk well is, well it’s got its own sweeteners and it’s natural, and things and cola well it’s not got it’s natural sweeteners and it’s just got lots and lots of sugar”

4) So teeth don’t fall out

Example: “because it’s really important and you don’t want your teeth to fall out and if you don’t brush your teeth all your teeth will fall out and you won’t be able to eat”

**Evidence-based explanations**

1) Proving a point / proving that the drink is/is not having an effect
Example: “if her teeth…don’t stay the same as those it will prove it’s not the cola, and if they stay the same it will prove that it is the cola”

2) Wants to know or show what is making teeth healthy/unhealthy

Example: “cause she wants to know… what is making her teeth healthy… if it is the milk”

No answer

1) Don’t Know / Can’t Explain

Example: “I’m not sure”

2) Restates answers

Example: “brush his teeth, go to the dentist”
Table 1

*The four story problems*

<table>
<thead>
<tr>
<th>Story Type</th>
<th>Brushing</th>
<th>Dentist</th>
<th>Drink</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belief-consistent; good outcome</td>
<td>Brushes teeth</td>
<td>Visits dentist</td>
<td>Milk</td>
<td>Healthy teeth</td>
</tr>
<tr>
<td>Belief-consistent; bad outcome</td>
<td>Doesn’t brush teeth</td>
<td>Doesn’t visit dentist</td>
<td>Cola</td>
<td>Unhealthy teeth</td>
</tr>
<tr>
<td>Belief-inconsistent; good outcome</td>
<td>Doesn’t brush teeth</td>
<td>Doesn’t visit dentist</td>
<td>Cola</td>
<td>Healthy teeth</td>
</tr>
<tr>
<td>Belief-inconsistent; bad outcome</td>
<td>Brushes teeth</td>
<td>Visits dentist</td>
<td>Milk</td>
<td>Unhealthy teeth</td>
</tr>
</tbody>
</table>
Table 2

*Success rates on good outcome task across age groups*

<table>
<thead>
<tr>
<th>Age</th>
<th>Belief-Consistent</th>
<th>Belief-Inconsistent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4-year-olds</td>
<td>6-year-olds</td>
</tr>
<tr>
<td>Pass</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>Fail</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td>12</td>
<td>12</td>
</tr>
</tbody>
</table>
### Table 3

**Success rates on bad outcome task across age groups**

<table>
<thead>
<tr>
<th></th>
<th>Belief-Consistent</th>
<th>Belief-Inconsistent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4-year-olds</td>
<td>6-year-olds</td>
</tr>
<tr>
<td>Pass</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Fail</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>Total</td>
<td>14</td>
<td>13</td>
</tr>
</tbody>
</table>
Table 4

Explanations given for answers on good outcome task.

<table>
<thead>
<tr>
<th></th>
<th>Belief-Consistent</th>
<th></th>
<th></th>
<th></th>
<th>Belief-Inconsistent</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4-year-olds</td>
<td>6-year-olds</td>
<td>8-year-olds</td>
<td>10-year-olds</td>
<td>4-year-olds</td>
<td>6-year-olds</td>
<td>8-year-olds</td>
</tr>
<tr>
<td>Evidence-based</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Theory-based</td>
<td>4</td>
<td>7</td>
<td>10</td>
<td>5</td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>No answer</td>
<td>11</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>15</td>
<td>13</td>
<td>20</td>
<td>23</td>
<td>17</td>
<td>12</td>
<td>19</td>
</tr>
</tbody>
</table>
Table 5

*Explanations given for answers on bad outcome task*

<table>
<thead>
<tr>
<th></th>
<th>Belief-Consistent</th>
<th>Belief-Inconsistent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4-year-olds</td>
<td>6-year-olds</td>
</tr>
<tr>
<td>Evidence-based</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Theory-based</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>No answer</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>15</td>
<td>13</td>
</tr>
</tbody>
</table>
Table 6

*Types of explanation given as a function of the types of answers for the good outcome condition*

<table>
<thead>
<tr>
<th>Condition</th>
<th>Belief-Consistent</th>
<th>Belief-Inconsistent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Answer</td>
<td>Evidence-based</td>
<td>Theory-based</td>
</tr>
<tr>
<td>VOTAT</td>
<td>3</td>
<td>17</td>
</tr>
<tr>
<td>HOTAT</td>
<td>20</td>
<td>7</td>
</tr>
<tr>
<td>Change All</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>23</td>
<td>25</td>
</tr>
</tbody>
</table>
Table 7

*Types of explanation given as a function of the types of answers for the bad outcome condition*

<table>
<thead>
<tr>
<th>Condition</th>
<th>Belief-Consistent</th>
<th>Belief-Inconsistent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Answer</td>
<td>Evidence-based</td>
<td>Theory-based</td>
</tr>
<tr>
<td>VOTAT</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>HOTAT</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Change All</td>
<td>3</td>
<td>45</td>
</tr>
<tr>
<td>Total</td>
<td>11</td>
<td>51</td>
</tr>
</tbody>
</table>
Figure Captions

*Figure 1.* Hypothesis testing task. Participants were presented with pictures relating to three oral health behaviours and presented with an outcome (healthy or unhealthy teeth). They were told that the character in the story believed that his healthy/unhealthy teeth resulted from his choice of drink. Participants were asked what the story character could do to prove his point and given three choices: Vary just the type of drink, vary the other two factors (toothbrushing and dentist), or vary all three factors. The labels VOTAT, HOTAT, and Change All are provided here for clarity; these were not seen by the participants.

*Figure 2.* Percentage of VOTAT answers for each condition (collapsed across age groups). Good/bad outcome is a within-participants variable; the grey bars represent the responses of participants in the belief-consistent condition and the black bars represent responses of participants in the belief-inconsistent condition.
Figure 1

3 choices:

- VOTAT
- HOTAT
- Change
- All

Figure 2

% pass

<table>
<thead>
<tr>
<th>Condition</th>
<th>% Pass</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good outcome / belief-consistent</td>
<td>40</td>
</tr>
<tr>
<td>Good outcome / belief-inconsistent</td>
<td>10</td>
</tr>
<tr>
<td>Bad outcome / belief-consistent</td>
<td>5</td>
</tr>
<tr>
<td>Bad outcome / belief-inconsistent</td>
<td>70</td>
</tr>
</tbody>
</table>