RESEARCH ARTICLE OPEN ACCESS

Measuring Rational Thinking in Adolescents: The Assessment of Rational Thinking for Youth (ART-Y)

Maggie E. Toplak[1] | Keith E. Stanovich[2]

[1]Department of Psychology, LaMarsh Centre for Child and Youth Research, York University, Toronto, Ontario, Canada | [2]Department of Applied Psychology and Human Development, University of Toronto, Toronto, Ontario, Canada

Correspondence: Maggie E. Toplak (mtoplak@yorku.ca)

Received: 5 September 2023 | **Revised:** 23 February 2024 | **Accepted:** 26 February 2024

Funding: This study was supported by Social Sciences and Humanities Research Council of Canada to Maggie Toplak (Grant File #: 435-2018-0874).

Keywords: adolescents and youth | assessment | rationality | real-world skills

ABSTRACT

There has been considerable conceptual and empirical progress on the measurement of rational thinking in adult samples. Studies in developmental samples have demonstrated that many of these domains and paradigms can also be assessed in children and youth, especially in adolescent samples. Here, we present an efficient rationality assessment battery for adolescents and youth—the Assessment of Rational Thinking for Youth (ART-Y). The ART-Y consists of five subtests: Probabilistic and Statistical Thinking, Scientific Thinking, Avoidance of Framing, Knowledge Calibration, and Rational Temporal Discounting. Two supplementary measures of thinking dispositions are included in the ART-Y: Actively Open-Minded Thinking (AOT) and Deliberative Thinking. The ART-Y battery was examined in a sample of 143 adolescents (mean age = 15.4 years). The five rational thinking subtests displayed intercorrelations largely consistent with those obtained in the adult literature. Age, cognitive ability, problem solving, probabilistic numeracy, and thinking dispositions predicted variance differently across the five subtests of the ART-Y, but again largely consistent with the adult literature. These measures, along with the ART-Y subtests, were examined as predictors of two real-world skills: financial literacy and academic achievement. Scientific thinking, knowledge calibration, and rational temporal discounting were significant unique predictors of financial literacy when statistically controlling for cognitive ability. Scientific thinking predicted academic achievement when statistically controlling for cognitive ability.

1 | Introduction

The influential heuristics and biases research program, inaugurated by Kahneman and Tversky in the early 1970s (Kahneman and Tversky 1972, 1973; Tversky and Kahneman 1974), has established that people's responses sometimes deviate from the axioms of rational choice on many reasoning tasks (Baron 2008, 2014; Evans 2014; Kahneman 2011; Stanovich 1999, 2011; Stanovich et al. 2016). These deviations vary from individual to individual, and this variance provides a means of measuring individual differences in rational thought. We utilized many of these tasks when constructing our Comprehensive Assessment of Rational Thinking (CART; Stanovich et al. 2016)—a very broad-based instrument for assessing rational thinking for adults. Here, drawing on our past work with the CART and developmental studies of younger subjects (Toplak 2022; Toplak et al. 2014a; Toplak and Flora 2021), we present a parallel rational assessment designed for youth called the Assessment of Rational Thinking for Youth (ART-Y). Rational thinking includes epistemic and instrumental rationality, referring respectively to the determination of *what is true* and *what to do* (Stanovich 2009; Stanovich et al. 2016). The subtests of the ART-Y tap epistemic issues such as the calibration of appropriate confidence in knowledge, and they tap instrumental issues of optimal decision making such as the framing of decision options and evaluating the utility of a delayed reward. Youth are increasingly exposed to complex and constantly changing environments where both components of rationality are important.

Although the ART-Y is derived from the CART, it is different in some ways. The biggest difference is in length. The CART was designed to be comprehensive, but that requirement results in a very long test for the purposes of research. For the ART-Y, we have chosen five subtests that measured disparate but absolutely critical aspects of the multifarious concept of rationality for a developmental sample. We carefully considered issues such as knowledge requirements, namely, knowledge that could be presumed to be acquired by youth, and considerations of stimulus equivalence (Stanovich et al. 2011; Toplak 2022). The five key subtests on the ART-Y are Probabilistic and Statistical Thinking, Scientific Thinking, Avoidance of Framing, Knowledge Calibration, and Rational Temporal Discounting. Like the CART, the ART-Y includes actively open-minded thinking and deliberative thinking as supplemental thinking dispositions measures. Each of the ART-Y subtests has a strong grounding in the literature on rational thinking in the adult literature (see Stanovich et al. 2016). The design of the ART-Y was based on research using developmentally suitable versions that have been studied in youth (see Toplak 2022, for a review).

The Probabilistic and Statistical Thinking and Scientific Thinking subtests of the CART are the most varied and content-packed of its subtests. They cover the most ground in terms of tapping different tasks from the heuristics and biases literature. They show the highest correlations with other subtests, and when recommending the shortest of the short-form versions of the CART, these are the two subtests that were chosen (Stanovich et al. 2016). Thus, they were the obvious places to start when constructing our adolescent version of a rational thinking battery. Our third choice for inclusion in the ART-Y—the Avoidance of Framing subtest—was driven by a generic, overarching fact about many measures of rational thinking: They assess whether the individual's choice remains the same when irrelevant contextual features are introduced (Stanovich 2013). Avoidance of framing effects was chosen for the ART-Y because it is one of the most reliable subtests tapping this property and has been used in several studies of children and adolescents (Reyna and Ellis 1994; Toplak et al. 2014a). To assess an important aspect of epistemic rationality, we included in the ART-Y a knowledge calibration subtest. Knowledge calibration errors have been found in a variety of real-life domains such as financial decisions and medical forecasts (Biais et al. 2005; Groopman 2007; Hilton et al. 2011; Tetlock 2005). The fifth subtest on the ART-Y—the Rational Temporal Discounting subtest—was designed to tap aspects of weakness of the will that lead people to make bad choices. We employed a temporal discounting task because it is logistically simpler to use in a multiple-task battery such as the ART-Y.

In terms of individual differences in the adult literature, the nomological network spanned by these five subtests forms a rather wide continuum (Stanovich et al. 2016). For example, whereas the probabilistic reasoning and scientific reasoning subtests show correlations with cognitive ability and with actively open-minded thinking in the range of 0.45 to 0.55, these relationships are in the range of 0.20 to 0.35 for the framing and knowledge calibration subtests of the CART. Correlations are even lower with these two cognitive indicators for the rational temporal discounting subtest (0.05 to 0.15). Likewise, the probabilistic reasoning and scientific reasoning subtests show a 0.56 correlation with each other in the CART but only in the range of 0.30 with framing and knowledge calibration subtests, which have more modest correlations with the 18 other subtests in the CART. Finally, temporal discounting shows low correlations with the other CART subtests—all associations lower than 0.20. We will examine below whether the landscape of individual differences across the ART-Y tasks is similar in adolescents.

In addition to these five subtests, the ART-Y also included measures of two thinking dispositions related to rationality. Table 1 includes a summary of the ART-Y subtests and rational thinking dispositions. In the next several subsections, we will provide more detail about the rationale for each of the subtests and thinking dispositions.

1.1 | Probabilistic and Statistical Reasoning

Probabilistic and statistical reasoning is one of the most thoroughly investigated areas in the heuristics and biases literature, especially in adult samples (Baron 2014; Kahneman 2011; Pohl 2017). This is not surprising because probability assessment is central to the achievement of both epistemic and instrumental rationality (Baron 2008; Stanovich 2010). The expected utility of an action involves multiplying the probability of an outcome by its utility and summing across possible outcomes. Thus, determining the best action involves estimating the probabilities of various outcomes. These probabilities are not typically conscious calculations of course—they are beliefs about states of the world and the confidence that a person has in them. To calibrate one's probabilistic beliefs rationally, just a few principles must be followed (Dawes 1998).

The Probabilistic and Statistical Reasoning subtest taps whether young adults tend to violate several of these strictures. On this subtest, several items tap the tendency to resist the gambler's fallacy (Ayton and Fischer 2004; Barron and Leider 2010; Burns and Corpus 2004; Croson and Sundali 2005; Roney and Sansone 2015; Xu and Harvey 2014)—a fallacy that reflects misunderstanding of randomness. Other items measure the tendency to avoid the conjunction effect (Mellers et al. 2001; Tversky and Kahneman 1983). Several items assess awareness of the fact that larger samples provide more accurate estimates of a parameter value.

On the ART-Y, four items assessed sensitivity to base rates using a paradigm that has been used in several investigations of children's reasoning (Davidson 1995; De Neys and Feremans 2013; De Neys and Vanderputte 2011; Kokis et al. 2002; Toplak 2022; Toplak et al. 2014a; Toplak and Flora 2021). The problems were structured so that the participant had to make an inductive inference in a simulation of a real-life decision. The information relevant to the decision was conflicting and of two different types. One type of evidence was statistical: either probabilistic or aggregate base rate information that favors

TABLE 1 | The five subtests and two thinking disposition scales of the Assessment of Rational Thinking for Youth (ART-Y).

ART-Y subtest or thinking disposition	Description, composition, and scoring
Subtests	
Probabilistic and Statistical Reasoning	Probabilistic and statistical reasoning is one of the most thoroughly investigated areas in the heuristics and biases literature, because probability assessment is central to the achievement of both epistemic and instrumental rationality. Several principles of probability assessment are assessed across 17 items, distributed as follows: avoiding probability matching tendencies (4 items), avoiding the gambler's fallacy (3 items), conjunction problems (3 items), sensitivity to causal base rates (4 items), and sensitivity to sample size/law of large numbers (3 items). The three conjunction problems and one of the sample size problems each have two parts, so the maximum score on this subtest is 21, a higher score indicates better performance.
Scientific Reasoning	The Scientific Reasoning subtest assesses skills of evidence evaluation, hypothesis formation and theory testing. There are 13 items on the subtest, measuring falsification tendencies on the four-card selection task (4 items); knowledge of the logic of converging evidence (2 items); measuring the tendency to avoid drawing causal inferences from correlational evidence (2 items); the use of control group reasoning (3 items); and covariation detection (2 items). A higher score indicates better performance.
Avoidance of Framing	The Avoidance of Framing subtest measures adherence to the principle of descriptive invariance that the preference between prospects should not depend on the manner in which they are described. Seven pairs of items with alternative frames are presented and the sum of the absolute differences in the responses on the different frames is computed. This score is reflected so that higher scores represent avoidance of framing and adherence to descriptive invariance.
Knowledge Calibration	This subtest assesses how well youth calibrate their degree of certainty about the things they think they know. Participants answer 42 true/false questions about general knowledge and, for each item, provide a confidence judgment indicating their subjective probability that their answer is correct. Mean accuracy is subtracted from mean confidence for each participant. The absolute score is used, as both overconfidence and underconfidence are deviations from perfect calibration. The absolute deviation score is subtracted from 100 so that higher scores on the subtest would indicate better knowledge calibration.
Rational Temporal Discounting	This subtest assesses prudent attitudes toward the future by asking participants to choose between smaller immediate rewards or much larger delayed rewards. Selection of the larger delayed reward is scored as more optimal, yielding a higher score. There are 27 scored items and 8 foil items that are not scored.
Rational thinking dispositions	
Actively Open-Minded Thinking Scale	This questionnaire measure assesses thinking dispositions such as the desire to act for good reasons; tolerance for ambiguity combined with a willingness to postpone closure; and the seeking and processing of information that disconfirms one's beliefs. Participants respond to 12 statements on a six-point scale ranging from disagree strongly to agree strongly. Ratings from the 12 items are summed for a total score.
Deliberative Thinking Scale	This questionnaire measure was designed to capture persistence in thinking, valuing thought, and intellectual engagement. Participants respond to 12 statements on a six-point scale ranging from disagree strongly to agree strongly. Ratings from the 12 items are summed for a total score.

one of the dichotomous decisions. The other type of evidence was a concrete case or a personal experience that points in the opposite direction (see the classic Volvo vs. Saab item in Fong et al. 1986). The problems assessed whether the subject had a tendency to rely on large-sample information rather than personal testimony. This subtest also contained four items tapping the tendency to maximize in probabilistic prediction tasks rather than probability match (Stanovich et al. 2016; West and Stanovich 2003). Developmentally suitable versions of all these problems have been studied in youth samples (Chiesi et al. 2011; Klaczynski 2001; Primi et al. 2017; Toplak 2022; Toplak et al. 2014a; Toplak and Flora 2021).

1.2 | Scientific Reasoning

The Scientific Reasoning subtest assesses key areas from the heuristics and biases literature that are of immense importance and not included in the Probabilistic and Statistical Reasoning subtest (Evans 1989; Manktelow 2012; Pohl 2017). These include the evaluation of evidence, hypothesis formation, and theory testing.

People tend to seek to confirm theories rather than falsify them. Performance on Wason's (Wason 1966) four-card selection task is often interpreted as indicating this tendency. Specifically, participants are asked to evaluate the following rule: "If a card has a vowel on its letter side, then it has an even number on its number side. Select the cards that prove the rule either true or false," Although there are many alternative theories that explain task performance (Evans 2007; Johnson-Laird 1999, 2006; Klauer et al. 2007; Oaksford and Chater 1994, 2007; Stanovich 1999), one of the oldest theories is that people focus on confirming the rule. Although people might perform poorly for many different reasons, regardless of which of these descriptive theories explains the error, there is no question that a concern for falsifiability would rectify the error. This subtest contains several of the types of four-card problems that have been studied in youth samples (Girotto et al. 1989), including deontic and nondeontic items (see Stanovich and West 1998a).

The principle of converging evidence describes how research results are synthesized in science: No one experiment is definitive, but each helps us to rule out at least some alternative explanations and, thus, aids in the process of homing in on the truth. The use of a variety of different methods makes scientists more confident that their conclusions rest on a solid empirical foundation. Research is highly convergent when a series of experiments consistently supports a given theory while collectively eliminating the most important competing theory. In the CART, Stanovich et al. (2016) developed items that measured the tendency to understand such convergence; two of these types of items were included in the ART-Y. Finally, two items measured the ability to avoid inferring causation from correlational evidence, three items tapped the tendency to use control-group reasoning (Lehman et al. 1988), and two items assessed covariation detection. Developmentally suitable versions of these tasks have been studied in youth samples (Toplak 2022; Toplak et al. 2014a; Toplak and Flora 2021).

1.3 | Avoidance of Framing

Under the standard view of so-called "rational man" in economics and decision theory, it is traditionally assumed that people have stable, underlying preferences for each of the options presented in a decision situation (Dawes 1998; Fishburn 1981, 1999; Resnik 1987; Starmer 2000; Thaler 2015). That is, it is assumed that a person's preferences for the options available for choice are complete, well-ordered, and well-behaved in terms of the axioms of choice. All of the axioms of choice (independence of irrelevant alternatives, transitivity, independence, reduction of compound lotteries, etc.), in one way or another, ensure that decisions are not influenced by irrelevant context (Stanovich 2013). Well-behaved internal preferences have the implication that a person is a utility maximizer—the person acts optimally to get what the person most wants.

In contrast to this model of ideal rationality, more than four decades of research has shown people's choices—sometimes choices about very important things—can be altered by irrelevant changes in how the alternatives are presented to them (Dawes 1998; Kahneman and Tversky 2000; Lichtenstein and Slovic 1971, 2006; Shafir and Tversky 1995; Slovic 1995). This problem is illustrated when people violate one of the simplest strictures of normative rationality, the principle of *descriptive invariance*: "that the preference order between prospects should not depend on the manner in which they are described" (Kahneman and Tversky 1984, 343). Such a violation is illustrated in the well-known "Asian disease problem" invented by Tversky and Kahneman (1981) and its analogues that have been extensively studied (Levin et al. 1998; Levin et al. 2002). We adapted seven such problems for the ART-Y, including both risky choice and attribute framing problems. Avoidance of framing has been examined in samples of youth (Parker and Weller 2015; Reyna and Ellis 1994; Toplak 2022; Toplak et al. 2014a; Toplak and Flora 2021; Weller et al. 2012, 2015, 2021).

1.4 | Knowledge Calibration

In this subtest, we assessed how well youth calibrate their degree of certainty about the things they think they know. Psychologists have done numerous studies using the so-called knowledge calibration paradigm (Fischhoff et al. 1977; Griffin and Tversky 1992; Hilton et al. 2011; Yates et al. 1997). A large set of such judgments can be evaluated for their epistemic rationality because, collectively, the set must adhere to certain statistical criteria. For example, if the weather forecaster says that there is a 90% chance of rain tomorrow and it is sunny and hot, there may be nothing wrong with that particular judgment. It just happened to be unexpectedly sunny on that particular day. However, if you found out that on half of the days the weatherperson said that there was a 90% chance of rain and it did not rain, then you would be justified in seriously questioning the accuracy of weather reports from this outlet.

In the most popular method for the assessment of knowledge calibration, people answer two-choice true/false questions and, for each item, provide a confidence judgment indicating their subjective probability that their answer is correct. Epistemic rationality is optimized only when one-to-one calibration is achieved—that the set of items assigned a subjective probability of 0.70 should be answered correctly 70% of the time, for example. This is what is meant by good knowledge calibration: A person must know what they know but also what they do not know. The standard finding across a wide variety of knowledge calibration experiments has been one of overconfidence, that subjective probability estimates are consistently higher than the obtained percentage correct.

Overconfidence effects have been found in perceptual and motor domains as well as in knowledge calibration paradigms (Baranski and Petrusic 1994, 1995; Mamassian 2008; West and Stanovich 1997). They are not just laboratory phenomena but have been found in a variety of real-life domains such as the prediction of sports outcomes (Ronis and Yates 1987), prediction of one's own behavior or life outcomes (Hoch 1985; Vallone et al. 1990), and economic and medical forecasts (Åstebro et al. 2007; Braun and Yaniv 1992; Groopman 2007; Tetlock 2005). Overconfidence is manifest in the so-called planning fallacy (see Buehler et al. 2002)—the fact that we often underestimate the time it will take to complete projects in the future (e.g., to complete an honors thesis, to complete this year's tax forms, and to finish a construction project). Overconfidence in knowledge calibration has been related to outcome variables such as financial decisions (Biais et al. 2005; Camerer and Lovallo 1999; Hilton et al. 2011; Russo and Schoemaker 1992).

The Knowledge Calibration subtest of the ART-Y employs a typical two-choice knowledge calibration paradigm. Using a two-choice format, subjects answer general knowledge questions that have been adapted for young adults. They then indicate their degree of confidence on a percentage scale that we have piloted for young adult subjects. Knowledge calibration has been examined in other youth batteries of decision-making tasks (Parker and Weller 2015; Weller et al. 2012, 2015, 2021).

1.5 | Rational Temporal Discounting

A prudent attitude toward the future that shifts psychological focus from the “here-and-now” to consideration of future outcomes has long been central to conceptions of rationality and wisdom (Baltes and Smith 2008; Loewenstein et al. 2003; Staudinger et al. 2005; Sternberg 2003; Strathman et al. 1994). The inability to properly value immediate versus delayed rewards keeps many people from maximizing their goal fulfillment. The logic of many addictions, such as alcoholism, overeating, and credit card shopping, illustrate this point. From a long-term perspective, a person definitely prefers sobriety, dieting, and keeping credit-card debt low. However, when immediately confronted with a stimulus that challenges this preference—a drink, a dessert, an item on sale—the long-term preference is trumped by the short-term desire.

Longitudinal studies have reported that positive long-term cognitive, educational, and career outcomes can be predicted from an early willingness to delay rewards (Mischel et al. 2011; Prencipe et al. 2006; Steinberg et al. 2009). In addition, temporal discounting performance has been significantly associated with excessive gambling (Petry 2001; Petry and Casarella 1999), drug addiction and smoking (Ainslie 2001, 2005; Kirby and Petry 2004), financial behavior (Ashraf et al. 2004; Meier and Sprenger 2012), prudent food-stamp usage (Shapiro 2005), educational success (Kirby et al. 2005), and a variety of other behaviors (Chabris et al. 2008; Reimers et al. 2009). Bartels et al. (2023) reported the most extensive study to date of how well laboratory-derived estimates of time preference predicted self-reports of 36 important outcome behaviors.

Temporal discounting measures and delay of gratification tasks have both been used to measure this attitude toward future rewards. In these tasks, participants are asked to choose between smaller immediate rewards or substantially larger delayed rewards. The selection of the larger delayed rewards is typically scored as more optimal (Ainslie 1975; Kirby 1997). Ayduk and Mischel (2002), Mischel (2014), Mischel et al. (1989), Mischel and Ebbesen (1970), and Rodriguez et al. (1989) pioneered the study of the delay of gratification paradigm with children.

There is a large literature on the extent to which adults discount monetary amounts into the future (“would you prefer $34 now or $50 in 30 days?”, see Kirby 2009). Many different paradigms have been used to assess how people compare a smaller reward immediately to a larger reward in the future—and how much larger the future reward has to be in order to tip the preference (Frederick et al. 2002; Green and Myerson 2004; Loewenstein et al. 2003; McClure et al. 2004). Many of these paradigms yield a curve with an individual's normalized indifference points plotted against time. Parameters are then derived from these curves. We use a simpler scoring scheme in the ART-Y, one which simply operationalizes certain levels of discounting as clearly less than rational. Our index is highly correlated with all the other discounting parameters (area under the curve, *k*-value, and indifference point) in the literature (Basile and Toplak 2015; Myerson et al. 2014).

1.6 | The Dispositions of Rationality Measured on the ART-Y: Actively Open-Minded Thinking and Deliberative Thinking Scales

Thinking disposition measures are not themselves components of rationality (Stanovich et al. 2016). Instead, they provide clues as to which underlying mechanisms are involved when suboptimal thinking is taking place (Baron 1993, 2008; Stanovich 2011). The distinction between cognitive capacities and thinking dispositions is an old one in psychology (Ackerman 1996; Ackerman and Heggestad 1997; Cronbach 1949). Cognitive capacities refer to the types of abilities underlying performance on IQ tests. They index cognitive power. Thinking dispositions, in contrast, are better viewed as cognitive styles. Rational thinking dispositions are those that relate specifically to the adequacy of belief formation and decision making. There is substantial evidence that thinking dispositions can explain variance in components of rational thinking after the variance due to cognitive ability has been controlled in adult samples (Stanovich et al. 2016; Viator et al. 2020) and in developmental samples (Kokis et al. 2002; Toplak 2022; Toplak et al. 2014a).

The thinking dispositions scales contained in the ART-Y are supplemental to the five primary subtests. They are not primary measures of rationality themselves because optimal functioning does not result from maximizing cognitive styles. Instead, rationality, plotted against most thinking dispositions, is an inverted U-shaped function. One does not maximize rationality by maximizing the reflectivity/impulsivity dimension, for example, because a person doing so might get lost in interminable pondering and never make a decision. One does not maximize the dimension of belief flexibility either, because such a person might end up with a pathologically unstable personality. Reflectivity and belief flexibility are “good” cognitive styles only in the sense that most people are too low on both dimensions (Baron 2008). Most people would be more rational if they increased their degrees of reflectivity and belief flexibility. But this does not mean that either of these thinking dispositions should always be maximized.

Not all of the thinking dispositions studied by psychologists relate to rationality. We chose two different thinking disposition scales for the ART-Y that reflect relatively disparate domains of cognitive regulation: the Actively Open-Minded Thinking (AOT) scale and the Deliberative Thinking scale. These thinking dispositions were also included in the CART. The ART-Y versions of these scales have been adapted for youth (Kokis et al. 2002; Toplak 2022; Toplak et al. 2014a; Toplak and Flora 2021).

We have been investigating actively open-minded thinking skills for almost two decades now. We were inspired to operationalize this concept by the work of Baron (1985, 1988, 1993), and we have refined the scale and examined its correlates in several studies (Stanovich and Toplak 2019, 2023; Stanovich and West 1997, 1998b, 2007). In Stanovich and West (1997), we conceptualized the AOT scale as a thinking disposition encompassing the cultivation of reflectiveness rather than impulsivity; the desire to act for good reasons; the tolerance for ambiguity combined with a willingness to postpone closure; and the seeking and processing of information that disconfirms one's beliefs. The items on the initial version (Stanovich and West 1997) of our AOT scale tapped reasoning styles such as the disposition toward reflectivity using items like: "If I think longer about a problem I will be more likely to solve it" and "Intuition is the best guide in making decisions" (the latter reverse scored). Other items tapped willingness to consider evidence contradictory to beliefs (e.g., "People should always take into consideration evidence that goes against their beliefs") and the willingness to consider alternative opinions and explanations ("A person should always consider new possibilities"). Some items tapped the willingness to postpone closure ("There is nothing wrong with being undecided about many issues"). Philosophically, the original scale focused strongly on issues of epistemic self-regulation raised in philosophical discussions (Goldman 1986; Harman 1995; Nozick 1993; Samuelson and Church 2015). The scale was a marker for the avoidance of epistemological absolutism, the willingness to perspective-switch and the tendency to consider alternative opinions and evidence.

The scale has gone through a series of refinements (see the discussion of that history in Stanovich and Toplak 2023). After iterating several versions, we currently recommend a 13-item version of the AOT for adults (Stanovich and Toplak 2023). That scale stresses the willingness to revise opinions and to be reflective about alternative theories and evidence. The ART-Y version has similar properties and is also based on our previous work with youth samples (Kokis et al. 2002; Toplak 2022; Toplak et al. 2014a; Toplak and Flora 2021).

The Deliberative Thinking scale on the ART-Y was designed to capture the type of cognitive variance tapped by need for cognition and typical intellectual engagement measures (Cacioppo et al. 1996; Goff and Ackerman 1992). Pilot versions of our scale and/or the original need for cognition scale were used in some of our earlier studies. Although not as potent a predictor as the AOT scale, deliberative thinking sometimes displays associations with heuristics and biases tasks included on the CART (Stanovich et al. 2016). The version used on the ART-Y was adapted for use with young subjects, also based on our studies with youth samples (Kokis et al. 2002; Toplak 2022; Toplak et al. 2014a).

In the remainder of this paper, we present some data on the characteristics of the ART-Y in a sample of 13- to 18-year-old adolescents. Specifically, we examined intercorrelations and predictors of the five key subtests on the ART-Y (Probabilistic and Statistical Thinking, Scientific Thinking, Avoidance of Framing, Knowledge Calibration, and Rational Temporal Discounting). The predictors included chronological age, thinking dispositions (AOT and Deliberating Thinking scales), a problem-solving task, probabilistic numeracy, and two measures of cognitive ability. The ART-Y subtests and these variables were further examined as predictors on two real-world skill measures: financial literacy and academic achievement. Given that some studies have examined financial literacy and economic understanding in children (Bruyneel et al. 2021; Echelbarger et al. 2020), financial literacy was included as a real-world skill measure in this study. We have found that rational thinking performance (resistance to cognitive biases) predicted self-reported academic achievement in a longitudinal sample of youth (Toplak and Flora 2021), and the current study provided the opportunity to examine whether the ART-Y would predict academic achievement based on a standardized performance measure.

2 | Methods

2.1 | Participants

The participants were 143 students from a high school (all males) in a rural area near a metropolitan city in the United States. The mean age of the participants was 15.4 years (SD = 1.2 years; range 13–18 years of age). There were 42 participants in grade 9 (29.4%), 39 participants in grade 10 (27.3%), 42 participants in grade 11 (29.4%), and 20 participants in grade 12 (14.0%). The sample was composed of different ethnicities, including 65 participants who reported Black ethnicity (45.5%), 44 reported Latino-Hispanic (30.8%), 15 reported Asian (10.5%), 8 reported White (5.6%), 1 reported South Asian (0.7%), and 10 reported Other (7.0%). Participants were reimbursed $40 for the study. The study was approved by the REB at York University. Participants 16 years and older signed consent forms, and participants under 16 years of age had signed consent provided by the school and signed assent forms. Four participants from the original sample were eliminated because they only completed one third of the study before choosing to discontinue.

Participants reported on the educational attainment of their parents. Of the 143 mothers, 36 (25.2%) had professional degrees, 43 (30.1%) completed college or university, 38 (26.6%) had some college or university, 12 (8.4%) completed high school and 14 (9.7%) completed less than high school. Of the fathers, 28 (19.6%) had professional degrees, 40 (28.0%) completed college or university, 30 (21.0%) had some college or university, 22 (15.4%) completed high school, 16 (11.2%) completed less than high school and 7 (4.9%) did not report their parents educational status.

2.2 | Tasks and Variables

2.2.1 | The Five Rational Thinking Subtests of the ART-Y

All of the subtest items and scoring criteria for the ART-Y are described in Data S1. Many items from our previous developmental work were included on the ART-Y (Toplak 2022). Other items were adapted from the CART and pilot tested to ensure that they were developmentally suitable and all had appropriate reading levels. The Flesch Reading Ease Score was 87.8 (potential range from 1 to 100, with 100 being highest readability score), and the Flesch–Kincaid Grade Level was 6.7 for the subtests and thinking dispositions included in the ART-Y.

2.2.1.1 | Probabilistic and Statistical Reasoning Subtest. The Probabilistic and Statistical Reasoning subtest was composed of 17 items, many of which were adapted from parallel items in the adult CART (Stanovich et al. 2016). The 17 items were distributed as follows: Four items assessed the ability to avoid probability matching tendencies and instead choose a maximizing strategy; three items assessed the ability to avoid the gambler's fallacy; three items assessed the ability to properly assign probabilities to conjunctions; four items assessed sensitivity to causal base rates; and three items assessed sensitivity to sample size/law-of-large numbers considerations. The items were intermixed with other items in the battery of tasks. The three conjunction problems and one of the sample size problems each had two parts, so the maximum score on this subtest was 21. Actual scores on this subtest ranged from 4 to 21. The reliability of this subtest was 0.70 (Cronbach's alpha). The mean score on the subtest was 10.71 (SD = 3.62).

2.2.1.2 | Scientific Reasoning Subtest. The Scientific Reasoning subtest was composed of 13 items, many of which were adapted from parallel items in the adult CART (Stanovich et al. 2016). The 13 items were distributed as follows: Four items measured falsification tendencies in the four-card selection task (two deontic and two nondeontic); two items assessed knowledge of the logic of converging evidence; two items assessed the tendency to avoid drawing causal inferences from correlation evidence; three items measured the tendency to use control-group reasoning; and two items measured covariation detection in a 2 × 2 matrix. The items were intermixed with other items in the battery of tasks. The mean score on this subtest was 8.10 (SD = 2.31), and higher scores indicate better performance on the task. The reliability of the Scientific Reasoning subtest was 0.56 (Cronbach's alpha). This reliability is somewhat low but is related to the number of items on this subtest and similar to the CART version. The reliability among adults completing the 17-item CART version (with an additional three points for a 25-item measure of covariation detection) was 0.67 (Stanovich et al. 2016).

2.2.1.3 | Avoidance of Framing Subtest. This test consisted of seven pairs of items that presented the same prospect in alternative frames. The members of each item pair were separated by many other tasks in the test battery and were presented on separate days. The subtest employed five pairs of items employing attribute framing and two pairs employing risky choice framing (see Levin et al. 1998, for a discussion of framing types). These items were adapted from studies of child and adolescent samples (Kokis et al. 2002; Toplak 2022; Toplak et al. 2014a; Toplak and Flora 2021; Weller et al. 2012) as well as studies of adult samples (Bruine de Bruin et al. 2007; Stanovich et al. 2016).

The items of each pair had the same narrative except for one frame that was worded in terms of a gain and the other frame was worded in terms of a loss (in the case of a risky choice item pair). In the case of the attribute framing item pairs, one frame was worded as a positive outcome and the other frame was worded as a negative outcome. For example, one attribute framing pair had as a positive frame: "The company *Intermail.com* ships online orders in the mail, and 95% of their orders arrive on time. How would you rate this company's shipping speed?" and as a negative frame: "The company *Intermail.com* ships online orders in the mail, and 5% of their orders arrive late. How would you rate this company's shipping speed?". Participants rated the shipping company's performance on a 6-point rating scale ranging from 1 (*very good performance*) to 6 (*very poor performance*).

Each item pair is scored by subtracting the positive (gain) frame of the pair from the negative (loss) frame. However, because the eventual score is an absolute value, the direction of the subtraction actually does not matter. This is because a framing effect in either direction represents a violation of the principle of descriptive invariance (Stanovich et al. 2016). The absolute value of the difference score represents the score for each item pair. The sum of the seven absolute difference scores ranged from 0 to 21 and displayed a mean of 7.19 (SD = 4.00). The absolute difference score was subtracted from 100 so that higher scores on the subtest would indicate avoidance of framing on this task, yielding a mean on this subtest of 92.81 (SD = 4.00). The reliability of the framing subtest was 0.48 (Cronbach's alpha). This is somewhat low but probably expected given that there are only seven scores contributing and that each of the scores represents a difference score. The reliability among adults completing the 11-pair CART version was only 0.61 (Stanovich et al. 2016). Bruine de Bruin et al. (2007) obtained a Cronbach's alpha of 0.62 and a test–retest reliability of 0.58 among adults using a 13-pair measure.

2.2.1.4 | Knowledge Calibration Subtest. The Knowledge Calibration subtest employs the typical two-choice calibration paradigm used extensively for many years in the classic literature on knowledge calibration (Lichtenstein and Fischhoff 1977, 1980; Ronis and Yates 1987; Yates et al. 1989). The subtest consisted of 42 items that followed one practice item. The general knowledge questions were selected from Tauber et al. (2013). The items and instructions were adapted based on studies with youth samples (e.g., Weller et al. 2012) and adult items in the Decision-Making Competence for adults (DMC; Bruine de Bruin et al. 2007) and in the CART (Stanovich et al. 2016).

An example of a typical item (for which the correct answer is A) is: What is the longest river in South America?

A Amazon*

B Nile

After answering each question, subjects indicated their degree of confidence in their answer on the following scale:

1. 100% chance that I answered correctly
2. 90% chance that I answered correctly
3. 80% chance that I answered correctly
4. 70% chance that I answered correctly
5. 60% chance that I answered correctly
6. 50% chance that I answered correctly

The mean confidence rating for these 42 items was 81.8% (SD = 7.9), and the mean percentage correct was 69.2% (SD = 10.2), yielding a substantial difference score of 12.6 (SD = 10.4), significantly different from zero, $t(142) = 14.52$, $p < 0.001$. The positive sign of the mean score indicates that the sample as a whole displayed overconfidence, the standard finding with items of this type (Stanovich et al. 2016). An overconfidence bias was displayed by 127 (88.8%) of the 143 subjects completing this task. However, because underconfidence is a failure of rational knowledge calibration just as much as is overconfidence, the score on the subtest is the absolute deviation from perfect calibration. Thus, the absolute value of the difference scores was calculated, turning the negative difference scores into positive deviations. The mean absolute accuracy/confidence deviation in the sample was 13.6 (SD = 9.2), significantly different from zero, $t(142) = 17.69$, $p < 0.001$. The absolute deviation score was subtracted from 100 so that higher scores on the subtest would indicate better knowledge calibration on this task, yielding a mean knowledge calibration score on this subtest of 86.4 (SD = 9.2). The split-half reliability (odd-even; Spearman–Brown corrected) of the absolute deviation score was 0.55. This is similar to the 36 item version of the subtest on the CART, which also had a reliability of 0.55 (Stanovich et al. 2016).

2.2.1.5 | Rational Temporal Discounting Subtest. The Rational Temporal Discounting subtest had 35 items. Participants were asked to make a choice between receiving a smaller amount of money now or a larger amount of money at a later time. An example item involved choosing between "$5 now or $10 in 2 days." There were four delay periods which corresponded with four delayed amounts ($10 in 2 days [5 items]; $25 in 1 month [9 items]; $100 in 3 months [9 items]; and $2000 in 365 days [12 items]). The items were presented in a fixed order. Two delay period/amounts ($10 in 2 days and $100 in 3 months) were presented with the immediate reward amounts in ascending order and two delay period/amounts ($25 in 1 month and $2000 in 365 days) were presented with the immediate reward amounts in descending order. The immediate amounts in the $10 in 2 days ascending condition were $1, $2.50, $5, $7.50, and $9.99*. The immediate amounts in the $25 in 1 month descending condition were $24.99*, $24.50*, $15, $12.50, $10, $7.50, $5, $2.50, and $1. The immediate amounts in the $100 in 3 months ascending condition were $1, $2.50, $5, $10, $25, $50, $75, $97.50*, and $99*. The immediate amounts in the $2000 in 365 days month descending condition were $1980*, $1950*, $1900*, $1500, $1250, $1000, $750, $500, $250, $100, $50, and $20.

Of these 35 items, 27 items were scored and eight were filler items not scored (noted by an asterisk above). The filler items all represented choices between an immediate amount and delayed amounts representing yearly interest rates of 24% or less. The scored items employed yearly interest rates of at least 33% and often vastly more. The median yearly interest rate of the scored items was 2800% and 24 of the 27 scored items had yearly interest rates of over 100%.

There were four response options for each item: (1) Strongly prefer $1 now, (2) Slightly prefer $1 now, (3) Slightly prefer $10 in 2 days, and (4) Strongly prefer $10 in 2 days. Strongly preferring the delayed reward was scored as 2 points, a slight preference for the delayed reward was scored as 1 point, and a score of 0 was given for the immediate/now choices. A total score was calculated for the 27 items, with a potential range of 0 to 54 points. The mean score on this subtest was 38.94 (SD = 11.08) and the observed range was 10–54 points. The reliability of this subtest was 0.94 (Cronbach's alpha). A higher score indicates more prudent temporal discounting.

2.2.2 | The Two Rational Thinking Dispositions in the ART-Y

Participants completed a self-report questionnaire that intermixed items from different thinking dispositions measures. They were asked to rate their agreement with each question using the following 6-point scale: *Disagree Strongly* (1), *Disagree Moderately* (2), *Disagree Slightly* (3), *Agree Slightly* (4), *Agree Moderately* (5), and *Agree Strongly* (6).

2.2.2.1 | Actively Open-Minded Thinking (AOT) Scale. The 12 items used in this scale were adapted from our previous studies of AOT utilizing child and adolescent samples (Kokis et al. 2002; Toplak 2022; Toplak et al. 2014a; Toplak and Flora 2021) as well as studies utilizing adult versions of the scale (Stanovich et al. 2016; Stanovich and West 2007). The version used here contained none of the problematic "belief items" discussed by Stanovich and Toplak (2019). The mean score on this subtest was 53.5 (SD = 5.9), and the observed range was 38 to 69. A higher score indicated more actively open-minded thinking. Cronbach's alpha was 0.62.

2.2.2.2 | Deliberative Thinking Scale. This scale was designed to tap variance similar to that captured by need for cognition and typical intellectual engagement measures (Cacioppo et al. 1996; Goff and Ackerman 1992). The 12 items used in this scale were adapted from our previous studies of need for cognition and related concepts that utilized child and adolescent samples (Kokis et al. 2002; Toplak 2022; Toplak et al. 2014a) and adult versions of the scale (Stanovich et al. 2016; Stanovich and Toplak 2019). The mean score on this subtest was 47.8 (SD = 8.6), and the observed range was 26 to 72. Cronbach's alpha was 0.84.

2.2.3 | Additional tasks

2.2.3.1 | Impression Management. In order to assess socially desirable responding, a 7-item impression management scale (Paulhus 1991) was administered as part of the thinking dispositions questionnaire. Examples of items are "I sometimes tell lies if I have to" (reverse scored) and "I never cover up my mistakes." The Cronbach's alpha was 0.60. A higher score indicated more socially desirable responding. The impression management scale did not correlate significantly with any of the five subtests of the ART-Y (correlations ranging from −0.12 to 0.06). The impression management scale did not correlate significantly with the AOT scale (0.12) but did show a significant correlation with performance on the Deliberative Thinking scale (0.27, $p < 0.001$).

2.2.3.2 | Problem-Solving Task. In a previous study, Frederick (2005) introduced a task that became widely used because of its posited ability to tap the cognitive reflection needed to inhibit intuitive responses. The task became so popular that we constructed our own version and incorporated it into the CART. Our 11-item measure was termed the Reflection versus Intuition Subtest (Stanovich et al. 2016). Since the publication of the CART, such tasks have been reconceptualized, based on new research. Although such tasks have often been treated as purely thinking dispositions or cognitive style measures, this is an inappropriate characterization, given what is now known about their properties. Such measures are psychometrically complex. They carry substantial variance that overlaps with intelligence, and they are strongly correlated with measures of numeracy (see Attali and Bar-Hillel 2020; Liberali et al. 2012; Otero et al. 2022; Patel et al. 2019; Sinayev and Peters 2015; Stanovich et al. 2016; Toplak et al. 2011, 2014b). To the extent that they assess a cognitive style of miserly processing (see Stanovich 2018), such variance can only be isolated after partialing out intelligence and numeracy (Otero et al. 2022). Because the task is a complex mixture of cognitive abilities, skills, knowledge, and dispositions, we have given the task the theoretical neutral label "problem solving task" in order to signal that it is not just an indicator of cognitive style.

There was a total of 11 items on the task used in this study. These items were adapted based on studies with child and adolescent samples (Toplak 2022) and on the Reflection versus Intuition Subtest of the adult CART (Stanovich et al. 2016). The mean score on this subtest was 3.64 ($SD = 3.23$). The reliability of this task (Cronbach's alpha) was 0.79. A higher score indicates better performance on this task.

2.2.3.3 | Probabilistic Numeracy. Levels of numeracy have been linked to several components of rational thinking (Cokely et al. 2012; Cokely and Kelley 2009; Jasper et al. 2013; Klaczynski 2014; Liberali et al. 2012; Peters 2020; Peters et al. 2006) and have also been found to be related to actual consumer, economic, and health decisions (Banks et al. 2010; Banks and Oldfield 2007; Låg et al. 2014; Peters 2012, 2020; Reyna et al. 2009; Reyna and Brainerd 2007). Probabilistic numeracy skills have also been examined in developmental samples in studies of judgment and decision-making (Donati et al. 2014; Fishbein and Schnarch 1997; Girotto 2014; Girotto and Gonzalez 2008; Primi et al. 2017; Reyna and Brainerd 2007; Schlottman and Wilkening 2012).

The items on many scales focus on what might be called probabilistic numeracy or statistical numeracy as they contain many items related to percentages and probabilities. The items chosen for this task were items of this type. The items from the Probabilistic Numeracy subtest of the adult CART (Stanovich et al. 2016) were adapted for younger subjects. There was a total of 15 items on the task. The mean score was 7.70 ($SD = 2.96$), and the reliability (Cronbach's alpha) was 0.73. A higher score indicated better performance on the task.

2.2.4 | Real-World Skill Measures

2.2.4.1 | Financial Literacy. The financial literacy task was an assessment of knowledge of financial tools for managing money and was used as a real-world skill correlate in the current study. These items were adapted from the Financial Literacy subtest of the CART used with adults (Stanovich et al. 2016). This task was composed of eight multiple choice questions. Examples of areas covered were understanding the difference between simple and compound interest, implications of early versus delayed retirement savings, defining important financial concepts (inflation, budgets, and income tax), prudent use of credit cards, and the purpose of car insurance. The observed range of scores was 0 to 8. The mean was 4.35 (SD = 1.84), and Cronbach's alpha was 0.56. A higher score indicated better financial literacy.

2.2.4.2 | Academic Achievement. Academic achievement scores were not part of the current research battery but were provided by the school for use as a real-world skill correlate for the current study. RIT scores from the Measures of Academic Progress (MAP) Growth achievement battery (NWEA 2019) were available for Language ($N = 122$), Math ($N = 120$), and Reading ($N = 121$). The Reading, Language Usage, and Mathematics tests were administered by the school in the fall of 2019, prior to the current study. The reliabilities across the content areas and grades (K to 12) are reported to be above 0.90 (NWEA 2019). In the current sample, the observed range of scores was 203 to 250, and the mean was 229.7 (SD = 9.3) for Language; the observed range was 214 to 294 and mean was 250.3 (SD = 14.2) for Math; and the observed range was 203 to 262 and mean was 231.1 (SD = 10.6) for Reading. The RIT scores on the three content areas were highly correlated (r's = 0.514 to 0.707). Each RIT score was standardized, and the three standardized scores were summed to create a composite academic achievement variable. A higher score indicated better academic achievement.

2.2.5 | Cognitive Ability Measures

2.2.5.1 | Shipley-2. The Shipley-2 is a standardized measure of crystallized and fluid cognitive abilities (Shipley et al. 2009). The Vocabulary and the Block Patterns scales were selected to assess verbal and nonverbal cognitive abilities in this study. Each of these scales had a 10-min time limit for administration. Both scales were group administered. The reliability estimates for the Vocabulary Scale ranged from 0.82 to 0.89; for the Block Patterns Scale, the reliabilities ranged from 0.88 to 0.94; and for the composite of Vocabulary and Block Patterns, the reliabilities ranged from 0.90 to 0.92 based on a sample of 13- to 19-year-olds (Shipley et al. 2009). The mean raw score for the Vocabulary scale was 26.80 ($SD = 3.95$) and for Block Patterns was 15.98 ($SD = 4.66$). The raw scores from the two scales were combined by standardizing and summing to generate a composite score.

2.2.5.2 | Verbal-Numerical Reasoning. A verbal-numerical reasoning measure of fluid intelligence measure from the UK Biobank (see Lyall et al. 2016) was also used as to assess fluid intelligence. These 12 items had verbal descriptions of problems followed by 4 to 6 multiple choice response options. Five of these items required knowledge of vocabulary, antonyms, and analogies of verbal concepts, and seven of the items required numerical calculations and sequences. Lyall et al. (2016) administered these verbal-numerical reasoning questions under a 2-min time limit in an online sample of adults between 40 and 69 years of age. However, no time limit was used in the current study. In our adolescent sample, the observed range was 2 to 12, and the mean was 8.24 (SD = 2.16). Cronbach's alpha was 0.62.

A Cognitive Ability Composite was used for some of the analyses. This composite included the sum of the Shipley-2 and Verbal-Numerical Reasoning standardized scores, which had a correlation of $r = 0.51$, $p < 0.001$. A higher score indicated better cognitive abilities.

2.3 | Procedure

The research study was conducted in-person on two consecutive days each lasting 1.5 h. Participants were tested in groups supervised by examiners and educators in each room. The following tasks were administered on the first day: probabilistic and statistical thinking, scientific thinking, problem solving, framing (part 1), knowledge calibration, temporal discounting, probabilistic numeracy, and demographic questions. The following tasks were administered on the second day: Shipley Vocabulary and Block Patterns, framing (part 2), verbal numerical reasoning, thinking dispositions, and financial literacy. All participants received the tasks in the same order.

3 | Results

The correlations among the five subtests of the ART-Y are presented in Table 2. The Probabilistic and Statistical Reasoning subtest displayed a significant 0.43 correlation with the Scientific Reasoning subtest. This correlation is consistent with that obtained with adult samples, although it is slightly lower than is typical. In adult samples, the correlation is typically in the range of 0.45 to 0.55 (Stanovich et al. 2016). The Avoidance of Framing subtest displayed significant correlations of 0.30 and 0.36 with the Probabilistic Reasoning and Scientific Reasoning subtests. These correlations are in the range of those obtained with adult subjects (Stanovich et al. 2016). The Knowledge Calibration subtest was significantly correlated with the Scientific Reasoning subtest (0.27), but not with the Probabilistic Reasoning or Avoidance of Framing subtests. The former result is in line with adult work. The 0.16 correlation between the Knowledge Calibration and Avoidance of Framing subtests, although not quite significant, is actually close to the 0.19 correlation that is obtained with adults on the CART (Stanovich et al. 2016). However, the nonsignificant 0.11 correlation between the Knowledge Calibration and Probabilistic Reasoning subtests is inconsistent with adult work, where a significant correlation is usually found. Finally, the Rational Temporal Discounting subtest shows no significant correlations with any of the other four subtests. This is largely consistent with adult work on temporal discounting measures, which usually finds correlations between 0.10 and 0.19 with other subtests on the CART (Stanovich et al. 2016). The final four rows of Table 2 display the means, standard deviations, range of scores, and Cronbach's alpha for each ART-Y subtest.

Table 3 presents correlations indicating how well other variables and tasks included in the study predicted each of the five subtests of the ART-Y. In this adolescent sample, age was a very modest predictor of subtest performance. It displayed significant correlations with the Probabilistic Reasoning subtest (0.17) and the Rational Temporal Discounting subtest (0.20), but had nonsignificant correlations with the remaining three subtests. The next two lines in the table examine how well the two thinking dispositions measured in the ART-Y—the AOT and Deliberative Thinking scales—predict performance on the five subtests. The two thinking disposition measures displayed similar patterns, showing modest correlations in the 0.21 to 0.34 range with the Probabilistic Reasoning and Scientific Reasoning subtests, but nonsignificant correlations with the other three subtests. These results contrast somewhat with what is obtained in adult subjects. In the CART, the AOT scale is a much more potent predictor than deliberative thinking. Additionally, it correlates significantly with the avoidance of framing effects and knowledge calibration in adult subjects, a finding we did not replicate with this adolescent version of the AOT scale. AOT correlations with probabilistic and scientific reasoning in adult samples are in the range of 0.40 to 0.50, higher than we obtained in the present study. Overall, our adolescent version of the AOT was not as strong a predictor of rational thinking performance as it has proven to be in adult studies (Stanovich et al. 2016; Stanovich and Toplak 2023).

The next two lines in the table indicate that performance on our problem-solving task and our probabilistic numeracy task showed moderately strong correlations with probabilistic and scientific reasoning (in the 0.44 to 0.61 range) and smaller but largely significant correlations with the other three subtests (in the 0.15 to 0.30 range). A parallel pattern was obtained with the two measures of cognitive ability (Shipley-2 and verbal numerical reasoning). Focusing on the composite cognitive ability measure, we see that the correlations with probabilistic reasoning and scientific reasoning were in the 0.51 to 0.61 range, and the correlations with the other three subtests in the range of 0.17 to 0.35. The correlations between the financial literacy measure and ART-Y subtests ranged from 0.24 to 0.41, and the correlations with academic achievement ranged from 0.15 to 0.66. Correlations between all of the other variables are shown in Data S2.

3.1 | Stepwise Regressions Predicting ART-Y Subtests

In Table 4, we examine which variables can uniquely explain performance on each of the five rational thinking subtests of the ART-Y. In a stepwise regression, we employed as predictors the cognitive ability composite score, the problem-solving task,

TABLE 2 | Means (SDs), reliability, and correlations between the five subtests of the ART-Y ($N = 143$).

	1.	2.	3.	4.	5.
1. Probabilistic and Statistical Reasoning	—				
2. Scientific Reasoning	0.43	—			
3. Avoidance of Framing	0.30	0.36	—		
4. Knowledge Calibration	0.11	0.27	0.16	—	
5. Rational Temporal Discounting	0.10	0.11	−0.01	0.06	—
Mean (SD) and reliability of each ART-Y subtest					
Mean	10.71	8.10	92.81	86.40	38.94
SD	3.62	2.31	4.00	9.20	11.08
Range of scores	4–21	2–13	55.95–100	79–100	10–54
Reliability	0.70	0.56	0.48	0.55	0.94

Note: Correlations >0.16 are significant at the 0.05 level. Correlations >0.21 are significant at the 0.01 level. Correlations >0.27 are significant at the 0.001 level.

TABLE 3 | Predictors of performance on the five rational thinking subtests on the ART-Y ($N = 143$).

Variables	Probabilistic and Statistical Reasoning	Scientific Reasoning	Avoidance of Framing	Knowledge Calibration	Rational Temporal Discounting
Chronological Age	0.17	0.08	0.02	0.15	0.20
ART-Y thinking dispositions					
Actively Open-Minded Thinking Scale	0.34	0.23	0.07	0.03	0.16
Deliberative Thinking Scale	0.29	0.21	0.10	0.02	0.13
Other predictor variables					
Problem Solving	0.61	0.55	0.30	0.23	0.15
Probabilistic Numeracy	0.49	0.44	0.30	0.22	0.24
Shipley-2	0.41	0.47	0.25	0.18	0.26
Verbal-Numerical Reasoning	0.48	0.60	0.36	0.15	0.03
Cognitive Ability Composite	0.51	0.61	0.35	0.19	0.17
Real-world skill variables					
Financial Literacy	0.30	0.41	0.28	0.30	0.24
Academic Achievement	0.47	0.66	0.39	0.15	0.17

Note: Correlations >0.16 are significant at the 0.05 level. Correlations >0.21 are significant at the 0.01 level. Correlations >0.27 are significant at the 0.001 level.

TABLE 4 | Beta weight in the final equation for a stepwise regression predicting each of the five ART-Y subtests by employing as predictors cognitive ability, problem solving, probabilistic numeracy, and actively open-minded thinking ($N = 143$).

Predictors	Probabilistic and Statistical Reasoning	Scientific Reasoning	Avoidance of Framing	Knowledge Calibration	Rational Temporal Discounting
Cognitive Ability Composite	—	0.449	0.348	—	—
Problem Solving	0.473	0.251		0.234	—
Probabilistic Numeracy	0.169	—	—	—	0.243
Actively Open-Minded Thinking Scale	0.138	—	—	—	—

probabilistic numeracy, and the AOT scale. Age and deliberative thinking were consistently the weakest predictors and thus were not entered. When the criterion variable was the Probabilistic and Statistical Reasoning subtest, the problem-solving task entered first into the regression equation, followed by probabilistic numeracy, and finally by the AOT scale. As Table 4 indicates, in the final regression equation, the problem-solving task was the dominant predictor.

When the criterion variable was the Scientific Reasoning subtest, the cognitive ability composite entered first into the regression equation followed by the problem-solving task. No other variables entered after that, and as Table 4 indicates, in the final equation, the cognitive ability composite was the dominant predictor.

When the Avoidance of Framing subtest was the criterion variable, only the cognitive ability composite entered the regression equation. No other variable explained significant variance after the first step. Regarding the Knowledge Calibration subtest, only the problem-solving task entered the regression equation. Finally, with regard to the Rational Temporal Discounting subtest, only the practical numeracy measure entered the regression equation.

The variance explained across the five subtests was 0.419, 0.413, 0.121, 0.055, and 0.059, respectively. Consistent with work on adult samples, performance on the Probabilistic Reasoning and Scientific Reasoning subtests was much more predictable from other variables than was performance on the other three subtests, which were quite independent of well-established cognitive indicators.

3.2 | Stepwise Regressions Predicting Real-World Skill Measures

In Table 5, we examine how the ART-Y subtests and the other variables performed as predictors of performance on one of our real-world skill measures, financial literacy. In the first stepwise

TABLE 5 | Beta weights in the final equation for a stepwise regression predicting financial literacy ($N=143$).

Predictors	Financial literacy	Predictors	Financial literacy
Cognitive Ability Composite	—	Cognitive Ability Composite	0.242
Problem Solving	—		
Probabilistic Numeracy	0.306		
Probabilistic and Statistical Reasoning	—	Probabilistic and Statistical Reasoning	—
Scientific Reasoning	0.224	Scientific Reasoning	0.187
Avoidance of Framing	—	Avoidance of Framing	—
Knowledge Calibration	0.175	Knowledge Calibration	0.197
Rational Temporal Discounting	—	Rational Temporal Discounting	0.169
Actively Open-Minded Thinking Scale	—	Actively Open-Minded Thinking Scale	—

TABLE 6 | Beta weights in the final equation for a stepwise regression predicting academic achievement composite ($N=120$).

Predictors	Beta weight
Cognitive Ability Composite	0.356
Problem Solving	0.181
Probabilistic Numeracy	0.169
Probabilistic and Statistical Reasoning	—
Scientific Reasoning	0.259
Avoidance of Framing Effects	—
Knowledge Calibration	—
Rational Temporal Discounting	—
Actively Open-Minded Thinking Scale	—

analysis, all five subtests of the CART were entered as predictors along with the cognitive ability composite, problem-solving task, probabilistic numeracy, and the AOT scale. The first variable to enter the regression equation was the probabilistic numeracy measure, followed by the Scientific Reasoning subtest and then the Knowledge Calibration subtest. No other variables entered the equation at that point. Those three variables predicted financial literacy with a multiple R^2 of 0.280. The beta weights of those three variables in the final equation are presented in Table 5.

The second stepwise regression analysis, presented to the right, entered just the five ART-Y subtests along with cognitive ability and the AOT scale. The first variable to enter the regression equation was cognitive ability, followed by the Knowledge Calibration subtest, then by the Temporal Discounting subtest, and then the Scientific Reasoning subtest. No other variable entered after that point. Those four variables predicted financial literacy with a multiple R^2 of 0.279—virtually the same as in the previous analysis. The beta weights of the four variables in the final equation are presented in Table 5.

In Table 6, we examine how the ART-Y subtests and the other variables performed as predictors of performance on the academic achievement composite measure. In a stepwise regression analysis, all five subtests of the ART-Y were entered as predictors along with the cognitive ability composite, problem-solving task, probabilistic numeracy, and the AOT scale. The first variable to enter the regression equation was cognitive ability, followed by the Scientific Reasoning subtest. The problem-solving task entered third and then probabilistic numeracy. No other variables entered the equation at that point. Those four variables predicted academic achievement with the multiple R^2 of 0.645. The beta weights of those four variables in the final equation are presented in Table 6.

4 | Discussion

The ART-Y was derived from and adapted from our adult measure, the CART (Stanovich et al. 2016). The five subtests that comprise the ART-Y are shown here to display relationships similar to their counterparts in the CART. Most notably, there is a relatively strong connection between the Probabilistic Reasoning subtest and the Scientific Reasoning subtest that mirrors the association observed in the CART. The Avoidance of Framing subtest also has moderate correlations with probabilistic and scientific reasoning. Performance on the Knowledge Calibration subtest, and even moreso the Rational Temporal Discounting subtest, is relatively unconnected to performance on the other subtests. This is consistent with results from the CART, where the Rational Temporal Discounting subtest displayed very modest correlations with the other subtests (range of $r=0.10$ to 0.19, in a sample of 747 adults). However, the slightly higher correlations in the CART may be partly attributable to the number of items scored (the ART-Y had 27 scored items and the CART had 38 items). Age was a modest predictor of ART-Y performance, consistent with other studies of rational thinking tasks in adolescent samples (Toplak 2022).

As Tables 3 and 4 demonstrate, the Probabilistic Reasoning and Scientific Reasoning subtests are predicted moderately well by other variables, such as cognitive ability and the problem-solving task. On the other hand, performance on the Knowledge Calibration and Rational Temporal Discounting subtests are not associated or display very weak associations, with many of the other predictor variables.

The ART-Y showed a modest ability to predict performance on two real-world skill measures— financial literacy and the academic achievement composite (see Tables 5 and 6). In terms of zero order correlations, the Scientific Reasoning subtest displayed a correlation of 0.41 with the former and 0.66 with the latter. As independent predictors, both the Scientific Reasoning and Knowledge Calibration subtests explained some unique variance in financial literacy when probabilistic numeracy was partialed out. Once probabilistic numeracy was removed as a predictor, three variables from the ART-Y were retained as unique predictors after cognitive ability entered the regression equation: the Scientific Reasoning subtest, the Knowledge Calibration subtest, and the Rational Temporal Discounting subtest. These findings are consistent with other studies that have shown that overconfidence in knowledge calibration has been related to financial decisions (Biais et al. 2005). Rational thinking (using subtests from the CART) has also been shown to predict financial behavior in adults (Toplak et al. 2017). Young adults have been shown to demonstrate particular risk for certain financial behaviors, including managing credit card debt and checks, especially if they also display low decision-making skills (Parker, Bruine de Bruin and Fischhoff 2015). Finally, results with the academic achievement measure converge with our previous findings that have shown that a battery of rational thinking tasks can predict, longitudinally, self and parent-reported academic achievement (Toplak and Flora 2021). Overall, these findings contribute to a growing body of research connecting rational thinking performance to real-world outcomes (Drummond and Fischhoff 2017; Sternberg and Sternberg 2017; Toplak et al. 2017).

The ART-Y does have some overlap with the Decision-Making Competence index for youths (Y-DMC) developed by Parker and Fischhoff (Parker and Fischhoff 2005; Parker et al. 2018; see also, Bruine de Bruin et al. 2007 for an adult version, the A-DMC, and Weller et al. 2012 for a preadolescent version, the PA-DMC). Two of the seven subtests of the Y-DMC—resistance to overconfidence bias and resistance to framing effects—are measured in a very similar manner in the ART-Y. One component of the Y-DMC, consistency in risk perception is measured exclusively with conjunction effect problems. Conjunction effect problems are part of the ART-Y, as they are included in the Probabilistic Reasoning subtest. But on that subtest, the ART-Y taps many more probabilistic reasoning effects that are not tapped on the DMC (e.g., avoiding the gambler's fallacy, sample size problems, and base rate problems). Conversely, although resistance to sunk costs is measured on the DMC but not the ART-Y, many aspects of the scientific reasoning subtest of the ART-Y are not covered by the DMC (e.g., falsifiability, converging evidence, and control group reasoning). Three of the subscales on the Y-DMC, applying decision rules, path independence, and recognizing social norms, have no direct counterpart in the ART-Y (or in the CART). However, the social norms subtest (testing the accuracy with which people perceive the social norms of their peers) does involve the metacognitive awareness of declarative knowledge in a manner that could be viewed as a form of knowledge calibration. Finally, the Y-DMC does not measure thinking dispositions as supplementary information, as does the ART-Y.

The work with DMC batteries converges with that from the ART-Y in showing that these tasks have associations with real-world skill and outcome variables, including risk behaviors and a broad range of self-reported outcomes (Bruine de Bruin et al. 2007; Parker and Fischhoff 2005; Weller et al. 2012, 2015). As in the current study, these correlations often remained significant even after statistically controlling for cognitive ability and numeracy skills. These findings reinforce the importance of examining reasoning skills beyond those typically assessed on intelligence tests—specifically, reasoning skills of the type assessed on the DMC batteries and ART-Y.

The ART-Y provides a reasonably wide sampling of rational thinking domains to assess in adolescents and young adults. The Probabilistic Reasoning and Scientific Reasoning subtests assess many of the well-studied effects in the heuristics and biases literature that serve to define rational thinking under some conceptualizations (Stanovich 2004, 2011, 2012). The Avoidance of Framing subtest assesses a foundational aspect of instrumental rationality–descriptive invariance in choice situations. The Rational Temporal Discounting subtest taps aspects of rational goal prioritization over time that are often studied in a literature quite separate from heuristics and biases work. The Knowledge Calibration subtest directly assesses an aspect of epistemic rationality, whereas many of the other subtests are concerned with making decisions and instrumental rationality.

The ART-Y is a promising battery to assess rational thinking in youth. It is based on extensive theoretical and empirical studies, including the CART in adults (Stanovich et al. 2016) and developmental research (Toplak 2022). The reliabilities of the rational thinking subtests were largely similar to the CART for adults, except for the Avoidance of Framing subtest. The framing subtest had a reliability of 0.48 on the ART-Y and 0.64 on the CART, but the CART also had 11 pairs of items (compared with 7 pairs on the ART-Y). Given that the framing items each have two parts, the addition of extra items is not optimal for an assessment intended to be shorter and more efficient for youth. The AOT scale in the ART-Y had a reliability of 0.62, but the AOT scale on the CART had a reliability of 0.85. The CART version of the AOT scale had 30 items, relative to 12 items on the ART-Y. This may partly explain why the adolescent version of the AOT was not as strong a predictor of rational thinking performance as in adult studies (Stanovich et al. 2016; Stanovich and Toplak 2023).

We have refrained from constructing a composite score of the five subtests from the ART-Y because, based on experience with the CART, we are concerned that such a score would be misinterpreted by some as indicating that there is a psychometric g of rationality. Nevertheless, it would be unobjectionable if investigators calculated a composite as long as they did not interpret it as a latent g-factor. Any composite construct computed from the ART-Y should be viewed as a formative construct, rather than a reflective one (Bollen and Lennox 1991; Diamantopoulos and Winklhofer 2001; Jarvis et al. 2003; Kovacs and Conway 2016, 2019). As Bollen and Lennox (1991) discuss, in the case of formative measurement models, indicators are defining characteristics of constructs and changes in indicators cause changes in the construct. In the contrasting case of reflective models, indicators are manifestations of the construct and changes in an indicator do not cause changes in the construct. A formative measurement model is more appropriate for a multidimensional construct like rational thinking. Any global notion of rational thinking that is defined by a composite ART-Y score would have to be understood in the manner of a formative concept—where the causal direction is from indicator to construct rather than from construct to indicator. Rational thinking is not a unified core ability defined by many different interchangeable indicators. It is disparate set of skills across broad domains such as those tapped by the ART-Y.

The strengths of the ART-Y are that it is based on a conceptual and theoretical framework deriving from models of rational thinking originating in the adult cognitive science literature

(Stanovich 2009; Stanovich et al. 2016). It is informed by extensive research on the study of these constructs and paradigms in developmental samples (Toplak 2022; Toplak et al. 2014a; Toplak and Flora 2021). Nevertheless, this particular study of the instrument had several limitations. First, the sample size in the current study was modest. Secondly, the schools participating educated only males, leaving investigation of performance on the ART-Y in female samples for future investigation. Our outcome variables were determined by availability and convenience and certainly could be improved upon. For example, while the finding of a unique contribution of the ART-Y subtests to predicting financial literacy is an important result, using a measure of real-life financial decisions would have been preferable. We did not use such a measure given the limited experience and lack of financial independence of youth at this period of development. However, given the significance of managing finances as a life domain and the findings in the current study, further examination of financial literacy should be conducted, with special consideration given to measurement issues (Hung et al. 2009). Future studies should also consider variables such as SES and neighborhood disadvantage, given the growing literature in developmental samples that have shown these variables are significant predictors of financial literacy (Lusardi et al. 2010) and decision-making competence (Weller et al. 2021, 2024).

By providing the full ART-Y in the supporting materials, our intention is to stimulate the use of this tool in a variety of studies of adolescent reasoning and decision-making. All studies attempting to link real-world skills and outcomes to thinking variation in children would be aided, we argue, by examining the subskills of rational thinking measured on the ART-Y. The Probabilistic Reasoning and Scientific Reasoning subtests themselves are important enough to serve as target indicators of the efficacy of educational interventions.

Acknowledgements

This paper is dedicated to our dear friend and collaborator, Richard F. West, who passed away on June 5, 2023. We would like to thank the students who assisted with data collection, Andrew Chan and Won Kim, and the students who assisted with data cleaning and scoring, Christiane Canillo, Kalin Mo, Joshua Rubenstein, and Emillie Vu. We would also like to thank the Alliance for Decision Education (AfDE) for assistance with recruitment of this sample and the faculty and staff at the school where the data were collected. The research was supported by a grant from the Social Sciences and Humanities Research Council of Canada to Maggie Toplak (Grant File #: 435-2018-0874).

Data Availability Statement

The data that support the findings of this study are openly available in Open Science Framework at https://osf.io/zmkyq/.

References

Ackerman, P. L. 1996. "A Theory of Adult Development:Process, Personality, Interests, and Knowledge." *Intelligence* 22: 227–257. https://doi.org/10.1016/S0160-2896(96)90016-1.

Ackerman, P. L., and E. D. Heggestad. 1997. "Intelligence, Personality, and Interests: Evidence for Overlapping Traits." *Psychological Bulletin* 121: 219–245. https://doi.org/10.1037/0033-2909.121.2.219.

Ainslie, G. 1975. "Specious Reward: A Behavioral Theory of Impulsiveness and Impulse Control." *Psychological Bulletin* 82, no. 4: 463–496. https://doi.org/10.1037/h0076860.

Ainslie, G. 2001. *Breakdown of Will*. Cambridge: Cambridge University Press.

Ainslie, G. 2005. "Précis of Breakdown of Will." *Behavioral and Brain Sciences* 28, no. 5: 635–649. https://doi.org/10.1017/S0140525X05000117.

Ashraf, N., D. Karlan, and W. Yin. 2004. *Tying Odysseus to the Mast: Evidence From a Commitment Savings Project in the Philippines*. Cambridge, MA: Harvard University Mimeograph.

Åstebro, T., S. A. Jeffrey, and G. K. Adomdza. 2007. "Inventor Perseverance After Being Told to Quit: The Role of Cognitive Biases." *Journal of Behavioral Decision Making* 20: 253–272. https://doi.org/10.1002/bdm.554.

Attali, Y., and M. Bar-Hillel. 2020. "The False Allure of Fast Lures." *Judgment and Decision Making* 15, no. 1: 93–111. https://doi.org/10.1017/S1930297500006938.

Ayduk, O., and W. Mischel. 2002. "When Smart People Behave Stupidly: Reconciling Inconsistencies in Social–Emotional Intelligence." In *Why Smart People Can Be So Stupid*, edited by R. J. Sternberg, 86–105. New Haven, CT: Yale University Press.

Ayton, P., and I. Fischer. 2004. "The Hot Hand Fallacy and the Gambler's Fallacy: Two Faces of Subjective Randomness?" *Memory & Cognition* 32: 1369–1378. https://doi.org/10.3758/BF03206327.

Baltes, P. B., and J. Smith. 2008. "The Fascination of Wisdom: Its Nature, Ontogeny, and Function." *Perspectives on Psychological Science* 3: 56–64. https://doi.org/10.1111/j.1745-6916.2008.00062.x.

Banks, J., C. O'Dea, and Z. Oldfield. 2010. "Cognitive Function, Numeracy and Retirement Saving Trajectories." *The Economic Journal* 120: F381–F410. https://doi.org/10.1111/j.1468-0297.2010.02395.x.

Banks, J., and Z. Oldfield. 2007. "Understanding Pensions: Cognitive Function, Numerical Ability and Retirement Saving." *Fiscal Studies* 28: 143–170. https://doi.org/10.1111/j.1475-5890.2007.00052.x.

Baranski, J. V., and W. M. Petrusic. 1994. "The Calibration and Resolution of Confidence in Perceptual Judgments." *Perception & Psychophysics* 55: 412–428. https://doi.org/10.3758/BF03205299.

Baranski, J. V., and W. M. Petrusic. 1995. "On the Calibration of Knowledge and Perception." *Canadian Journal of Experimental Psychology* 49: 397–407. https://doi.org/10.1037/1196-1961.49.3.397.

Baron, J. 1985. *Rationality and Intelligence*. New York: Cambridge University Press.

Baron, J. 1988. *Thinking and Deciding*. First ed. New York: Cambridge University Press.

Baron, J. 1993. "Why Teach Thinking?--An Essay." *Applied Psychology. An International Review* 42: 191–214.

Baron, J. 2008. *Thinking and Deciding*. Fourth ed. New York: Cambridge University Press.

Baron, J. 2014. "Heuristics and Biases." In *The Oxford Handbook of Behavioral Economics and the Law*, edited by E. Zamir and D. Teichman, 3–27. Oxford: Oxford University Press.

Barron, G., and S. Leider. 2010. "The Role of Experience in the Gambler's Fallacy." *Journal of Behavioral Decision Making* 23, no. 1: 117–129. https://doi.org/10.1002/bdm.676.

Bartels, D. M., Y. Li, and S. Bharti. 2023. "How Well Do Laboratory-Derived Estimates of Time Preference Predict Real-World Behaviors? Comparisons to Four Benchmarks." *Journal of Experimental Psychology. General* 152: 2651–2665. No Pagination Specified-No Pagination Specified. https://doi.org/10.1037/xge0001380.

Basile, A. G., and M. E. Toplak. 2015. "Four Converging Measures of Temporal Discounting and Their Relationships with Intelligence, Executive Functions, Thinking Dispositions, and Behavioral Outcomes." *Frontiers in Psychology* 6: 13. https://doi.org/10.3389/fpsyg.2015.00728.

Biais, B., D. Hilton, S. Pouget, and K. Mazurier. 2005. "Judgmental Overconfidence, Self-Monitoring and Trading Performance in an Experimental Financial Market." *Review of Economic Studies* 72: 297–312.

Bollen, K., and R. Lennox. 1991. "Conventional Wisdom on Measurement: A Structural Equation Perspective." *Psychological Bulletin* 110: 305–314. https://doi.org/10.1037/0033-2909.110.2.305.

Braun, P. A., and I. Yaniv. 1992. "A Case Study of Expert Judgment: Economists' Probabilities Versus Base-Rate Model Forecasts." *Journal of Behavioral Decision Making* 5: 217–231. https://doi.org/10.1002/bdm.3960050306.

Bruine de Bruin, W., A. M. Parker, and B. Fischhoff. 2007. "Individual Differences in Adult Decision-Making Competence." *Journal of Personality and Social Psychology* 92, no. 5: 938–956. https://doi.org/10.1037/0022-3514.92.5.938.

Bruyneel, S., L. Cherchye, S. Cosaert, B. De Rock, and S. Dewitte. 2021. "Verbal Aptitude Hurts Children's Economic Decision-Making Accuracy." *Journal of Behavioral Decision Making* 34, no. 2: 199–212. https://doi.org/10.1002/bdm.2204.

Buehler, R., D. Griffin, and M. Ross. 2002. "Inside the Planning Fallacy: The Causes and Consequences of Optimistic Time Predictions." In *Heuristics and Biases: The Psychology of Intuitive Judgment*, edited by T. Gilovich, D. Griffin, and D. Kahneman, 250–270. New York: Cambridge University Press.

Burns, B. D., and B. Corpus. 2004. "Randomness and Inductions from Streaks: "Gambler's Fallacy" Versus "Hot Hand"." *Psychonomic Bulletin & Review* 11: 179–184. https://doi.org/10.3758/BF03206480.

Cacioppo, J. T., R. E. Petty, J. A. Feinstein, and W. B. G. Jarvis. 1996. "Dispositional Differences in Cognitive Motivation: The Life and Times of Individuals Varying in Need for Cognition." *Psychological Bulletin* 119, no. 2: 197–253. https://doi.org/10.1037/0033-2909.119.2.197.

Camerer, C., and D. Lovallo. 1999. "Overconfidence and Excess Entry: An Experimental Approach." *American Economic Review* 89, no. 1: 306–318. https://doi.org/10.1257/aer.89.1.306.

Chabris, C. F., D. Laibson, C. L. Morris, J. P. Schuldt, and D. Taubinsky. 2008. "Individual Laboratory-Measured Discount Rates Predict Field Behavior." *Journal of Risk and Uncertainty* 37, no. 2–3: 237–269. https://doi.org/10.1007/s11166-008-9053-x.

Chiesi, F., C. Primi, and K. Morsanyi. 2011. "Developmental Changes in Probabilistic Reasoning: The Role of Cognitive Capacity, Instructions, Thinking Styles, and Relevant Knowledge." *Thinking & Reasoning* 17, no. 3: 315–350. https://doi.org/10.1080/13546783.2011.598401.

Cokely, E. T., M. Galesic, E. Schulz, S. Ghazal, and R. Garcia-Retamero. 2012. "Measuring Risk Literacy: The Berlin Numeracy Test." *Judgment and Decision Making* 7: 25–47. https://doi.org/10.1017/S1930297500001819.

Cokely, E. T., and C. M. Kelley. 2009. "Cognitive Abilities and Superior Decision Making Under Risk: A Protocol Analysis and Process Model Evaluation." *Judgment and Decision Making* 4: 20–33. https://doi.org/10.1017/S193029750000067X.

Cronbach, L. J. 1949. *Essentials of Psychological Testing*. New York: Harper.

Croson, R., and J. Sundali. 2005. "The Gambler's Fallacy and the Hot Hand: Empirical Data from Casinos." *Journal of Risk and Uncertainty* 30: 195–209. https://doi.org/10.1007/s11166-005-1153-2.

Davidson, D. 1995. "The Representativeness Heuristic and the Conjunction Fallacy Effect in Children's Decision Making." *Merrill-Palmer Quarterly* 41, no. 3: 328–346.

Dawes, R. M. 1998. "Behavioral Decision Making and Judgment." In *The Handbook of Social Psychology*, edited by D. T. Gilbert, S. T. Fiske, and G. Lindzey, vol. 1, 497–548. Boston: McGraw-Hill.

De Neys, W., and V. Feremans. 2013. "Development of Heuristic Bias Detection in Elementary School." *Developmental Psychology* 49, no. 2: 258–269. https://doi.org/10.1037/a0028320.

De Neys, W., and K. Vanderputte. 2011. "When Less Is Not Always More: Stereotype Knowledge and Reasoning Development." *Developmental Psychology* 47: 432–441. https://doi.org/10.1037/a0021313.

Diamantopoulos, A., and H. M. Winklhofer. 2001. "Index Construction With Formative Indicators: An Alternative to Scale Development." *Journal of Marketing Research* 38: 269–277. https://doi.org/10.1509/jmkr.38.2.269.18845.

Donati, M. A., A. Panno, F. Chiesi, and C. Primi. 2014. "A Mediation Model to Explain Decision Making Under Conditions of Risk among Adolescents: The Role of Fluid Intelligence and Probabilistic Reasoning." *Journal of Clinical and Experimental Neuropsychology* 36, no. 6: 588–595. https://doi.org/10.1080/13803395.2014.918091.

Drummond, C., and B. Fischhoff. 2017. "Development and Validation of the Scientific Reasoning Scale." *Journal of Behavioral Decision Making* 30, no. 1: 26–38. https://doi.org/10.1002/bdm.1906.

Echelbarger, M., K. Good, and A. Shaw. 2020. "Will She Give You Two Cookies for One Chocolate? Children's Intuition About Trades." *Judgment and Decision Making* 15, no. 6: 959–971. https://doi.org/10.1017/S1930297500008160.

Evans, J. S. B. T. 1989. *Bias in Human Reasoning: Causes and Consequences*. Hove, UK: Lawrence Erlbaum Associates.

Evans, J. S. B. T. 2007. *Hypothetical Thinking: Dual Processes in Reasoning and Judgement*. New York: Psychology Press.

Evans, J. S. B. T. 2014. *Reasoning, Rationality and Dual Processes*. London: Psychology Press.

Fischhoff, B., P. Slovic, and S. Lichtenstein. 1977. "Knowing With Certainty: The Appropriateness of Extreme Confidence." *Journal of Experimental Psychology. Human Perception and Performance* 3: 552–564.

Fishbein, E., and D. Schnarch. 1997. "The Evolution With Age of Probabilistic, Intuitively-Based Misconceptions." *Journal for Research in Mathematics Education* 28: 96–105.

Fishburn, P. C. 1981. "Subjective Expected Utility: A Review of Normative Theories." *Theory and Decision* 13: 139–199. https://doi.org/10.1007/BF00134215.

Fishburn, P. C. 1999. "The Making of Decision Theory." In *Decision Science and Technology: Reflections on the Contributions of Ward Edwards*, edited by J. Shanteau, B. A. Mellers, and D. A. Schum, 369–388. Boston: Kluwer Academic Publishers.

Fong, G. T., D. H. Krantz, and R. E. Nisbett. 1986. "The Effects of Statistical Training on Thinking About Everyday Problems." *Cognitive Psychology* 18, no. 3: 253–292. https://doi.org/10.1016/0010-0285(86)90001-0.

Frederick, S. 2005. "Cognitive Reflection and Decision Making." *The Journal of Economic Perspectives* 19: 25–42. https://doi.org/10.1257/089533005775196732.

Frederick, S., G. Loewenstein, and T. O'Donoghue. 2002. "Time Discounting and Time Preference: A Critical Review." *Journal of Economic Literature* XL: 351–401. https://doi.org/10.1257/jel.40.2.351.

Girotto, V. 2014. "Probabilistic Reasoning: Rational Expectations in Young Children and Infants." In *New Approaches in Reasoning Research*, edited by W. De Neys and M. Osman. New York: Psychology Press.

Girotto, V., M. Gilly, A. Blaye, and P. Light. 1989. "Children's Performance in the Selection Task: Plausibility and Familiarity." *British*

Journal of Psychology 80, no. 1: 79–95. https://doi.org/10.1111/j.2044-8295.1989.tb02304.x.

Girotto, V., and M. Gonzalez. 2008. "Children's Understanding of Posterior Probability." *Cognition* 106, no. 1: 325–344. https://doi.org/10.1016/j.cognition.2007.02.005.

Goff, M., and P. L. Ackerman. 1992. "Personality-Intelligence Relations: Assessment of Typical Intellectual Engagement." *Journal of Educational Psychology* 84: 537–552. https://doi.org/10.1037/0022-0663.84.4.537.

Goldman, A. I. 1986. *Epistemology and Cognition*. Cambridge, MA: Harvard University Press.

Green, L., and J. Myerson. 2004. "A Discounting Framework for Choice With Delayed and Probabilistic Rewards." *Psychological Bulletin* 130: 769–792. https://doi.org/10.1037/0033-2909.130.5.769.

Griffin, D., and A. Tversky. 1992. "The Weighing of Evidence and the Determinants of Confidence." *Cognitive Psychology* 24: 411–435. https://doi.org/10.1016/0010-0285(92)90013-R.

Groopman, J. 2007. *How Doctors Think*. Boston: Houghton Mifflin.

Harman, G. 1995. "Rationality." In *Thinking*, edited by E. E. Smith and D. N. Osherson, vol. 3, 175–211. Cambridge, MA: The MIT Press.

Hilton, D., I. Regner, L. Cabantous, L. Charalambides, and S. Vautier. 2011. "Do Positive Illusions Predict Overconfidence in Judgment? A Test Using Interval Production and Probability Evaluation Measures of Miscalibration." *Journal of Behavioral Decision Making* 24: 117–139. https://doi.org/10.1002/bdm.678.

Hoch, S. J. 1985. "Counterfactual Reasoning and Accuracy in Predicting Personal Events." *Journal of Experimental Psychology. Learning, Memory, and Cognition* 11: 719–731.

Hung, A., A. M. Parker, and J. Yoong 2009. "Defining and Measuring Financial Literacy." RAND Working Paper Series WR-708.

Jarvis, C. B., S. B. MacKenzie, and P. M. Podsakoff. 2003. "A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research." *Journal of Consumer Research* 30: 199–218. https://doi.org/10.1086/376806.

Jasper, J. D., C. Bhattacharya, I. P. Levin, L. Jones, and E. Bossard. 2013. "Numeracy as a Predictor of Adaptive Risky Decision Making." *Journal of Behavioral Decision Making* 26: 164–173. https://doi.org/10.1002/bdm.1748.

Johnson-Laird, P. N. 1999. "Deductive Reasoning." *Annual Review of Psychology* 50: 109–135. https://doi.org/10.1146/annurev.psych.50.1.109.

Johnson-Laird, P. N. 2006. *How We Reason*. Oxford: Oxford University Press.

Kahneman, D. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus & Giroux.

Kahneman, D., and A. Tversky. 1972. "Subjective Probability: A Judgment of Representativeness." *Cognitive Psychology* 3: 430–454. https://doi.org/10.1016/0010-0285(72)90016-3.

Kahneman, D., and A. Tversky. 1973. "On the Psychology of Prediction." *Psychological Review* 80: 237–251. https://doi.org/10.1037/h0034747.

Kahneman, D., and A. Tversky. 1984. "Choices, Values, and Frames." *American Psychologist* 39: 341–350. https://doi.org/10.1037/0003-066X.39.4.341.

Kahneman, D., and A. Tversky, eds. 2000. *Choices, Values, and Frames*. Cambridge: Cambridge University Press.

Kirby, K. 1997. "Bidding on the Future: Evidence Against Normative Discounting of Delayed Rewards." *Journal of Experimental Psychology. General* 126: 54–70.

Kirby, K. N. 2009. "One-Year Temporal Stability of Delay-Discount Rates." *Psychonomic Bulletin & Review* 16: 457–462. https://doi.org/10.3758/PBR.16.3.457.

Kirby, K. N., and N. Petry. 2004. "Heroin and Cocaine Abusers Have Higher Discount Rates for Delayed Rewards Than Alcoholics or Non-drug-Using Controls." *Addiction* 99: 461–471. https://doi.org/10.1111/j.1360-0443.2003.00669.x.

Kirby, K. N., G. C. Winston, and M. Santiesteban. 2005. "Impatience and Grades: Delay-Discount Rates Correlate Negatively With College GPA." *Learning and Individual Differences* 15: 213–222. https://doi.org/10.1016/j.lindif.2005.01.003.

Klaczynski, P. A. 2001. "Analytic and Heuristic Processing Influences on Adolescent Reasoning and Decision-Making." *Child Development* 72, no. 3: 844–861. https://doi.org/10.1111/1467-8624.00319.

Klaczynski, P. A. 2014. "Heuristics and Biases: Interactions Among Numeracy, Ability, and Reflectiveness Predict Normative Responding." *Frontiers in Psychology* 5: 1–13. https://doi.org/10.3389/fpsyg.2014.00665.

Klauer, K. C., C. Stahl, and E. Erdfelder. 2007. "The Abstract Selection Task: New Data and an Almost Comprehensive Model." *Journal of Experimental Psychology. Learning, Memory, and Cognition* 33: 688–703.

Kokis, J., R. Macpherson, M. Toplak, R. F. West, and K. E. Stanovich. 2002. "Heuristic and Analytic Processing: Age Trends and Associations With Cognitive Ability and Cognitive Styles." *Journal of Experimental Child Psychology* 83: 26–52. https://doi.org/10.1016/S0022-0965(02)00121-2.

Kovacs, K., and A. R. A. Conway. 2016. "Process Overlap Theory: A Unified Account of the General Factor of Intelligence." *Psychological Inquiry* 27, no. 3: 151–177. https://doi.org/10.1080/1047840x.2016.1153946.

Kovacs, K., and A. R. A. Conway. 2019. "What Is IQ? Life Beyond "General Intelligence"." *Current Directions in Psychological Science* 28, no. 2: 189–194. https://doi.org/10.1177/0963721419827275.

Låg, T., L. Bauger, M. Lindberg, and O. Friborg. 2014. "The Role of Numeracy and Intelligence in Health-Risk Estimation and Medical Data Interpretation." *Journal of Behavioral Decision Making* 27: 95–108. https://doi.org/10.1002/bdm.1788.

Lehman, D. R., R. O. Lempert, and R. E. Nisbett. 1988. "The Effects of Graduate Training on Reasoning: Formal Discipline and Thinking About Everyday-Life Events." *American Psychologist* 43, no. 6: 431–442. https://doi.org/10.1037/0003-066X.43.6.431.

Levin, I. P., G. J. Gaeth, J. Schreiber, and M. Lauriola. 2002. "A New Look at Framing Effects: Distribution of Effect Sizes, Individual Differences, and Independence of Types of Effects." *Organizational Behavior and Human Decision Processes* 88: 411–429. https://doi.org/10.1006/obhd.2001.2983.

Levin, I. P., S. L. Schneider, and G. J. Gaeth. 1998. "All Frames Are Not Created Equal: A Typology and Critical Analysis of Framing Effects." *Organizational Behavior and Human Decision Processes* 76: 149–188. https://doi.org/10.1006/obhd.1998.2804.

Liberali, J. M., V. F. Reyna, S. Furlan, L. M. Stein, and S. T. Pardo. 2012. "Individual Differences in Numeracy and Cognitive Reflection, With Implications for Biases and Fallacies in Probability Judgment." *Journal of Behavioral Decision Making* 25: 361–381. https://doi.org/10.1002/bdm.752.

Lichtenstein, S., and B. Fischhoff. 1977. "Do Those Who Know More Also Know More About How Much They Know?" *Organizational Behavior and Human Performance* 20: 159–183. https://doi.org/10.1016/0030-5073(77)90001-0.

Lichtenstein, S., and B. Fischhoff. 1980. "Training for Calibration." *Organizational Behavior and Human Performance* 26: 149–171. https://doi.org/10.1016/0030-5073(80)90052-5.

Lichtenstein, S., and P. Slovic. 1971. "Reversal of Preferences Between Bids and Choices in Gambling Decisions." *Journal of Experimental Psychology* 89: 46–55. https://doi.org/10.1037/h0031207.

Lichtenstein, S., and P. Slovic, eds. 2006. *The Construction of Preference*. Cambridge: Cambridge University Press.

Loewenstein, G. F., D. Read, and R. Baumeister, eds. 2003. *Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice*. New York: Russell Sage.

Lusardi, A., O. S. Mitchell, and V. Curto. 2010. "Financial Literacy Among the Young." *Journal of Consumer Affairs* 44, no. 2: 358–380. https://doi.org/10.1111/j.1745-6606.2010.01173.x.

Lyall, D. M., B. Cullen, M. Allerhand, et al. 2016. "Cognitive Test Scores in UK Biobank: Data Reduction in 480,416 Participants and Longitudinal Stability in 20,346 Participants." *PLoS ONE* 11, no. 4: e0154222. https://doi.org/10.1371/journal.pone.0154222.

Mamassian, P. 2008. "Overconfidence in an Objective Anticipatory Motor Task." *Psychological Science* 19: 601–606. https://doi.org/10.1111/j.1467-9280.2008.02129.x.

Manktelow, K. I. 2012. *Thinking and Reasoning*. Hove, England: Psychology Press.

McClure, S. M., D. I. Laibson, G. Loewenstein, and J. D. Cohen. 2004. "Separate Neural Systems Value Immediate and Delayed Monetary Rewards." *Science* 306: 503–507. https://doi.org/10.1126/science.1100907.

Meier, S., and C. D. Sprenger. 2012. "Time Discounting Predicts Creditworthiness." *Psychological Science* 23: 56–58. https://doi.org/10.1177/0956797611425931.

Mellers, B., R. Hertwig, and D. Kahneman. 2001. "Do Frequency Representations Eliminate Conjunction Effects? An Exercise in Adversarial Collaboration." *Psychological Science* 12: 269–275. https://doi.org/10.1111/1467-9280.00350.

Mischel, W. 2014. *The Marshmallow Test. Why Self-Control Is the Engine of Success*. New York: Little, Brown and Company.

Mischel, W., O. Ayduk, M. G. Berman, et al. 2011. "'Willpower' Over the Life Span: Decomposing Self-Regulation." *Social Cognitive and Affective Neuroscience* 6, no. 2: 252–256. https://doi.org/10.1093/scan/nsq081.

Mischel, W., and E. B. Ebbesen. 1970. "Attention in Delay of Gratification." *Journal of Personality and Social Psychology* 16: 329–337. https://doi.org/10.1037/h0029815.

Mischel, W., Y. Shoda, and M. L. Rodriguez. 1989. "Delay of Gratification in Children." *Science* 244: 933–938. https://doi.org/10.1126/science.2658056.

Myerson, J., A. A. Baumann, and L. Green. 2014. "Discounting of Delayed Rewards: (a) Theoretical Interpretation of the Kirby Questionnaire." *Behavioural Processes* 107: 99–105. https://doi.org/10.1016/j.beproc.2014.07.021.

Nozick, R. 1993. *The Nature of Rationality*. Princeton, NJ: Princeton University Press.

NWEA. 2019. *MAP® Growth™ Technical Report*. Portland, OR: Author.

Oaksford, M., and N. Chater. 1994. "A Rational Analysis of the Selection Task as Optimal Data Selection." *Psychological Review* 101: 608–631. https://doi.org/10.1037/0033-295X.101.4.608.

Oaksford, M., and N. Chater. 2007. *Bayesian Rationality: The Probabilistic Approach to Human Reasoning*. Oxford: Oxford University Press.

Otero, I., J. F. Salgado, and S. Moscoso. 2022. "Cognitive Reflection, Cognitive Intelligence, and Cognitive Abilities: A Meta-Analysis." *Intelligence* 90, no. 101: 614. https://doi.org/10.1016/j.intell.2021.101614.

Parker, A. M., W. Bruine de Bruin, and B. Fischhoff. 2015. "Negative Decision Outcomes Are More Common Among People with Lower Decision-Making Competence: An Item-Level Analysis of the Decision Outcome Inventory (DOI)." *Frontiers in Psychology* 6: 363. https://doi.org/10.3389/fpsyg.2015.00363.

Parker, A. M., W. Bruine de Bruin, B. Fischhoff, and J. Weller. 2018. "Robustness of Decision-Making Competence: Evidence From Two Measures and an 11-Year Longitudinal Study." *Journal of Behavioral Decision Making* 31, no. 3: 380–391. https://doi.org/10.1002/bdm.2059.

Parker, A. M., and B. Fischhoff. 2005. "Decision-Making Competence: External Validation Through an Individual Differences Approach." *Journal of Behavioral Decision Making* 18: 1–27. https://doi.org/10.1002/bdm.481.

Parker, A. M., and J. A. Weller. 2015. "Greater Decision-Making Competence Is Associated With Greater Expected-Value Sensitivity, but Not Overall Risk Taking: An Examination of Concurrent Validity." *Frontiers in Psychology* 6: 717. https://doi.org/10.3389/fpsyg.2015.00717.

Patel, N., S. G. Baker, and L. D. Scherer. 2019. "Evaluating the Cognitive Reflection Test as a Measure of Intuition/Reflection, Numeracy, and Insight Problem Solving, and the Implications for Understanding Real-World Judgments and Beliefs." *Journal of Experimental Psychology. General* 148, no. 12: 2129–2153. https://doi.org/10.1037/xge0000592.

Paulhus, D. L. 1991. "Measurement and Control of Response Bias." In *Measures of Personality and Social Psychological Attitudes*, edited by J. P. Robinson, P. Shaver, and L. S. Wrightsman, 17–59. San Diego, CA: Academic Press.

Peters, E. 2012. "Beyond Comprehension: The Role of Numeracy in Judgments and Decisions." *Current Directions in Psychological Science* 21, no. 1: 31–35. https://doi.org/10.1177/0963721411429960.

Peters, E. 2020. *Innumeracy in the Wild: Misunderstanding and Misusing Numbers*. New York: Oxford University Press.

Peters, E., D. Västfjäll, P. Slovic, C. K. Mertz, K. Mazzocco, and S. Dickert. 2006. "Numeracy and Decision Making." *Psychological Science* 17: 407–413. https://doi.org/10.1111/j.1467-9280.2006.01720.x.

Petry, N. M. 2001. "Substance Abuse, Pathological Gambling, and Impulsivity." *Drug and Alcohol Dependence* 63: 29–38. https://doi.org/10.1016/S0376-8716(00)00188-5.

Petry, N. M., and T. Casarella. 1999. "Excessive Discounting of Delayed Rewards in Substance Abusers With Gambling Problems." *Drug and Alcohol Dependence* 56: 25–32. https://doi.org/10.1016/S0376-8716(99)00010-1.

Pohl, R., ed. 2017. *Cognitive Illusions*. Second ed. New York: Routledge.

Prencipe, A., A. Kesek, J. Cohen, C. Lamm, M. D. Lewis, and P. D. Zelazo. 2006. "Development of Hot and Cool Executive Function During the Transition to Adolescence." *Journal of Experimental Child Psychology* 108: 621–637.

Primi, C., K. Morsanyi, M. A. Donati, S. Galli, and F. Chiesi. 2017. "Measuring Probabilistic Reasoning: The Construction of a New Scale Applying Item Response Theory." *Journal of Behavioral Decision Making* 30, no. 4: 933–950. https://doi.org/10.1002/bdm.2011.

Reimers, S., E. A. Maylor, N. Stewart, and N. Chater. 2009. "Associations Between a One-Shot Delay Discounting Measure and Age, Income, Education and Real-World Impulsive Behavior." *Personality and Individual Differences* 47: 973–978. https://doi.org/10.1016/j.paid.2009.07.026.

Resnik, M. D. 1987. *Choices: An Introduction to Decision Theory*. Minneapolis: University of Minnesota Press.

Reyna, V. F., and C. J. Brainerd. 2007. “The Importance of Mathematics in Health and Human Judgment: Numeracy, Risk Communication, and Medical Decision Making.” *Learning and Individual Differences* 17, no. 2: 147–159. https://doi.org/10.1016/j.lindif.2007.03.010.

Reyna, V. F., and S. Ellis. 1994. “Fuzzy-Trace Theory and Framing Effects in Children's Risky Decision Making.” *Psychological Science* 5: 275–279. https://doi.org/10.1111/j.1467-9280.1994.tb00625.x.

Reyna, V. F., W. L. Nelson, P. K. Han, and N. F. Dieckmann. 2009. “How Numeracy Influences Risk Comprehension and Medical Decision Making.” *Psychological Bulletin* 135: 943–973. https://doi.org/10.1037/a0017327.

Rodriguez, M. L., W. Mischel, and Y. Shoda. 1989. “Cognitive Person Variables in Delay of Gratification of Older Children at Risk.” *Journal of Personality and Social Psychology* 57: 358–367. https://doi.org/10.1037/0022-3514.57.2.358.

Roney, C. J. R., and N. Sansone. 2015. “Explaining the Gambler's Fallacy: Testing a Gestalt Explanation Versus the “Law of Small Numbers”.” *Thinking & Reasoning* 21: 193–205.

Ronis, D. L., and J. F. Yates. 1987. “Components of Probability Judgment Accuracy: Individual Consistency and Effects of Subject Matter and Assessment Method.” *Organizational Behavior and Human Decision Processes* 40: 193–218. https://doi.org/10.1016/0749-5978(87)90012-4.

Russo, J. E., and P. J. H. Schoemaker. 1992. “Managing Overconfidence.” *Sloan Management Review* 33, no. 2: 7–17.

Samuelson, P. L., and I. M. Church. 2015. “When Cognition Turns Vicious: Heuristics and Biases in Light of Virtue Epistemology.” *Philosophical Psychology* 28: 1095–1113.

Schlottman, A., and F. Wilkening. 2012. “Judgment and Decision Making in Young Children.” In *Judgment and Decision Making as a Skill: Learning, Development, and Evolution*, edited by M. K. Dhami, A. Schlottman, and M. R. Waldmann, 55–83. New York: Cambridge University Press.

Shafir, E., and A. Tversky. 1995. “Decision Making.” In *Thinking*, edited by E. E. Smith and D. N. Osherson, vol. 3, 77–100. Cambridge, MA: The MIT Press.

Shapiro, J. M. 2005. “Is There a Daily Discount Rate? Evidence From the Food Stamp Nutrition Cycle.” *Journal of Public Economics* 89: 303–325. https://doi.org/10.1016/j.jpubeco.2004.05.003.

Shipley, W. C., C. P. Gruber, T. A. Martin, and A. M. Klein. 2009. *Shipley-2 Manual*. Los Angeles, California: Western Psychological Services.

Sinayev, A., and E. Peters. 2015. “Cognitive Reflection Vs. Calculation in Decision Making.” *Frontiers in Psychology* 6, no. 532. https://doi.org/10.3389/fpsyg.2015.00532.

Slovic, P. 1995. “The Construction of Preference.” *American Psychologist* 50: 364–371. https://doi.org/10.1037/0003-066X.50.5.364.

Stanovich, K. E. 1999. *Who Is Rational? Studies of Individual Differences in Reasoning*. Mahwah, NJ: Erlbaum.

Stanovich, K. E. 2004. *The Robot's Rebellion: Finding Meaning in the Age of Darwin*. Chicago: University of Chicago Press.

Stanovich, K. E. 2009. *What Intelligence Tests Miss: The Psychology of Rational Thought*. New Haven, CT: Yale University Press.

Stanovich, K. E. 2010. *Decision Making and Rationality in the Modern World*. New York: Oxford University Press.

Stanovich, K. E. 2011. *Rationality and the Reflective Mind*. New York: Oxford University Press.

Stanovich, K. E. 2012. “On the Distinction Between Rationality and Intelligence: Implications for Understanding Individual Differences in Reasoning.” In *The Oxford Handbook of Thinking and Reasoning*, edited by K. Holyoak and R. Morrison, 343–365. New York: Oxford University Press.

Stanovich, K. E. 2013. “Why Humans Are (Sometimes) Less Rational Than Other Animals: Cognitive Complexity and the Axioms of Rational Choice.” *Thinking & Reasoning* 19: 1–26. https://doi.org/10.1080/13546783.2012.713178.

Stanovich, K. E. 2018. “Miserliness in Human Cognition: The Interaction of Detection, Override and Mindware.” *Thinking & Reasoning* 24: 423–444. https://doi.org/10.1080/13546783.2018.1459314.

Stanovich, K. E., and M. E. Toplak. 2019. “The Need for Intellectual Diversity in Psychological Science: Our Own Studies of Actively Open-Minded Thinking as a Case Study.” *Cognition* 187: 156–166. https://doi.org/10.1016/j.cognition.2019.03.006.

Stanovich, K. E., and M. E. Toplak. 2023. “Actively Open-Minded Thinking and Its Measurement.” *Journal of Intelligence* 11, no. 2: 27. https://doi.org/10.3390/jintelligence11020027.

Stanovich, K. E., and R. F. West. 1997. “Reasoning Independently of Prior Belief and Individual Differences in Actively Open-Minded Thinking.” *Journal of Educational Psychology* 89: 342–357. https://doi.org/10.1037/0022-0663.89.2.342.

Stanovich, K. E., and R. F. West. 1998a. “Cognitive Ability and Variation in Selection Task Performance.” *Thinking and Reasoning* 4: 193–230.

Stanovich, K. E., and R. F. West. 1998b. “Individual Differences in Rational Thought.” *Journal of Experimental Psychology. General* 127: 161–188.

Stanovich, K. E., and R. F. West. 2007. “Natural Myside Bias Is Independent of Cognitive Ability.” *Thinking & Reasoning* 13: 225–247. https://doi.org/10.1080/13546780600780796.

Stanovich, K. E., R. F. West, and M. E. Toplak. 2011. “The Complexity of Developmental Predictions From Dual Process Models.” *Developmental Review* 31: 103–118. https://doi.org/10.1016/j.dr.2011.07.003.

Stanovich, K. E., R. F. West, and M. E. Toplak. 2016. *The Rationality Quotient: Toward a Test of Rational Thinking*. Cambridge, MA: MIT Press.

Starmer, C. 2000. “Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice Under Risk.” *Journal of Economic Literature* 38: 332–382. https://doi.org/10.1257/jel.38.2.332.

Staudinger, U. M., J. Dorner, and C. Mickler. 2005. “Wisdom and Personality.” In *A Handbook of Wisdom: Psychological Perspectives*, edited by R. J. Sternberg and J. Jordan, 191–219. New York: Cambridge University Press.

Steinberg, L., S. Graham, L. O'Brien, J. Woorlard, E. Cauffman, and M. Banich. 2009. “Age Differences in Future Orientation and Delay Discounting.” *Child Development* 80, no. 1: 28–44. https://doi.org/10.1111/j.1467-8624.2008.01244.x.

Sternberg, R. J. 2003. *Wisdom, Intelligence, and Creativity Synthesized*. Cambridge: Cambridge University Press.

Sternberg, R. J., and K. Sternberg. 2017. “Measuring Scientific Reasoning for Graduate Admissions in Psychology and Related Disciplines.” *Journal of Intelligence* 5, no. 3: 29. https://doi.org/10.3390/jintelligence5030029.

Strathman, A., F. Gleicher, D. S. Boninger, and C. Scott Edwards. 1994. “The Consideration of Future Consequences: Weighing Immediate and Distant Outcomes of Behavior.” *Journal of Personality and Social Psychology* 66: 742–752. https://doi.org/10.1037/0022-3514.66.4.742.

Tauber, S. K., J. Dunlosky, K. A. Rawson, M. G. Rhodes, and D. M. Sitzman. 2013. “General Knowledge Norms: Updated and Expanded from the Nelson and Narens (1980) Norms.” *Behavior Research Methods* 45: 1115–1143. https://doi.org/10.3758/s13428-012-0307-9.

Tetlock, P. E. 2005. *Expert Political Judgment*. Princeton, NJ: Princeton University Press.

Thaler, R. H. 2015. *Misbehaving: The Making of Behavioral Economics*. New York: Norton.

Toplak, M. E. 2022. *Cognitive Sophistication and the Development of Judgment and Decision-Making*. London, UK: Elsevier.

Toplak, M. E., and D. B. Flora. 2021. "Resistance to Cognitive Biases: Longitudinal Trajectories and Associations With Cognitive Abilities and Academic Achievement Across Development." *Journal of Behavioral Decision Making* 34, no. 3: 344–358. https://doi.org/10.1002/bdm.2214.

Toplak, M. E., R. F. West, and K. E. Stanovich. 2011. "The Cognitive Reflection Test as a Predictor of Performance on Heuristics and Biases Tasks." *Memory & Cognition* 39: 1275–1289. https://doi.org/10.3758/s13421-011-0104-1.

Toplak, M. E., R. F. West, and K. E. Stanovich. 2014a. "Rational Thinking and Cognitive Sophistication: Development, Cognitive Abilities, and Thinking Dispositions." *Developmental Psychology* 50: 1037–1048.

Toplak, M. E., R. F. West, and K. E. Stanovich. 2014b. "Assessing Miserly Processing: An Expansion of the Cognitive Reflection Test." *Thinking & Reasoning* 20: 147–168.

Toplak, M. E., R. F. West, and K. E. Stanovich. 2017. "Real-World Correlates of Performance on Heuristics and Biases Tasks in a Community Sample." *Journal of Behavioral Decision Making* 30: 541–554. https://doi.org/10.1002/bdm.1973.

Tversky, A., and D. Kahneman. 1974. "Judgment Under Uncertainty: Heuristics and Biases." *Science* 185: 1124–1131. https://doi.org/10.1126/science.185.4157.1124.

Tversky, A., and D. Kahneman. 1981. "The Framing of Decisions and the Psychology of Choice." *Science* 211: 453–458. https://doi.org/10.1126/science.7455683.

Tversky, A., and D. Kahneman. 1983. "Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment." *Psychological Review* 90: 293–315. https://doi.org/10.1037/0033-295X.90.4.293.

Vallone, R., D. W. Griffin, S. Lin, and L. Ross. 1990. "Overconfident Prediction of Future Actions and Outcomes by Self and Others." *Journal of Personality and Social Psychology* 58: 582–592. https://doi.org/10.1037/0022-3514.58.4.582.

Viator, R. E., N. L. Harp, S. B. Rinaldo, and B. B. Marquardt. 2020. "The Mediating Effect of Reflective-Analytic Cognitive Style on Rational Thought." *Thinking & Reasoning* 26: 381–413. https://doi.org/10.1080/13546783.2019.1634151.

Wason, P. C. 1966. "Reasoning." In *New Horizons in Psychology I*, edited by B. M. Foss, 106–137. London, UK: Penguin.

Weller, J. A., I. P. Levin, J. P. Rose, and E. Bossard. 2012. "Assessment of Decision-Making Competence in Preadolescence." *Journal of Behavioral Decision Making* 25, no. 4: 414–426. https://doi.org/10.1002/bdm.744.

Weller, J. A., M. Moholy, E. Bossard, and I. P. Levin. 2015. "Preadolescent Decision-Making Competence Predicts Interpersonal Strengths and Difficulties: A 2-Year Prospective Study." *Journal of Behavioral Decision Making* 28, no. 1: 76–88. https://doi.org/10.1002/bdm.1822.

Weller, J. A., A. M. Parker, M. Reynolds, and L. Kirisci. 2021. "Developmental Trajectory Classes in Psychological Dysregulation Predict Later Decision-Making Competence." *Addictive Behaviors* 112, no. 106: 650.

Weller, J. A., A. M. Parker, M. Reynolds, L. Kirisci, and L. Michaels. 2024. "Longitudinal Associations Between Parenting Practices and Child's Later Decision-Making Competence." *Applied Developmental Science*. https://doi.org/10.1080/10888691.2023.2287203.

West, R. F., and K. E. Stanovich. 1997. "The Domain Specificity and Generality of Overconfidence: Individual Differences in Performance Estimation Bias." *Psychonomic Bulletin & Review* 4: 387–392. https://doi.org/10.3758/BF03210798.

West, R. F., and K. E. Stanovich. 2003. "Is Probability Matching Smart? Associations Between Probabilistic Choices and Cognitive Ability." *Memory & Cognition* 31: 243–251. https://doi.org/10.3758/BF03194383.

Xu, J., and N. Harvey. 2014. "Carry on Winning: The Gamblers' Fallacy Creates Hot Hand Effects in Online Gambling." *Cognition* 131: 173–180. https://doi.org/10.1016/j.cognition.2014.01.002.

Yates, J. F., J. Lee, and J. G. Bush. 1997. "General Knowledge Overconfidence: Cross-National Variations, Response Style, and "Reality"." *Organizational Behavior and Human Decision Processes* 70: 87–94. https://doi.org/10.1006/obhd.1997.2696.

Yates, J. F., Y. Zhu, D. Ronis, D. Wang, H. Shinotsuka, and M. Toda. 1989. "Probability Judgment Accuracy:China, Japan, and the United States." *Organizational Behavior and Human Decision Processes* 43: 145–171. https://doi.org/10.1016/0749-5978(89)90048-4.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.