

Middle School Students' Approaches to Reasoning about Disconfirming Evidence

Keisha Varma¹, Martin Van Boekel¹ & Sashank Varma¹

¹University of Minnesota, 250 Education Sciences Building, Minneapolis, MN 55455, USA

Correspondence: Keisha Varma, Educational Psychology, 250 Education Sciences Building, 56 East River Rd, Minneapolis, MN 55455, USA. E-mail: keisha@umn.edu

Received: November 10, 2017 Accepted: December 31, 2017 Online Published: January 19, 2018

doi:10.5539/jedp.v8n1p28

URL: <http://doi.org/10.5539/jedp.v8n1p28>

Abstract

This study investigated differences in how middle school children reason about disconfirming evidence. Scientists evaluate hypotheses against evidence, rejecting those that are disconfirmed. Although this *instant rationality* propels empirical science, it works less for theoretical science, where it is often necessary to *delay rationality* – to tolerate disconfirming evidence in the short run. We used behavioral measures to identify two groups of middle-school children: *strict reasoners* who prefer instant rationality and quickly dismiss disconfirmed hypotheses, and *permissive reasoners* who prefer delayed rationality and retain disconfirmed hypotheses for further evaluation. We measured their scientific reasoning performance as well as their cognitive ability and motivational orientation. What distinguished the groups was not overall differences in these variables, but their predictive relation. For strict reasoners, better scientific reasoning was associated with faster processing, whereas for permissive reasoners, better scientific reasoning was associated with more deliberate thinking – slower processing and broader consideration of both disconfirmed and alternate hypotheses. These findings expand our understanding of “normative” scientific reasoning.

Keywords: scientific reasoning, middle school students, executive function, working memory

1. Introduction

This study investigated differences in how middle school children reason about disconfirming evidence. We hypothesized that some children are naturally inclined towards fallibilism (Popper, 1963). These *strict reasoners* prefer instant rationality, and quickly dismiss hypotheses that have been disconfirmed. We hypothesized that other children are *permissive reasoners*: They prefer delaying rationality, and are willing to retain hypotheses in the face of disconfirming evidence for further evaluation.

We developed behavioral measures of strict versus permissive reasoning about disconfirming evidence by recasting a finding from the judgment and decision-making literature. *Hindsight bias* is the tendency for recollections of earlier predictions to shift towards the actual outcome of the event after outcome knowledge is provided (Fischhoff, 1975; Slovic & Fischhoff, 1977). It is typically understood as a deviation from normative thinking. We propose re-interpreting it as evidence of a preference for delaying rationality. When permissive reasoners retain a theory after seeing disconfirming evidence, or when they revise the theory to accommodate the evidence, they display a constructive form of hindsight bias.

From a fallibilist perspective, strict reasoners are “good” and permissive reasoners are “bad.” By contrast, we propose that both groups are rational, but that their rationality operates over different time scales (short versus long, respectively) and towards different goals (hypotheses evaluation versus theory development, respectively). Our prediction, then, is that strict reasoners and permissive reasoners will not differ in overall cognitive ability, motivational orientation, or scientific reasoning, but rather in the predictive relation between these factors.

We measured basic cognitive abilities that have been proposed as central to intelligence and as driving cognitive development: processing speed (Coyle, Pillow, Snyder, & Kochunov, 2011; Fry & Hale, 1996), working memory (WM; Kyllonen & Christal, 1990), and selective attention (Richland & Burchinal, 2013; Rueda, Fan, McCandliss, Halparin, Gruber, Lercari, et al., 2004). We measured complex cognitive abilities using the Wisconsin Card Sort Task (WCST; Grant & Berg, 1948). This task requires evaluating logical hypotheses against evidence and, when they are disconfirmed, shifting to alternate hypotheses. Shifting is a component of executive function that is

positively associated with intelligence (Miyake & Friedman, 2012; Miyake, Friedman, Emerson, Witzki, Howerter, & Wager, 2000) and that improves over development (Huizinga, Dolan, & van der Molen, 2006). We predict that strict and permissive reasoners will be comparable in these basic and complex cognitive abilities. This prediction is surprising from the fallibilist position that instant rationality is normative, which expects strict reasoners to be “better” than permissive reasoners.¹

With respect to motivation, we focused on Dweck’s (2000; Dweck & Leggett, 1988) proposal that people possess implicit theories of intelligence that govern how they respond to failure. Entity theorists believe that intelligence is fixed, and that failures cannot be overcome. By contrast, incremental theorists believe that intelligence is malleable, and that failures serve as feedback in the learning process – and can therefore be overcome. We predict that incremental theorists will be better scientific reasoners than entity theorists, and that this should be true for both strict and permissive reasoners. This follows from our proposal that the groups do not differ on whether they adjust their beliefs based on disconfirming evidence. Rather, they differ on the time course of adjustment – instantly or after a delay.

Although we predict no differences between strict and permissive reasoners in overall cognition, motivation, and scientific reasoning, we do predict that group will modulate the relation between cognitive ability and motivational orientation on one hand and scientific reasoning on the other. Strict reasoners prefer instant rationality, and therefore faster processors will be better scientific reasoners, because it is the fluency with which hypotheses are evaluated against evidence – and potentially disconfirmed – that is paramount. By contrast, permissive reasoners prefer to delay rationality, and therefore we predict that better scientific reasoning will be associated with less impulsive and more deliberate thinking as indexed by slower processing speed and systematic consideration of – and shifting between – alternate hypotheses in the WCST.

We focus on children rather than adults for several reasons. The newly revised national science standards emphasize the need for K-12 students to engage in authentic science experimentation to develop scientific practices such as designing and carrying out valid investigations, analyzing and interpreting data, constructing explanations, and drawing evidence-based conclusions (NRC, 2012). Moreover, there is a correlation between hindsight bias, which we interpret positively as a form of permissive reasoning, and executive function, a basic cognitive ability, in preschool children (Bernstein, Atance, Meltzoff, & Loftus, 2007). We further focused on middle-school students because prior research has documented relationships between the measures relevant to our research questions in this age group. There is a correlation between scientific reasoning and WCST, a measure of complex cognitive ability, in middle- and high-school students (Kwon & Lawson, 2000). In addition, prior research has found a relation between achievement in mathematics, a subject closely related to science, and motivational orientation – possessing an incremental versus entity theory of implicit intelligence – in middle-school students (Blackwell, Trzesniewski, & Dweck, 2007).

2. Method

2.1 Participants

Participants were 109 7th grade students (age: $M = 13.37$ years, $SD = 0.29$; 55 males, 54 females) from a racially and ethnically diverse middle school in a Midwestern suburb.

2.2 Design

Each participant completed all measures. Participants were classified as either *strict* or *permissive* reasoners based on their performance on two measures of response to disconfirming evidence, as described below.

2.3 Measures

We collected measures of response to disconfirming evidence, simple and complex cognitive ability, motivational orientation, and scientific reasoning. All measures were designed or adapted to be administered in a group setting.

2.3.1 Prediction Evaluation Task

We designed two measures of permissive reasoning (i.e., hindsight bias) in scientific reasoning. The first adapted a measure from the social psychology and neuroscience literature (Batson, Thompson, Seufferling, Whitney, & Strongman, 1999; Greene & Paxton, 2009). On each trial, participants first predicted the outcome of a coin flip. The experimenter then flipped the coin and announced the outcome. Finally, participants evaluated their prediction by circling “correct” or “incorrect” on a response sheet. There were 20 trials. Participants could not see the actual outcome of each flip, which was done at the front of the class. Instead, they relied on the experimenter’s announcement of the outcome. In fact, the announced outcomes were predetermined so that equal numbers of heads and tails occurred over trials 1-10 and over trials 11-20. Participants who evaluated more than 15 predictions

correctly (< 1% chance) were defined as permissive reasoners.

2.3.2 Science Quiz

We designed a second measure of permissive reasoning in scientific reasoning.² Participants completed a science quiz of 10 multiple-choice questions; see Appendix A for example items. They were told these were college-level questions, and that they were unlikely to know any of the answers (“even your teacher would have difficulty with this quiz”), but that they should try their best. They were also told that the experimenter had made a mistake and had not printed the quiz, but rather the answer key to the quiz. Thus, the “answers” were at the bottom of the quiz, printed upside down and backward. Participants were told not to look at them. These “answers” were actually incorrect, and participants who gave six or more of them (< 1% chance) were defined as permissive reasoners.

2.3.3 Processing Speed

We measured processing speed using the “Coding B” scale of the WISC-IV®, adapted for group administration as specified by Varma, Varma, Van Boekel, and Wang (2016). Participants were given a response sheet. A key at the top mapped the digits 1-9 to arbitrary symbols. Several rows of digits appeared at the bottom; below each digit was an empty box. Participants had two minutes to translate as many digits to symbols as possible, working left-to-right and top-to-bottom. The dependent variable was the number of correctly translated symbols.

2.3.4 Working Memory

We measured WM using the “Forward Digit Span” scale of the WISC-IV®, adapted for group administration as specified by Varma et al. (2016). On each trial, the experimenter read aloud a sequence of digits at a rate of approximately one per second. Participants were then cued to reproduce the sequence from memory on their response sheet. There were two sequences of each of the lengths 2-9. The dependent variable was the total number of correctly recalled digits. A limitation of this measure is that it only taps the storage component of WM, not the processing component (Baddeley & Hitch, 1974; Just & Carpenter, 1992). It was chosen because it proved difficult in pilot testing to implement more conventional measures of WM such as complex span (Daneman & Carpenter, 1980; Turner & Engle, 1989) and backward digit span (from the WISC-IV®) in a group setting without widespread cheating.

2.3.5 Selective Attention

We measured selective attention using a paper-and-pencil version of the flanker task (Eriksen & Eriksen, 1974) adapted for group administration (Varma et al., 2016). Participants first completed a page of 32 neutral stimuli, each consisting of a central arrow pointing to the left or right, flanked on each side by two asterisks (e.g., “**←**”). “L” and “R” appeared below the leftmost and rightmost asterisk, respectively, and participants circled the one indicating the direction of the central arrow. After completing the page, participants consulted a stopwatch projected at the front of the classroom and recorded their completion time. They then completed a page of 32 interference stimuli, where the central arrow was flanked by arrows that either did (e.g., “←←←←←”) or did not (e.g., “→→←→→”) point in the same direction, and again recorded their completion time. The dependent variable was the interference completion time minus the neutral completion time.

2.3.6 Cognitive Flexibility

We measured cognitive flexibility using a version of the WCST shortened and adapted for group administration (Varma et al., 2016). 48 stimuli were projected at the front of the classroom. Each showed five cards, one target and four standards labeled A-D; see Appendix A for an example stimulus. Each card varied on four levels of each of three dimensions: color, number, and shape. Participants judged which of the four standard cards the target card “was most similar to” by circling one of A-D on their response sheet. Feedback was then provided by masking all but the correct standard card. The rule silently shifted every eight cards, cycling twice through same color, same number, and same shape. We coded the most common measure of WCST performance, number of perseverative errors, defined as errors immediately following a rule shift resulting from applying the disconfirmed rule. We also coded one novel measure, number of systematic errors, defined as errors immediately following a rule shift resulting from applying a logically possible rule (i.e., *not* the disconfirmed rule). Whereas perseverative errors are “bad,” systematic errors are “good,” indicating deliberate search of the hypothesis problem space (Klahr & Dunbar, 1988).

2.3.7 Implicit Theory of Intelligence

We measured implicit theory of intelligence by adapting the Dweck (2000) measure. We modified the three items to be specific to science and simplified the wording in consultation with the classroom teacher; see Appendix A. Higher scores indicate an incremental theory and lower scores an entity theory.

2.3.8 Self-Efficacy

To ensure that it is the implicit theory of intelligence that is important and not motivation more generally, we also measured self-efficacy. We adapted Dweck's (2000) measure of self-confidence in intelligence, modifying the items to be specific to science; see Appendix A. Higher scores indicate higher self-efficacy.

2.3.9 Scientific Reasoning

Existing measures of scientific reasoning in middle school children are based on Piagetian concepts (e.g., Lawson, 1978). We utilized a new measure of scientific reasoning (Varma et al., 2013) derived from an extensive review of the current psychological and educational literatures on scientific reasoning, and designed to be consistent with the National Science Standards (NRC, 2012). It spans five facets of scientific reasoning:

- 1) Hypothesis generation: the ability to observe a situation or event, recognize the difference between existing understanding and what more needs to be learned, and clearly articulate a question that can direct an empirical investigation.
- 2) Hypothesis testing: the ability to design valid tests of a hypothesis that correctly identify and manipulate all relevant variables in order to produce empirical evidence that will allow one to answer questions.
- 3) Reasoning from evidence: the ability to interpret the results of an investigation and to draw justified inferences and/or conclusions based upon the data.
- 4) Providing explanations: the ability to coordinate theory and evidence to draw inferences about causal or statistical relations.
- 5) Coordinating theory and evidence: the ability to evaluate a theory in light of experimental outcomes, reconcile new evidence with prior beliefs, and (if required) revise one's theory and generate new predictions.

The measure consists of ten multiple-choice and explanation items adapted from research on scientific reasoning (i.e., Tschirgi, 1980) and released standardized tests (i.e., TIMSS, 1995; 2003); see Appendix B for example items. Higher scores indicate better scientific reasoning.

2.4 Procedure

The scientific reasoning measure was administered by the classroom teacher during one class period; participants required approximately 20 minutes to complete it. All other measures were administered by experimenters during another class period, and required approximately 45 minutes to complete. Measures were administered in a fixed sequence designed (based on pilot testing) to maximize the interest of the participants. For example, they were more engaged by the "race against the clock" nature of the processing speed measure and less engaged by the self-efficacy measure. A drawback of the design decision to maximize engagement is that it may have introduced order effects.

3. Results

Because all measures were group administered, it was possible for disinterested participants to disengage. We therefore trimmed participants who made more than five errors on the digit-symbol substitution measure, more than 50% errors on the flanker, and fewer than five errors on the WCST.³ The trimmed sample consisted of 94 participants (age: $M = 13.36$ years, $SD = 0.28$; 49 females, 45 males). The 29 participants who demonstrated hindsight bias on the prediction evaluation task or science quiz formed the permissive reasoning group. The remaining 65 participants formed the strict reasoning group.⁴

3.1 Overall Differences Between Strict and Permissive Reasoners

We first evaluated whether the strict and permissive reasoning groups differed on cognitive ability, motivational orientation, or scientific reasoning, comparing them on each measure using independent *t*-tests (Table 1). The groups did not differ on any of the measures ($p > 0.30$ for all). This is inconsistent with fallibilism, which equates strict reasoning with normative reasoning, and therefore expects strict reasoners to have better scientific reasoning and cognitive ability scores than permissive reasoners. However, it is consistent with the proposed distinction between strict and permissive reasoning. In particular, we identified these groups using measures derived from the hindsight bias literature, and prior research has found no correlation between hindsight bias and intelligence as measured by Raven's Progressive Matrices (Pohl & Eisenhauer, 1995).

Table 1. Independent *t*-tests comparing the strict and permissive groups on all variables

Variable	Reasoning Group		<i>t</i> (92)	<i>p</i>
	Strict	Permissive		
Age	13.38 (0.28)	13.31 (0.29)	0.958 ^a	0.341
Processing Speed	59.08 (11.87)	60.45 (13.68)	-0.493	0.623
WM	58.45 (11.32)	59.34 (11.13)	-0.357	0.722
Flanker	9.48 (9.16)	7.55 (6.32)	1.026	0.308
WCST Perseveration Errors	0.71 (0.53)	0.86 (0.63)	-0.326	0.745
WCST Systematic Errors	0.20 (0.78)	0.24 (0.91)	-0.836	0.405
Implicit Intelligence	10.95 (4.00)	11.28 (3.65)	-0.369	0.713
Self-Efficacy	8.35 (2.92)	8.72 (2.64)	-0.584	0.561
Science Reasoning (<i>z</i>)	-0.015 (1.02)	0.017 (0.92)	-0.147	0.883

Note: All measures reported *M* (*SD*).

^a Two participants failed to report their age. Therefore, the degrees of freedom on this test was 90.

3.2 Relational Differences Between Strict and Permissive Reasoners

We next evaluated the prediction that strict and permissive reasoners differ in the relation between cognitive ability and motivational orientation on one hand and scientific reasoning on the other. For each group, we regressed scientific reasoning on the cognitive and motivational variables using a hierarchical approach to first control for (1) age, and then to evaluate the additional predictive power provided by (2) basic cognitive ability, (3) complex cognitive ability, and (4) motivational orientation. The results are shown in Table 2. For strict reasoners, the basic cognitive variables and motivational orientation variables each explained significant additional variance, and the final, full model accounted for 37.5% of the variance in scientific reasoning. For permissive reasoners, age and the motivational orientation variables each explained significant additional variance, and the final, full model accounted for 58.0% of the variance. To put these fits in context, only one previous educational psychology study has used cognitive measures to predict scientific reasoning ability in middle (and high) school students. Kwon and Lawson (2000) found that age, WCST, Tower of London, Group Embedded Figures, and “mental capacity” accounted for 56.1% of the variance on a measure of logico-scientific reasoning (Lawson, 1978).

Table 2. Hierarchical multiple regressions predicting scientific reasoning for the strict and permissive groups.

Step	Variable(s)	Reasoning Group					
		Strict			Permissive		
		ΔR^2	R^2	<i>p</i>	ΔR^2	R^2	<i>p</i>
1	Age	2.4%	2.4%	.226	20.5%	20.5%	.014
2	Basic Cognitive	24.2%	26.6%	.001	4.6%	25.1%	.692
3	Complex Cognitive	1.8%	28.4%	.498	11.2%	36.3%	.169
4	Motivation	9.1%	37.5%	.026	21.7%	58.0%	.016

The final, full regression models for each group are shown in Table 3. For strict reasoners, better scientific reasoning was associated with faster processing speed. This is consistent with the theoretical proposal that strict reasoners prefer instant rationality. For this group, faster processing enables more efficient evaluation of hypotheses against evidence and dismissal of those that are disconfirmed. Better scientific reasoning was also associated with having a more incremental theory of intelligence, a finding we return to in the Discussion.

Table 3. Final (step 4) multiple regression models predicting scientific reasoning for the strict and permissive groups.

Variable	Reasoning Group					
	Strict			Permissive		
	B (z)	t	p	B (z)	t	p
(intercept)		-1.123	.266		-1.671	.110
Age	0.094	0.829	.411	0.239	1.198	.245
Processing Speed	0.371	2.728	.009	-0.523	-2.152	.044
WM	-0.178	-1.515	.136	0.604	3.114	.005
Flanker	-0.070	-0.593	.556	0.225	1.371	.186
WCST Perseveration Errors	0.060	0.519	.606	0.373	2.232	.037
WCST Systematic Errors	0.207	1.773	.082	0.496	2.911	.009
Implicit Intelligence	0.308	2.500	.015	0.715	2.848	.010
Self-Efficacy	-0.122	-0.859	.394	-0.250	-1.654	.114

For permissive reasoners, the results were more complex – and more interesting. Better scientific reasoning was associated with slower processing speed, higher WM capacity, and more systematic errors on the WCST. These findings are consistent with the theoretical proposal that permissive reasoners prefer delaying rationality. They are willing to retain hypotheses in the face of disconfirming evidence, at least in the short run. Higher WM capacity and more systematic search of the hypothesis problem space enable a more thorough evaluation of the disconfirmed hypothesis and consideration of alternative hypotheses when confronted with disconfirming evidence. For this group, faster processing might be a detriment, leading to more impulsive decision-making. Better scientific reasoning was also associated with more perseverative errors. Although this finding was not predicted *a priori*, it can be interpreted as a consequence of delaying rationality. Retaining a disconfirmed hypothesis is the wrong choice when the disconfirming evidence is in fact veridical, as is the case for perseverative errors in the WCST. However, it is the correct choice when the disconfirming evidence represents a Type I error, as is sometimes the case in scientific reasoning (Simmons, Nelson, & Simonsohn, 2011). Finally, better scientific reasoning was associated with having a more incremental theory of intelligence, a finding we return to in the Discussion.

4. Discussion

In this study we proposed the existence of two groups, strict reasoners who prefer instant rationality and permissive reasoners who prefer delayed rationality. We identified the permissive reasoners behaviorally by their willingness to engage in hindsight bias on a prediction evaluation task or science quiz, which we reconceptualized as a willingness to retain or revise hypotheses given disconfirming evidence. Fallibilism expects permissive reasoners to be worse scientific reasoners than strict reasoners. In fact, Popper (1963) famously dismissed Freudianism and Marxism based on the permissive reasoning of their advocates – their willingness to retain these theories in the face of disconfirming evidence. Fallibilism also expects permissive reasoners to be of lower cognitive ability than strict reasoners. These expectations were not supported: the two groups did not differ on any of the scientific reasoning, cognitive ability, and motivational orientation measures.

By contrast, the theoretical proposal that strict reasoners prefer instant rationality and permissive reasoners prefer delayed rationality was supported by the finding that group membership modulated the relation between cognitive and motivational factors on one hand and scientific reasoning on the other. That strict reasoners prefer instant rationality was supported by the finding that better scientific reasoning was associated with faster processing speed. For this group, faster processing enables more fluent evaluation of hypotheses against evidence, freeing up limited cognitive resources for higher-order scientific inferencing.

That permissive reasoners prefer delayed rationality was supported by the finding that better scientific reasoning was associated with slower processing, greater WM capacity, and making more systematic errors on the WCST. We interpret this pattern as evidence of a speed-accuracy trade-off. More deliberate permissive reasoners cautiously retain and carefully revise hypotheses in the face of disconfirming evidence, whereas their more impulsive counterparts breeze by it without fully considering its implications, resulting in worse scientific

reasoning. This is consistent with the recent demonstration that slower data presentation rates enable people to defer hypothesis evaluation (Lange, Thomas, Buttaccio, Illingworth, & Davelaar, 2013).

That better scientific reasoning was associated with making more perseverative errors on the WCST was surprising given that such errors are associated with lesions to prefrontal cortex (Milner, 1963), with lower executive function and psychometric intelligence (Miyake et al., 2000), and with worse logico-scientific reasoning in middle and high school children (Kwon & Lawson, 2000). We offered a *post hoc* interpretation of this finding above, which we articulate further here. Scientific reasoning is *probabilistic*. Disconfirming evidence usually indicates that a hypothesis is incorrect – but it can also represent a Type I error. In the latter case, retaining a disconfirmed hypothesis in the short-term (i.e., perseverating) is adaptive, a form of replication indicative of deliberate thinking (Simmons et al., 2011). By contrast, the WCST is *deterministic*. Perseveration is never adaptive; a logical error is, and will continue to be, a logical error. Thus, making perseverative errors can be associated with better performance on more probabilistic scientific reasoning tasks, even if it is associated with worse performance on more deterministic logical tasks.

For both strict and permissive reasoners, better scientific reasoning was associated with having a more incremental implicit theory of intelligence. This finding does not follow specifically from our definitions of strict and permissive reasoners. Rather, we interpret it as an example of the more general finding in the motivation literature that people with incremental theories respond more adaptively to failure than people with fixed theories, and that is associated with better performance in domains such as mathematics (Blackwell et al., 2007). The current study extends this finding to the domain of science.

5. Limitations and Future Directions

The current study suffers from a number of limitations that should be addressed in future research. The size of the permissive reasoning group was (29 participants) was rather small given the number of independent variables in the regression analyses, and it is therefore important to replicate this study with a larger sample.

One question for future research is the role of inhibition in scientific reasoning. Prior research has shown that when correct scientific theories are learned, they do not supplant incorrect naïve beliefs, but rather suppress them during scientific reasoning, especially under speeded conditions (Goldberg & Thompson-Schill, 2009; Kelemen & Rosset, 2009; Knobe & Samuels, 2013; Shtulman & Valcarcel, 2012). This suggests a central role for inhibition when evaluating theoretical assumptions against disconfirming evidence. Given that we identified strict versus permissive reasoners using measures derived from the hindsight bias literature, and given the association between hindsight bias and inhibition (Bernstein et al., 2007), it is possible that strict reasoners and permissive reasoners use inhibition differently. Strict reasoners might use inhibition to suppress a disconfirmed hypothesis in the presence of disconfirming evidence, to be able to consider alternate hypotheses. By contrast, permissive reasoners might use inhibition to suppress disconfirming evidence so that a disconfirmed hypothesis can be maintained in the presence of alternate hypotheses. Future research should evaluate these predictions, and the differential roles of inhibition for strict versus permissive reasoners more generally.

This study focused on individual differences in the *processes* involved in scientific reasoning (e.g., Klahr & Dunbar, 1988). Accordingly, our measures were knowledge-lean. Further research is needed to extend our findings to knowledge-rich contexts, where participants engage in scientific reasoning around scientific *concepts*. Relevant here are cognitive development studies of the relationship between conceptual knowledge and patterns of reasoning. For example, Howe and colleagues have investigated how elementary and middle school age children reason about physics concepts (e.g., Howe, Taylor Tavares, & Devine, 2014). Their work distinguishes between deliberate engagement with explicit concepts versus less reflective engagement utilizing tacit knowledge.

Another question for future research is whether strict versus permissive reasoners differ on non-cognitive dimensions. We identified these two groups using measures derived from the hindsight bias literature. Some theories of hindsight bias stress the role of non-cognitive factors such as motivated sense-making (Pezzo & Pezzo, 2007) and metacognitive surprise (Müller & Stahlberg, 2007). Future research should investigate the relative contributions of cognitive and non-cognitive factors to strict versus permissive reasoning.

The current work has implications for the development of scientific reasoning. Better scientific reasoning was associated with cognitive abilities that improve over development: processing speed for strict reasoners, and WM and WCST performance for permissive reasoners. The question of whether these associations change over development could be addressed in a cross-sectional or longitudinal study. The question of whether a child's membership in either the strict or permissive group is stable over development or whether it fluctuates could be addressed in a longitudinal study.

The current work also has implications for the developmental neuroscience of scientific reasoning and cognitive abilities more generally (Kwon & Lawson, 2000). Neuroimaging studies of children and adults have established prefrontal cortex as a neural correlate of WM (Darki & Klingberg, 2015), inhibition (Rueda et al., 2004), and WCST performance (Konishi et al., 2008). Prefrontal cortex is also a neural correlate of reasoning about disconfirming evidence in logical (Goel & Dolan, 2003) and scientific (Fugelsang & Dunbar, 2005) contexts. Finally, prefrontal cortex is recruited as people deepen their understanding of physical systems (Mason & Just, 2015). One question, then, is whether strict vs. permissive reasoners rely on different brain areas when reasoning about disconfirming evidence, reflecting the employment of different cognitive processes as documented in the current study. For example, WM capacity was positively associated with scientific reasoning for permissive reasoners (but not strict reasoners). The prediction, then, is that permissive reasoners (but not strict reasoners) might show greater recruitment of lateral PFC areas associated with WM when reasoning about disconfirming evidence.

Addressing these questions will set the stage for future research on science education. One goal would be to evaluate whether group differences in strict versus permissive reasoning (and individual differences in cognitive ability and motivational orientation) predict science achievement as measured by standardized tests. Another goal is to design instruction that encourages children to reason more permissively about disconfirming evidence than is emphasized by the current science standards, where fallibilism is the norm (NRC, 2012). The result would be a shift from an emphasis on hypothesis evaluation to one on theory development (Chinn & Brewer, 1993).

6. Conclusion

The most important contribution of the current research is the proposal that people who reason permissively about disconfirming evidence are different from – not worse than – people who reason strictly. We identified permissive reasoners in a novel way, by recasting hindsight bias as a behavioral indicator of comfort with contradiction rather than a deviation from rationality and normativity, as it is viewed from the fallibilist perspective on science (Popper, 1963) and in the reasoning and decision-making literatures (Fischhoff, 1975; Nickerson, 1998; Slovic & Fischhoff, 1977). Permissive reasoners sometimes use disconfirming evidence to revise existing theoretical assumptions to better approximate reality, and sometimes set aside such evidence pending future replication. Their delayed rationality, necessary for theory development, balances the instant rationality prized by experimentalists.

Acknowledgements

The authors thank the teachers and students who participated in this research.

Notes

Note 1. We say “expects” rather than “predicts” because fallibilism is a philosophical program, not a scientific theory.

Note 2. This measure was inspired by discussions with Stephanie Carlson.

Note 3. Making fewer than five errors on the WCST indicates cheating because participants should make errors on, at a minimum, the five trials immediately following rule shifts.

Note 4. Note that a participant had to behave permissively on only one of the two measures to be included in the permissive reasoning group. A reviewer raised the interesting question of whether the results differed for participants who behaved permissively on both measures versus participants who behaved permissively on only one measure. Unfortunately, the size of the permissive reasoning group was too small ($N = 29$) to support analyses of these subgroups.

References

- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation* (pp. 47-89). New York: Academic. [https://doi.org/10.1016/S0079-7421\(08\)60452-1](https://doi.org/10.1016/S0079-7421(08)60452-1)
- Batson, C. D., Thompson, E. R., Seufferling, G., Whitney, H., & Strongman, J. A. (1999). Moral hypocrisy: Appearing moral to oneself without being so. *Journal of Personality and Social Psychology*, *77*, 525-537. <https://doi.org/10.1037/0022-3514.77.3.525>
- Bernstein, D. M., Atance, C., Meltzoff, A. N., & Loftus, G. R. (2007). Hindsight bias and developing theories of mind. *Child Development*, *78*, 1374-1394. <https://doi.org/10.1111/j.1467-8624.2007.01071.x>
- Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development*, *78*, 246-263. <https://doi.org/10.1111/j.1467-8624.2007.00995.x>

- Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research*, *63*, 1-49. <https://doi.org/10.3102/00346543063001001>
- Coyle, T. R., Pillow, D. R., Snyder, A. C., & Kochunov, P. (2011). Processing speed mediates the development of general intelligence (g) in adolescence. *Psychological Science*, *22*, 1265-1269. <https://doi.org/10.1177/0956797611418243>
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450-466. [https://doi.org/10.1016/S0022-5371\(80\)90312-6](https://doi.org/10.1016/S0022-5371(80)90312-6)
- Darki, F., & Klingberg, T. (2015). The role of fronto-parietal and fronto-striatal networks in the development of working memory: A longitudinal study. *Cerebral Cortex*, *25*, 1587-1595. <https://doi.org/10.1093/cercor/bht352>
- Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality, and development*. Philadelphia, PA: Psychology Press.
- Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, *95*, 256-273. <https://doi.org/10.1037/0033-295X.95.2.256>
- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Attention, Perception, & Psychophysics*, *16*, 143-149. <https://doi.org/10.3758/BF03203267>
- Fischhoff, B. (1975). Hindsight \neq foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, *1*, 288-299. <https://doi.org/10.1037/0096-1523.1.3.288>
- Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science*, *7*, 237-241. <https://doi.org/10.1111/j.1467-9280.1996.tb00366.x>
- Fugelsang, J. A., & Dunbar, K. N. (2005). Brain-based mechanisms underlying complex causal thinking. *Neuropsychologia*, *43*, 1204-1213. <https://doi.org/10.1016/j.neuropsychologia.2004.10.012>
- Goel, V., & Dolan, R. J. (2003). Explaining modulation of reasoning by belief. *Cognition*, *87*, B11-B22. [https://doi.org/10.1016/S0010-0277\(02\)00185-3](https://doi.org/10.1016/S0010-0277(02)00185-3)
- Goldberg, R. F., & Thompson-Schill, S. L. (2009). Developmental "roots" in mature biological knowledge. *Psychological Science*, *20*, 480-487. <https://doi.org/10.1111/j.1467-9280.2009.02320.x>
- Grant, D. A., & Berg, E. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental Psychology*, *38*, 404-411. <https://doi.org/10.1037/h0059831>
- Greene, J. D., & Paxton, J. M. (2009). Patterns of neural activity associated with honest and dishonest moral decisions. *Proceedings of the National Academy of Sciences USA*, *106*, 12506-12511. <https://doi.org/10.1073/pnas.0900152106>
- Howe, C., Taylor Tavares, J., & Devine, A. (2014). Children's conceptions of physical events: Explicit and tacit understanding of horizontal motion. *British Journal of Developmental Psychology*, *32*(2), 141-162. <https://doi.org/10.1111/bjdp.12026>
- Hume, D. (1748). *Philosophical essays concerning human understanding*. London: A. Millar.
- Huizinga, M., Dolan, C. V., & van der Molen (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia*, *44*, 2017-2036. <https://doi.org/10.1016/j.neuropsychologia.2006.01.010>
- Ioannidis, J. P. (2005). Why most published research findings are false. *Public Library of Science Medicine*, *2*, e124. <https://doi.org/10.1371/journal.pmed.0020124>
- Kelemen, D., & Rosset, E. (2009). The human function compunction: Teleological explanation in adults. *Cognition*, *11*, 138-143. <https://doi.org/10.1016/j.cognition.2009.01.001>
- Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, *12*, 1-48. https://doi.org/10.1207/s15516709cog1201_1
- Knobe, J., & Samuels, R. (2013). Thinking like a scientist: Innateness as a case study. *Cognition*, *126*, 72-86.

- <https://doi.org/10.1016/j.cognition.2012.09.003>
- Konishi, S., Morimoto, H., Jimura, K., Asari, T., Chikazoe, J., Yamashita, K., et al. (2008). Differential superior prefrontal activity on initial versus subsequent shifts in naive subjects. *NeuroImage*, *41*, 575-580. <https://doi.org/10.1016/j.neuroimage.2008.02.037>
- Kuhn, T. S. (1996). *The structure of scientific revolutions* (3rd ed.). Chicago: University of Chicago Press. <https://doi.org/10.7208/chicago/9780226458106.001.0001>
- Kwon, Y. J., & Lawson, A. E. (2000). Linking brain growth with the development of scientific reasoning ability and conceptual change during adolescence. *Journal of Research in Science Teaching*, *37*, 44-62. [https://doi.org/10.1002/\(SICI\)1098-2736\(200001\)37:1<44::AID-TEA4>3.0.CO;2-J](https://doi.org/10.1002/(SICI)1098-2736(200001)37:1<44::AID-TEA4>3.0.CO;2-J)
- Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?!. *Intelligence*, *14*, 389-433. [https://doi.org/10.1016/S0160-2896\(05\)80012-1](https://doi.org/10.1016/S0160-2896(05)80012-1)
- Lange, N. D., Thomas, R. P., Buttaccio, D. R., Illingworth, D. A., & Davelaar, E. J. (2013). Working memory dynamics bias the generation of beliefs: The influence of data presentation rate on hypothesis generation. *Psychonomic Bulletin & Review*, *20*, 171-176. <https://doi.org/10.3758/s13423-012-0316-9>
- Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91-196). Cambridge, UK: Cambridge University Press. <https://doi.org/10.1017/CBO9781139171434.009>
- Laudan, R. (1977). Ideas and organizations in British geology: A case study in institutional history. *Isis*, *68*, 527-538. <https://doi.org/10.1086/351872>
- Lawson, A. E. (1978). The development and validation of a classroom test of formal reasoning. *Journal of Research in Science Teaching*, *15*, 11-24. <https://doi.org/10.1002/tea.3660150103>
- Mason, R. A., & Just, M. A. (2015). Physics instruction induces changes in neural knowledge representation during successive stages of learning. *NeuroImage*, *111*, 36-48. <https://doi.org/10.1016/j.neuroimage.2014.12.086>
- Milner, B. (1963). Effects of different brain lesions on card sorting: The role of the frontal lobes. *Archives of Neurology*, *9*, 100-110. <https://doi.org/10.1001/archneur.1963.00460070100010>
- Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, *21*, 8-14. <https://doi.org/10.1177/0963721411429458>
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49-100. <https://doi.org/10.1006/cogp.1999.0734>
- Müller, P. A., & Stahlberg, D. (2007). The role of surprise in hindsight bias: A metacognitive model of reduced and reversed hindsight bias. *Social Cognition*, *25*, 165-184. <https://doi.org/10.1521/soco.2007.25.1.165>
- National Research Council (2012). Committee on Conceptual Framework for the New K-12 Science Education Standards. *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. National Academies Press.
- Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W. G. Chase (Ed.), *Visual information processing* (pp. 283-308). New York: Academic Press. <https://doi.org/10.1016/B978-0-12-170150-5.50012-3>
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*, 175-220. <https://doi.org/10.1037/1089-2680.2.2.175>
- Pezzo, M. V., & Pezzo, S. P. (2007). Making sense of failure: A motivated model of hindsight bias. *Social Cognition*, *25*, 147-164. <https://doi.org/10.1521/soco.2007.25.1.147>
- Pohl, R., & Eisenhauer, M. (1995). Hindsight bias in locating cities on a map. *Zeitschrift für Experimentelle Psychologie*, *42*, 63-93.
- Popper, K. R. (1963). *Conjectures and refutations: The growth of scientific knowledge*. Routledge & K. Paul. <https://doi.org/10.1063/1.3050617>
- Richland, L. E., & Burchinal, M. R. (2013). Early executive function predicts reasoning development.

- Psychological Science*, 24, 87–92. <https://doi.org/10.1177/0956797612450883>
- Rueda, M. R., Fan, J., McCandliss, B. D., Halparin, J. D., Gruber, D. B., Lercari, L. P., & Posner, M. I. (2004). Development of attentional networks in childhood. *Neuropsychologia*, 42, 1029-1040. <https://doi.org/10.1016/j.neuropsychologia.2003.12.012>
- Shtulman, A., & Valcarcel, J. (2012). Scientific knowledge suppresses but does not supplant earlier intuitions. *Cognition*, 124, 209-215. <https://doi.org/10.1016/j.cognition.2012.04.005>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22, 1359-1366. <https://doi.org/10.1177/0956797611417632>
- Slovic, P., & Fischhoff, B. (1977). On the psychology of experimental surprises. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 544–551. <https://doi.org/10.1037/0096-1523.3.4.544>
- Tschirgi, J. E. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child Development*, 51, 1-10. <https://doi.org/10.2307/1129583>
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, 28, 127-154. [https://doi.org/10.1016/0749-596X\(89\)90040-5](https://doi.org/10.1016/0749-596X(89)90040-5)
- Varma, K., Ross, P., Lawrenz, F., Roehrig, G., Huffman, D., McGuire, L., & Chen, Y. (2013). *Unpacking the elements of scientific reasoning*. Paper presented at the National Association for Research in Science Teaching, Puerto Rico.
- Varma, K., Varma, S., Van Boekel, M., & Wang, J. (2016). Studying individual differences in a middle school classroom context: Considering research design, student experience, and teacher knowledge. *Sage Research Methods Cases – Education*.

Appendix A

Example Items, Stimuli, and Measures

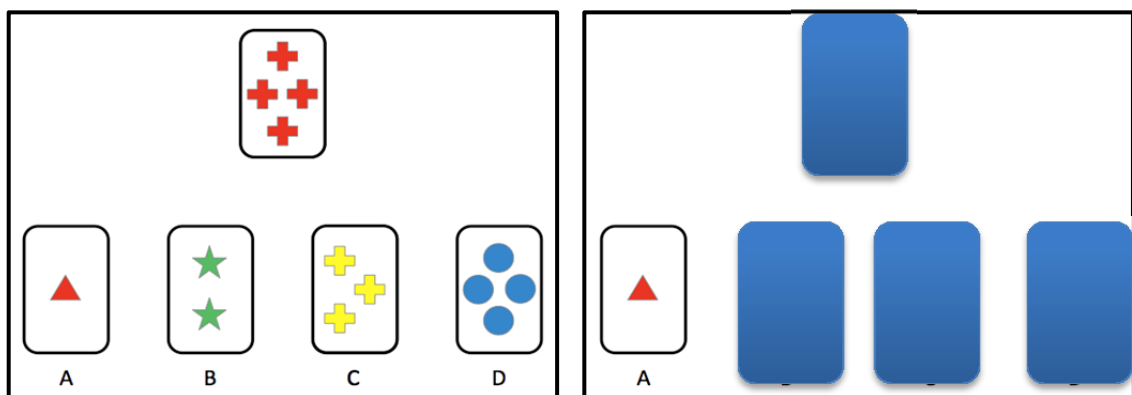
Hindsight Bias: Example items from Science Quiz

1. How many stars are in the Milky Way galaxy?
 - A. 30 million
 - B. 300 million
 - C. 3 billion
 - D. 300 billion

2. Monarch butterflies travel at an average of _____ miles per day during migration.
 - A. 10
 - B. 20
 - C. 30
 - D. 40

3. How many lenses are in a dragonfly's eye?
 - A. 30
 - B. 300
 - C. 3,000
 - D. 30,000

Cognitive Flexibility: Example stimulus from WCST



 Implicit Theory of Intelligence Measure

1. You have a certain amount of intelligence, and you really can't do much to change it.

1	2	3	4	5	6
Strongly Agree	Agree	Mostly Agree	Mostly Disagree	Disagree	Strongly Disagree

2. Your intelligence is something about you that you can't change very much.

1	2	3	4	5	6
Strongly Agree	Agree	Mostly Agree	Mostly Disagree	Disagree	Strongly Disagree

3. You can learn new things, but you can't really change your basic intelligence.

1	2	3	4	5	6
Strongly Agree	Agree	Mostly Agree	Mostly Disagree	Disagree	Strongly Disagree

 Self-Efficacy Measure

1. My aim is to completely master the material presented in this class.

1	2	3	4	5	6
Strongly Agree	Agree	Mostly Agree	Mostly Disagree	Disagree	Strongly Disagree

2. I am striving to do well compared to other students.

1	2	3	4	5	6
Strongly Agree	Agree	Mostly Agree	Mostly Disagree	Disagree	Strongly Disagree

3. My goal is to learn as much as possible.

1	2	3	4	5	6
Strongly Agree	Agree	Mostly Agree	Mostly Disagree	Disagree	Strongly Disagree

4. My aim is to perform well relative to other students.

1	2	3	4	5	6
Strongly Agree	Agree	Mostly Agree	Mostly Disagree	Disagree	Strongly Disagree

Appendix B

Example Items from the Scientific Reasoning Measure

Baking a Cake (Facet 2) (Note B1)

John decided to bake a cake. The cake recipe called for butter for the shortening, sugar for the sweetening, and white flour. John decided to make change the recipe.

- He used margarine instead of butter for the shortening
- He used honey instead of sugar for the sweetening, and
- He used brown whole wheat flour instead of regular white flour

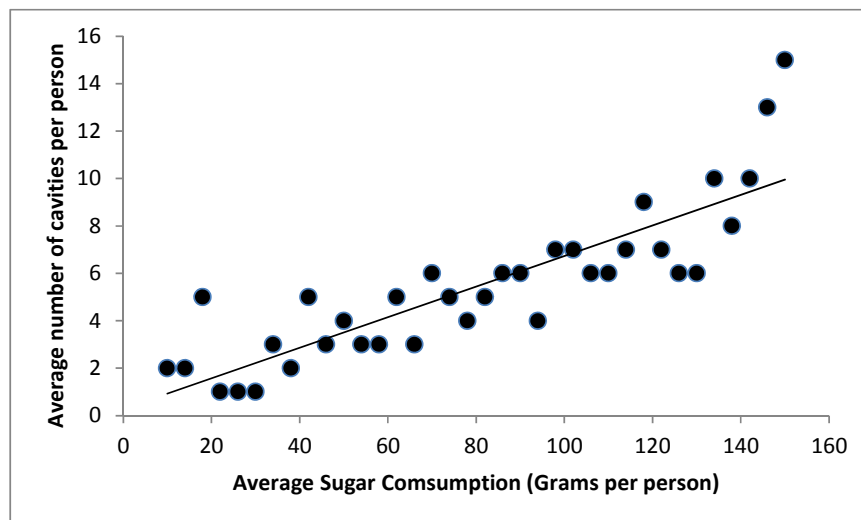
1. The cake turned our great. It was so moist. John thought that the reason the cake was so great was the honey. He thought that the type of shortening (butter or margarine) or the type of flour really didn't matter. What should he do to prove this point? (*Circle the best answer*)

- A. Bake the cake again but use sugar instead of honey, and still use the margarine and brown whole wheat flour.
- B. Bake the cake again but this time use sugar, butter, and regular white flour.
- C. Bake the cake again still using honey, but this time using butter and regular white flour.

2. Explain your selection

Cavities by Country (Facet 3) (Note B2)

The following graph shows the consumption of sugar and the number of cavities (decayed teeth) in different countries. Each country is represented by a dot in the graph.



1. Which one of the following statements is *best* supported by the data given in the graph? *Choose one answer.*

The “Baking a Cake” item was adopted from Tschirgi (1980) and addresses facet 2.

- A. In some countries, people brush their teeth more frequently than in other countries.
- B. The more sugar people eat, the more likely they are to get cavities.
- C. In recent years, the rate of cavities has increased in many countries.
- D. In recent years, the consumption of sugar has increased in many countries.

2. Explain your choice.

Note B1. The “Baking a Cake” item was adopted from Tschirgi (1980) and addresses facet 2.

Note B2. The “Cavities by Country” item was adopted from the Trends in International Mathematics and Science Study (1995; 2003).

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).