Thinking & Reasoning

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ptar20

Understanding the relationship between rationality and intelligence: a latent-variable approach

Alexander P. Burgoyne, Cody A. Mashburn, Jason S. Tsukahara, David Z. Hambrick & Randall W. Engle

To cite this article: Alexander P. Burgoyne, Cody A. Mashburn, Jason S. Tsukahara, David Z. Hambrick & Randall W. Engle (2021): Understanding the relationship between rationality and intelligence: a latent-variable approach, Thinking & Reasoning, DOI: 10.1080/13546783.2021.2008003

To link to this article: https://doi.org/10.1080/13546783.2021.2008003

Published online: 26 Nov 2021.

THINKING & REASONING
https://doi.org/10.1080/13546783.2021.2008003

Understanding the relationship between rationality and intelligence: a latent-variable approach

Alexander P. Burgoyne[a], Cody A. Mashburn[a], Jason S. Tsukahara[a], David Z. Hambrick[b] and Randall W. Engle[a]

[a]Georgia Institute of Technology, Atlanta, GA, USA; [b]Michigan State University, East Lansing, Michigan, USA

ABSTRACT
A hallmark of intelligent behavior is *rationality* – the disposition and ability to think analytically to make decisions that maximize expected utility or follow the laws of probability. However, the question remains as to whether rationality and intelligence are empirically distinct, as does the question of what cognitive mechanisms underlie individual differences in rationality. In a sample of 331 participants, we assessed the relationship between rationality and intelligence. There was a common ability underpinning performance on some, but not all, rationality tests. Latent factors representing rationality and general intelligence were strongly correlated ($r = .54$), but their correlation fell well short of unity. Rationality correlated significantly with fluid intelligence ($r = .56$), working memory capacity ($r = .44$), and attention control ($r = .49$). Attention control fully accounted for the relationship between working memory capacity and rationality, and partially accounted for the relationship between fluid intelligence and rationality. We conclude by speculating about factors rationality tests may tap that other cognitive ability tests miss, and outline directions for further research.

ARTICLE HISTORY Received July 19, 2021; Accepted November 12, 2021

KEYWORDS Rationality; intelligence; attention control; working memory capacity; fluid intelligence

A hallmark of intelligent behavior is *rationality* – the disposition and ability to think analytically to make decisions that maximize expected utility or follow the laws of probability, and therefore align with normative principles of decision making (Evans & Over, 1996; Stanovich, 2011). Examples range from the mundane, such as deciding whether to purchase a discounted item at the grocery store, to the momentous, such as deciding how to

CONTACT Alexander P. Burgoyne burgoyn4@gmail.com School of Psychology, Georgia Institute of Technology, Atlanta, GA, USA.

invest for retirement. Indeed, measures of rationality predict important real-world outcomes; low decision-making competence predicts juvenile delinquency in adolescents (Parker & Fischhoff, 2005) and bankruptcy and other adverse consequences in adults (Parker et al., 2015).

Rationality is a complex cognitive construct that has been studied in many psychological domains. For example, rationality is *relative*; decisions that benefit oneself or one's group may be at odds with one another, so whether a decision is considered optimal may depend on the context in which the decision was made, or the perspective of the evaluator (Gigerenzer, 1996). Furthermore, scholars have distinguished between at least two types of rationality (for a review, see Plantinga, 1993). *Epistemic rationality* refers to making decisions or holding beliefs based on strong support from one's evidence (Kelly, 2003). *Instrumental rationality*, on the other hand, refers to making decisions that satisfy one's goals (Kelly, 2003). Finally, rationality and related measures have been applied beyond traditional decision-making contexts involving probabilistic reasoning, to domains ranging from religious belief and conspiratorial ideation (Pennycook et al., 2020) to susceptibility to "fake news" (Pennycook & Rand, 2019). Findings in each of these research contexts have emphasized the importance of one's *disposition* to think analytically, in addition to one's ability to do so (Pennycook et al., 2015). Given the scope of rationality as a cognitive construct, for the present purposes, we focus specifically on the instrumental view of rationality, with an emphasis on the disposition and ability to engage in analytic thinking and apply the laws of probability in decision-making contexts.

From this standpoint, rationality tests typically take the form of logic problems that require the solver to avoid using heuristics and biases. Incorrect responses can result from neglecting base rates, committing the conjunction fallacy (as in the classic "Linda" problem), relying on misleading anchors, displaying overconfidence, and so on, depending on the test. While research on the cognitive processes underpinning rationality has shed light on heuristics, biases, and individual differences in susceptibility to them (Bruine de Bruin et al., 2020; Tversky & Kahneman, 1974), an important question remains: Are rationality and intelligence empirically distinct?

On its face, rationality might appear to be subsumed by *fluid intelligence* – the ability to reason to solve novel problems, particularly those that cannot be solved automatically (McGrew, 2009). And, in fact, measures of rationality and fluid intelligence correlate positively and significantly (Bruine de Bruin et al., 2007; Stanovich & West, 1998a; Toplak et al., 2011; 2014a), and rationality tests are usually novel to the problem solver (but see Toplak et al., 2014b). At a theoretical level, fluid intelligence supports rationality via hypothesis generation and disengagement from unsuccessful solution attempts (McGrew, 2009; Shipstead et al., 2016). That is, people with higher fluid intelligence are better able to "overcome [incorrect] initial impressions and hypotheses" that are counterproductive to efficient search of the problem space (Shipstead et al., 2016, p. 776). Despite robust correlations and conceptual overlap, however, variance in rationality remains unaccounted for by fluid intelligence (e.g., Bruine de Bruin et al., 2007). What's more, individual differences in decision-making competence predict negative life outcomes (as indexed by the Decision Outcomes Inventory) even after controlling for fluid intelligence (Bruine de Bruin et al., 2007). These results suggest that rationality tests capture something that fluid intelligence tests do not. One perspective, described in detail in the "Dual Process Theory" section below, is that rationality measures uniquely capture a person's disposition or tendency to think reflectively, particularly in contexts that evoke an automatic (i.e., non-reflective or analytic) response (c.f., Stanovich et al., 2016).

That said, fluid intelligence is not the only broad cognitive ability that might contribute to rationality; working memory and attention control may also play a role (Evans & Stanovich, 2013). Working memory is the cognitive system responsible for the temporary maintenance of information in a highly accessible state (Baddeley, 1992), and is hypothesized to support rationality via *cognitive decoupling* – maintaining a secondary representation of a real-world scenario for the purposes of hypothesis testing and simulation (Leslie, 1987). Working memory capacity has been found to correlate with performance on rationality tests (De Neys et al., 2005; Handley et al., 2004; Markovits & Doyon, 2004; Toplak et al., 2011; 2014a), and manipulations that place cognitive load on the central executive – and therefore demand attention control – negatively impact decision making (De Neys, 2006a, 2006b; De Neys et al., 2005).

Broadly speaking, working memory is supported by the interplay between a central executive attention component and short-term storage components (Baddeley, 1992). Our view, termed the executive attention theory of working memory capacity (Kane & Engle, 2002), is that attention control largely accounts for the predictive validity of working memory capacity measures. *Attention control* refers to the ability to guide thoughts and behavior in a goal-directed manner, particularly under conditions of interference between automatic and controlled processes (Burgoyne & Engle, 2020; Engle, 2018). Evidence for the executive attention view is provided by latent variable analyses demonstrating that controlled attention largely accounts for working memory capacity's relationships with other cognitive constructs. For example, Engle et al. (1999) estimated relationships between latent factors representing working memory capacity, short-term memory, and fluid intelligence. They found that working memory capacity significantly predicted fluid intelligence even after accounting for short-term memory, whereas short-term memory did not predict fluid intelligence after accounting for working memory capacity. This suggests that the controlled attention required by working memory tasks, which challenge participants to flexibly shift attention between storing and processing information, drives the relationship between working memory capacity and fluid intelligence.

As another example, Draheim et al. (2020) estimated the relationships between working memory capacity, attention control, and fluid intelligence at the latent level in a sample of 396 participants. Attention control was measured using a battery of new and improved tasks, including the *antisaccade task*, in which participants must inhibit a prepotent response (i.e., *don't look at the flickering asterisk*) and generate and execute a countervailing controlled response (i.e., *do look away from the flickering asterisk*; Unsworth et al., 2004). Other tasks measuring attention control, such as the adaptive-difficulty Stroop and flanker tasks, similarly required the performer to suppress an automatic response and provide a controlled alternative. Draheim et al. (2020) found that attention control could mediate the relationship between working memory capacity and fluid intelligence. That is, depending on how attention control was measured, the relationship between working memory capacity and fluid intelligence was no longer statistically significant after accounting for attention control. This suggests that the primary reason working memory capacity and fluid intelligence are correlated is because they both tap cognitive functions supported by controlled attention.

At a theoretical level, attention control supports two distinct but complementary cognitive functions that are theorized to play a role in working memory tasks, fluid intelligence tasks, and tests of rational thinking: maintenance and disengagement (Burgoyne & Engle, 2020; Shipstead et al., 2016). *Maintenance* refers to cognitive operations that involve keeping track of information, such as items that must be remembered, task goals, or details about the problem to be solved. Maintenance is required by complex span tests of working memory capacity because participants must keep track of items (e.g., letters, digits) to successfully perform the task. Maintenance also plays a role in fluid intelligence tests, because participants must use problem-relevant information to generate and evaluate hypotheses. Similarly, maintenance plays a role in tests of rational thinking, which often require the maintenance of secondary representations for hypothesis testing (Barrett et al., 2004; Engle, 2002; Evans, 2003; Shipstead et al., 2016). *Disengagement*, on the other hand, refers to cognitive operations that involve removing outdated information from active processing, and preventing outdated information from re-entering the focus of attention. Disengagement plays a role in complex span tests of working memory capacity because participants must disengage from (i.e., discard) the to-be-remembered items from previous trials. Disengagement also contributes to fluid intelligence test performance because participants must remove hypotheses that have been ruled out from active processing. Finally, disengagement appears to play an important role in rationality tests; rationality items are often designed to automatically cue an incorrect response, but this response must be inhibited for the problem to be solved correctly.

Thus, attention control may contribute to individual differences in rationality in multiple ways. To reiterate, it may support the maintenance and manipulation of information in service of rational thought, such as intermediate products and expected utilities. A prediction that follows from this line of reasoning is that attention control should account for the relationship between working memory capacity and rationality. This prediction remains untested, although studies that have burdened the central executive via a secondary task and found decrements in performance on rationality tests provide supporting evidence (e.g., De Neys et al., 2005). Attention control may also support the overriding of automatic responses on rationality items, and disengaging from hypotheses that have been ruled out. A prediction that follows from this argument is that attention control should account for the relationship between fluid intelligence and rationality. While these possible roles of attention control in supporting rationality have yet to be disentangled empirically, differences in attention control (i.e., executive functions) have been shown to predict performance on rationality tests (Basile & Toplak, 2015; Del Missier et al., 2010, 2012; Handley et al., 2004; Toplak et al., 2014a). One purpose of the present study is to provide the first empirical test of these predictions using latent variable analyses.

Dual process theory

The theoretical contribution of attention control to rationality is supported by dual-process theories of cognition. Dual process theory posits that cognitive operations can be categorized into one of two modes: *Type 1 processes* are described as fast, parallel, automatic, intuitive, and seemingly effortless, whereas *Type 2 processes* are described as slow, sequential, controlled, reflective, and effortful (Evans, 2008; Evans & Stanovich, 2013; Frankish, 2010). The defining characteristic of Type 1 processing is *autonomy* (Stanovich & Toplak, 2012). That is, Type 1 processes require minimal controlled attention and, as a result, place minimal demands on working memory (Evans & Stanovich, 2013). By contrast, the defining characteristic of Type 2 processing is the deliberate and effortful thought process that requires the engagement of attention control and working memory (Evans, 2008). Although Type 1 processes provide quick and intuitive responses, these automated responses are not always optimal. Thus, one of the primary roles of Type 2 processing is to block or override Type 1 processing when appropriate (Toplak et al., 2014a).

Attention control tasks measure one's *ability* to override automatic responses, which is similar to but not the same as one's *disposition* towards overriding automatic responses. Although individual differences in attention control correlate with self-reported thinking dispositions (Basile & Toplak, 2015; Toplak et al., 2014a, but see Toplak et al., 2011), it has been argued that thinking dispositions are a critical construct tapped by rationality tests that cognitive ability tests miss (Stanovich, 2009; Stanovich et al., 2016). Referring to the distinction between cognitive ability and thinking dispositions, Evans (2008) stated: "The difference is between what people are able to do and *what they are inclined to do*" (p. 262, italics added).

By way of example, consider the "Bat and Ball" problem from the Cognitive Reflection Test (Frederick, 2005): "A bat and a ball cost \$1.10 in total. The bat costs \$1.00 more than the ball. How much does the ball cost?" The problem is designed to evoke an intuitive yet incorrect response, and indeed, people often report that "10 cents" immediately springs to mind (Frederick, 2005). The correct answer, however, is not 10 cents, but 5 cents: (\$0.05 + \$1.05 = \$1.10). Given the fairly minimal computational demands of the solution, performance on this problem has been taken to measure one's disposition and ability to inhibit an automatic response and think reflectively before responding (Campitelli & Gerrans, 2014; Evans, 2008; Frederick, 2005; Stanovich, 2012; Toplak et al., 2011).

Errors on rationality tests can arise for a number of reasons within the dual-process framework. For example, the problem solver may lack the ability or motivation to override Type 1 processing or not recognize the need to, they may lack the control of attention needed to maintain decoupling operations, or they may lack the requisite computational ability, strategy, or knowledge to solve the problem using Type 2 processing (Stanovich, 2012; Stanovich et al., 2016). This multifactorial perspective on errors, coupled with the different heuristics or cognitive biases tapped by the tasks, might explain why measures of rationality sometimes correlate weakly with each other and show differential relations to cognitive ability. For example, factor analyses of rationality tests often extract more than one factor (Parker & Fischhoff, 2005; Teovanović et al., 2015; Weaver & Stewart, 2012), and while many measures of rationality correlate with cognitive ability, others, such as *myside bias* – the tendency to evaluate propositions in a manner biased towards one's own opinions – do not (Stanovich & West, 2007; see also Del Missier et al., 2012).

Likewise, correct responses on rationality tests can arise from a number of sources within the dual-process framework. Specifically, a recent study by Thompson et al. (2018) revealed that participants with greater cognitive ability as measured by the Shipley-2 standardized intelligence test (Shipley et al., 2009) were more likely to provide accurate Type 1 responses to reasoning problems than participants who were lower in cognitive ability. One interpretation of this result is that smarter participants are better able to successfully perform probabilistic reasoning in an automatic or intuitive fashion because this analytical approach to problem solving has become natural through practice (Stanovich et al., 2011). Thus, contrary to the notion that correct responses on reasoning problems can only occur through Type 2 deliberative processing, there is some evidence that Type 1 processes can provide correct (and analytically grounded) responses to reasoning problems for some participants. This poses a wrinkle for understanding the link between rationality and intelligence within the dual-process framework, because it suggests that Type 1 processes may not necessarily be a source of faulty reasoning among those with greater cognitive ability.

The present study

In the present study, we examined relationships among broad cognitive ability factors and a rationality factor using latent variable analyses. Latent variables capture variance common to a set of measures and are thus theoretically free of measurement error and more closely approximate the cognitive constructs of interest than observed measures (Kline, 2015). Although other studies have assessed the relationship between cognitive ability and rationality, few have done so using latent variables, and none have examined whether attention control accounts for the working memory capacity-rationality relationship, or the fluid intelligence-rationality relationship. Therefore, the present study goes beyond prior work by clarifying the relative contribution of latent cognitive ability factors to individual differences in rationality using a sample of more than 300 participants representing a wide range of ability.

Our cognitive ability task battery included three measures of fluid intelligence, three measures of working memory capacity, and four measures of attention control. Our tests of rationality included the Wason selection task, the conjunction fallacy test, and two tests assessing base rate neglect. We chose these tasks to test the viability of a latent rationality factor. Given limited testing time and the large number of potential heuristics and cognitive biases available for study, a comprehensive assessment of all aspects of rational thinking was not feasible for this study. For example, Stanovich (2016) lists over 30 heuristics and biases, and the Comprehensive

Assessment of Rational Thinking contains 20 subtests. Furthermore, the data reported in the present study were collected prior to the publication of the Comprehensive Assessment of Rational Thinking. If there is a coherent rationality construct to be measured, however, the tasks we selected should provide evidence for it. The base rate neglect and conjunction fallacy tasks are indicators of probabilistic reasoning, and might be expected to cohere even in the absence of an underlying rationality factor. The Wason selection task, on the other hand, is a measure of scientific thinking and confirmation bias. In the absence of a common rationality factor, there is little reason to expect it to cohere with the probabilistic reasoning tests after accounting for individual differences in general intelligence.

In summary, we addressed five questions: (1) Do measures of rational thinking form a latent factor? (2) To what extent are general intelligence and rationality related at the latent level? (3) Do rationality measures form a latent factor after accounting for individual differences in general intelligence? (4) Does attention control account for the relationship between working memory capacity and rationality? And (5) Does attention control account for the relationship between fluid intelligence and rationality? To address these questions, we performed confirmatory factor analysis on the rationality measures and used structural equation modeling to investigate relationships among cognitive abilities and performance on rationality tests.

Method

Participants

Our initial sample consisted of 352 participants recruited from the Georgia Institute of Technology and the Atlanta community. All participants provided informed consent. All participants were native English speakers (i.e., learned English before age 5), ranged in age from 18-35, and had not participated in a study in our lab before. Subjects were paid $30 at the end of each session and received an additional $10 check as a bonus at the end of the fourth session. Georgia Tech students could choose to receive 2 hours of course credits per session instead of monetary compensation. The data reported in this project were collected as part of a larger research effort; further details and an updated reference list of all publications resulting from this larger research effort can be found at https://osf.io/5da6j/.

We excluded cases with excessive missing data (see below), either due to equipment failures, attrition, or outlying scores. Values falling 3.5 standard deviations above or below the mean were identified as univariate outliers and set to missing. This procedure was applied to all tasks except the Wason selection task, which had a pronounced floor effect ($M = 0.74$ out of a possible 10) producing five positive outliers. Excluding these five cases

Table 1. Demographic information.

Demographic	Statistic
Age (years)	Mean: 24.5 *SD*: 4.6 Range: 18–35
Gender	Male: 49% Female: 51%
At least some college?	Yes: 51% No: 49%
Ethnicity	White: 60.4% Black or African American: 18.4% Asian or Pacific Islander: 9.7% Other*: 9.1% N/A: 1.8%

Note. Demographic information was unavailable for two participants. *Other includes, Hispanic or Latino, Native American, and Other.

attenuated the relationships among rationality and cognitive ability measures, suggesting that these five participants performed well on the task battery. We elected to include these cases in analyses.

Missing data led to the exclusion of nine participants who were missing scores on two or more measures of working memory capacity or fluid intelligence, or three or more measures of attention control. The different exclusion criteria for the cognitive ability constructs stems from the different number of tasks defining latent variables, but share the rationale that, to be included in the analysis, participants must not have missing data for a majority of the tasks reflecting a construct. Of the remaining 343 participants, we excluded 12 who were missing scores on one of the rationality measures. The final sample consisted of 331 participants. Demographic information is summarized in Table 1.

Procedure

Participants completed computerized tests of cognitive ability and rationality in small groups over the course of four two-hour sessions. Participants scheduled their testing sessions based on their availability, under the constraint that they could not complete two sessions on the same day. The first three sessions included tasks that are relevant to the present study. During Session 1, participants completed Advanced Operation Span, Raven's Advanced Progressive Matrices, Antisaccade, Flanker, Stereotype Base Rate Neglect, Diagnostic Base Rate Neglect, and Selective Visual Arrays. During Session 2, they completed Advanced Symmetry Span, Number Series, Stroop, and the Conjunction Fallacy task. During Session 3, they completed Advanced Rotation Span, Letter Sets, and the Wason Selection task. For more information about the study procedures and additional tasks completed by participants, please see https://osf.io/5da6j/.

Fluid intelligence

Raven's advanced progressive matrices (Raven & Court, 1998). Participants were presented with 3×3 arrays of geometric patterns. Each array contained a missing item, and participants were to select the pattern that best completed the array. Participants were given 10 minutes to complete the 18 odd-numbered items from Raven's Advanced Progressive Matrices. The measure was the number correct.

Letter sets (Ekstrom, 1976). Participants were presented with five sets of four letters arranged in a row and were to select the set that did not follow the same pattern as the other four. For example, for the sets NLIK, PLIK, QLIK, THIK, and VLIK, the correct response is THIK because the other sets all contain L. Participants were given 5 minutes to complete 30 items. The measure was the number correct.

Number series (Thurstone, 1938). Participants were presented with a series of numbers, and were to select which of four alternatives logically completed the series. Participants were given 5 minutes to complete 15 items. The measure was the number correct.

Working memory capacity

Advanced operation span (Draheim et al., 2018; Unsworth et al., 2005). Participants solved math equations and remembered a letter that followed each equation. After a series of trials, participants recalled the letters in the presented order. Set sizes ranged from 3 to 8 letters. Each set occurred 2 times, with the exception of the 8 letter set size, which occurred 4 times. The measure was the total number of letters recalled in the correct order.

Advanced symmetry span (Draheim et al., 2018; Unsworth et al., 2005). Participants made symmetry judgements about patterns and remembered the location of a square that appeared after each pattern. After a series of trials, participants recalled the location of the squares in the presented order. Set sizes ranged from 2 to 7 spatial locations and each set occurred 2 times.

Advanced rotation span (Draheim et al., 2018; Kane et al., 2004). Participants remembered a series of directional arrows (8 directions) of varying size (small or large) in alternation with a mental rotation task in which they had to mentally rotate and decide if a letter was mirror-reversed or not. Set-sizes ranged from 2 to 7 memory items and each set occurred 2 times.

Attention control

Antisaccade (Hallett, 1978; Hutchison, 2007). Participants identified a "Q" or "O" that appeared briefly on the opposite side of the screen as a distractor stimulus. After a central fixation cross appeared for 1000 ms or 2000 ms, an asterisk (*) flashed at 12.3° visual angle to the left or right of the central fixation for 100 ms. Afterward, the letter "Q" or "O" was presented on the opposite side at 12.3° visual angle of the central fixation for 100 ms, immediately followed by a visual mask (##). Participants indicated whether the letter was a "Q" or an "O". They completed 16 slow practice trials during which letter duration was set to 750 ms, followed by 48 test trials. The measure was the proportion correct.

Stroop (Stroop, 1935). Participants were presented with a word ("RED", "GREEN", or "BLUE") and indicated its hue (red, green, or blue). On each trial, there was a central fixation point (400-700ms) followed by a centrally presented word. The participant pressed one of three keys labeled with the colors green, blue, and red to indicate their response. They completed 486 trials. For two-thirds of the trials, the hue and word were congruent. For the other trials, the hue and word were incongruent. The measure was the mean reaction time on congruent trials subtracted from the mean reaction time on incongruent trials. Higher scores indicate a larger interference effect (i.e., worse performance). Only accurate trials were used to calculate mean reaction times.

Flanker (Eriksen & Eriksen, 1974). Participants identified the direction (left or right) of a middle arrow that was flanked by arrows pointing either the same or opposite direction. On each trial, there was a central fixation point (900 ms) followed by a centrally presented row of 5 items. The middle item was an arrow pointing to the left or right. Participants indicated the direction of the middle arrow as quickly and accurately as possible by pressing the "z" (left) or "/" (right) key. The arrows flanking the middle arrow were either facing the same direction as the middle arrow (congruent trials), facing the opposite direction as the middle arrow (incongruent trials), or were replaced by horizontal lines (neutral trials). There were 216 trials with 72 trials of each type. The measure was the mean reaction time on congruent trials subtracted from the mean reaction time on incongruent trials. Higher scores indicate a larger interference effect (i.e., worse performance). Only accurate trials were used to calculate mean reaction times.

Selective visual arrays (Luck & Vogel, 1997; Martin et al., 2019; Shipstead et al., 2014). After a central fixation of 1000 ms, a cue word ("RED" or "BLUE") appeared instructing the participant to attend to either red or blue rectangles. Next, a target array of red and blue rectangles of different orientations (horizontal, left diagonal, right diagonal, and vertical) was presented for 250 ms, followed by a blank screen for 900 ms. Next, a probe array with only the cued-color rectangles was presented, with one rectangle highlighted by a white dot. The orientation of the highlighted rectangle was either the same as it was in the target array, or different, with equal likelihood. The participant indicated with the keyboard whether the orientation of the highlighted rectangle had changed or stayed the same. The target array contained either 5 or 7 rectangles per color (10 and 14 total). There were 48 trials per array set size. The measure was a capacity score (k), calculated using the single-probe correction (Cowan et al., 2005): set size $*$ (hit rate $+$ correction rejection rate $-$ 1). The measure was the mean k estimate for the two set sizes.

Rationality

Wason selection (Wason, 1968). Participants selected two cards to determine whether a rule was true or not. On each problem, participants were shown four cards with a rule beneath them. They were told that each card has two sides with different information presented on it, and to select the two cards that needed to be turned over to determine whether the rule was true or false. The rule for each problem was conditional logic that if there was certain information on one side of a card there must be certain information on the other side (e.g., "If a card has an A on one side, then it has an even number on the other side"; the face-up cards for this problem were "A", "D", "4", and "7"). The first card choice is easy; most people select the card that has an "A" to test if there is an even number on the other side. The second card choice is more difficult. People often incorrectly choose the card that appears to provide confirmatory evidence for the rule, but cannot falsify it (i.e., confirmation bias). For example, selecting the card that has the even number "4" on it cannot falsify the rule, because even if this card has a non-A letter on the other side, the rule does not state any relationship between non-A cards and even/odd numbers. The correct response is the card that provides disconfirming evidence, in this case, the card that has the odd number "7" on it. If this card has an A on the other side, then this proves that the rule is false. There were 10 problems with no time limit. There were two problem types. Some were abstract (e.g., the "A" problem above), whereas others were deontic (e.g., a rule about a patron's age if they are drinking beer). The measure was the number of problems answered correctly. Items and item-level details are included in the Appendix.

Conjunction fallacy (Tversky & Kahneman, 1974). Participants were given a scenario and decided which of two statements was more likely. One of the statements was a conjunction that included the other statement. The conjunction of two probabilities is always less likely than either of the constituent probabilities considered alone. The conjunction statement, however, was framed to seem more probable given the scenario. For instance, the "Linda" problem describes a stereotypical feminist activist; the two statements are "Linda is a bank teller" and "Linda is a bank teller and is active in the feminist movement". Most people incorrectly choose the latter option – the conjunction – because it describes a feminist activist. The more probable statement, however, is that "Linda is a bank teller". There were seven problems with no time limit. The measure was the number of problems answered correctly. Items and item-level details are included in the Appendix.

Stereotype base rate neglect (De Neys & Glumicic, 2008; Tversky & Kahneman, 1974). Participants were given a scenario and decided which of two statements was more likely. For each scenario, they were given base rate information about a sample (e.g., "there are 4 men and 996 women in a sample of 1000") and a description of a person randomly selected from that sample. The description of the randomly selected person contained stereotypes pertaining to either the majority or minority group within the sample (e.g., a woman or man). The participant was asked to decide which group the randomly selected person was most likely to belong to. In De Neys and Glumicic (2008) version of the task, the correct response is always the one that is consistent with the base rate information (e.g., "woman" is the correct answer because 996/1000 of the sample were women and the person was selected randomly), although this response option is often at odds with the stereotype information. There are two item types; on non-conflict trials the stereotypical information was consistent with the base rate information, whereas on conflict trials the stereotypical information led participants towards the wrong answer relative to the base rate information. Therefore, on conflict trials, a wrong answer suggests the participant neglected the base rate information. The measure was the number of correct responses on conflict items. Items and item-level details are included in the Appendix.

Diagnostic base rate neglect (Bar-Hillel, 1977). Participants were given a scenario and determined the likelihood of an event. The scenario included base rate information (e.g., "1% of population has skin cancer") and diagnostic information (e.g., "a test will detect skin cancer 99.5% of the time, which means that 0.5% of the time the test will come back positive for skin cancer even though the patient does not have it"). The participant must estimate the probability of an event based on this information (e.g., the probability that a patient has skin cancer if the test is positive). Neglecting the base rate information (e.g., responding "99.5%") results in a large error with respect to the correct answer (e.g., 67%). The measure was the mean absolute difference between the response and the correct answer. Items and item-level details are included in the Appendix.

Fit statistics and modelling details

For all models, we report the Satorra–Bentler scaled chi-square value (χ^2_{SB}) and ROBUST estimates of model fit (Satorra & Bentler, 1994). ROBUST statistics account for violations of multivariate normality, but cannot be calculated for incomplete data. Although we excluded participants who had missing values on the rationality tests, there were a few missing data points on the other measures (1.21 – 3.93%). These missing values were imputed using an expectation maximization algorithm for the confirmatory factor analyses and structural equation models reported below.

We report three fit indices. The first, χ^2_{SB}, is an absolute fit index gauging the fit of the specified model to the observed covariance matrix. A significant χ^2_{SB} value indicates lack of fit. However, the χ^2_{SB} test statistic is heavily influenced by sample size; in large samples, very minor discrepancies between the data and model can lead to a significant statistic. Thus, on its own, a significant χ^2_{SB} value is insufficient for rejecting a model. We also report the ROBUST comparative fit index (*CFI), which compares the fit of the specified model to a null model in which covariation between measures is restricted to zero. *CFIs are bound between 0 and 1, and large values indicate better fit, with .95 or higher indicating good fit. Finally, we report the ROBUST root mean square error of approximation (*RMSEA), which provides information about the reasonableness with which a given model approximates a population covariance matrix (Byrne, 2013). Thus, the *RMSEA is often interpreted as an index of error, where small values, ideally below .05, indicate a well-fitting model. Models are evaluated based on all three fit statistics. Model comparisons are reported as $\Delta\chi^2_{SB}$ and account for any violations of normality (Satorra & Bentler, 2001). Wald *Z* statistics were calculated with ROBUST standard errors.

Results

Descriptive statistics for the cognitive ability and rationality measures are presented in Table 2. Correlations are presented in Table 3. In general,

Table 2. Descriptive statistics for the rationality and cognitive ability measures.

Measure	*N*	*M*	*SD*	Skewness	Kurtosis	Reliability
Diagnostic base rate neglect	331	41.41	12.36	−0.15	−1.01	.44$^{\alpha}$
Stereotype base rate neglect	331	0.37	0.23	0.91	−0.48	.46$^{\alpha}$
Conjunction fallacy	331	2.33	1.93	0.92	0.01	.69$^{\alpha}$
Wason selection	331	0.74	1.29	2.64	9.19	.66$^{\alpha}$
Raven's matrices	325	8.67	3.70	0.12	−0.92	.81$^{\alpha}$
Number series	318	8.03	3.20	0.31	−0.79	.86$^{\alpha}$
Letter sets	328	14.39	4.82	−0.06	−0.38	.88$^{\alpha}$
Operation span	321	49.10	17.96	−0.31	−0.71	.86^{b}
Symmetry span	326	23.09	10.13	0.30	−0.30	.80^{b}
Rotation span	324	21.01	9.64	0.25	−0.31	.83^{b}
Antisaccade	328	0.78	0.19	−0.97	0.28	.91$^{\alpha}$
Selective visual arrays	327	1.08	1.21	0.50	−0.03	.72$^{\alpha}$
Flanker effect	323	96.34	45.86	1.07	1.59	.82^{b}
Stroop effect	326	153.08	97.74	0.53	0.48	.74^{b}

Note. $^{\alpha}$ = Cronbach's alpha reliability. ^{b} = Spearman-Brown corrected split-half reliability.

Table 3. Correlations among rationality and cognitive ability measures.

Measure	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
1. Diagnostic BRN	–												
2. Stereotype BRN	**.16**	–											
3. Conjunction fallacy	**.17**	**.49**	–										
4. Wason selection	**.12**	**.19**	**.29**	–									
5. Raven's matrices	**.32**	**.33**	**.39**	**.19**	–								
6. Number series	**.28**	**.23**	**.40**	**.16**	**.66**	–							
7. Letter sets	**.27**	**.20**	**.27**	.07	**.56**	**.63**	–						
8. Operation span	**.18**	**.16**	**.26**	**.13**	**.48**	**.53**	**.49**	–					
9. Symmetry span	**.16**	**.20**	**.28**	**.19**	**.54**	**.47**	**.46**	**.57**	–				
10. Rotation span	**.26**	**.17**	**.33**	.10	**.62**	**.51**	**.50**	**.54**	**.69**	–			
11. Antisaccade	**.23**	**.15**	**.24**	.11	**.49**	**.39**	**.36**	**.28**	**.38**	**.45**	–		
12. Visual arrays	**.21**	**.23**	**.32**	**.16**	**.52**	**.47**	**.49**	**.33**	**.41**	**.46**	**.40**	–	
13. Flanker effect	**−.14**	**−.16**	**−.20**	−.08	**−.34**	**−.26**	**−.25**	**−.22**	**−.31**	**−.24**	**−.31**	**−.20**	–
14. Stroop effect	−.04	**−.14**	**−.12**	−.10	**−.24**	**−.21**	**−.23**	**−.18**	**−.23**	**−.25**	**−.17**	**−.12**	**.11**

Note. BRN = Base Rate Neglect. **Bold,** $p < .05$. Correlations computed using pairwise deletion (*ns* range from 311 to 331).

correlations among the rationality and cognitive ability measures were small-to-moderate in size and significant, indicating better performance on the rationality tests by higher-ability participants. Performance on the diagnostic base rate neglect task, however, correlated *opposite the predicted direction* with all measures. That is, participants with better performance on the other rationality and cognitive ability tests gave responses that were numerically further from the correct answer on the diagnostic base rate neglect task, a result which Stanovich and West (1998b) observed as well. Alternative scoring procedures for the diagnostic base rate neglect task yielded analogous patterns of correlations. One interpretation of this result is that it provides evidence against an underlying rationality ability tapped by the tasks, because individuals who were "more rational" on three tasks were "less rational" on a fourth. For this reason, we excluded the diagnostic base rate neglect task as an indicator of rationality in the latent variable analyses that follow, because its inclusion would make scores on the latent rationality factor difficult to interpret. We address this issue further in the Discussion.

The rationality factor

First, we used confirmatory factor analysis to test whether variance common to three rationality measures – stereotype base rate neglect, conjunction fallacy, and Wason selection – could be modeled as a latent factor. As previously noted, the bivariate correlations among rationality measures were small-to-moderate in size and statistically significant, suggesting that they may share variance in common. Strong evidence for an underlying rationality factor would be provided by large and statistically significant

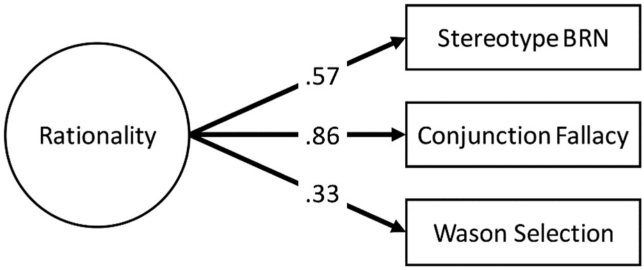

Figure 1. The rationality factor measurement model. Stereotype BRN = stereotype base rate neglect.

factor loadings (>.50), with weaker evidence provided by small-to-moderate loadings (>.35; see Kline, 2015).

The rationality factor is depicted in Figure 1. The stereotype base rate neglect and conjunction fallacy tasks had large standardized loadings (.57 and .86), whereas the Wason selection task had a relatively low loading (.33). All loadings were statistically significant, but their discrepant magnitude suggests that our rationality factor primarily reflects probabilistic reasoning. The rationality factor accounted for 33% of the variance in performance on the stereotype base rate neglect task, 74% of the variance in the conjunction fallacy task, and 11% of the variance in the Wason selection task (Table 4; note that these proportions of variance are calculated by squaring the value of each standardized factor loading). Thus, there is common variance among the rationality tests which was captured by the latent rationality factor, allowing us to examine its relation to general intelligence in subsequent analyses.

Rationality and intelligence

Next, we estimated the relationship between latent factors representing general intelligence and rationality. We specified the general intelligence factor to have loadings on all 10 cognitive ability measures and the rationality factor to have loadings on the three rationality measures. The two latent factors were allowed to correlate. We were particularly interested in the strength of the relationship between these latent factors – the degree to which general intelligence and rationality share reliable variance. We tested whether this latent correlation could be set to 1.0 without loss in model fit. If so, this would provide evidence that rationality and intelligence are not empirically distinct.

The model is depicted in Figure 2. There was a strong and statistically significant correlation between the general intelligence and rationality factors (r = .54). In other words, 29% of the variance was shared among the

Table 4. Standardized factor loadings, significance tests, and coefficients of determination for the rationality measures in the measurement model.

Measure	Factor loading	Wald *Z*	*p*	R^2
Stereotype base rate neglect	.57	6.20	<.001	.33
Conjunction fallacy	.86	8.08	<.001	.74
Wason selection	.33	3.06	.002	.11

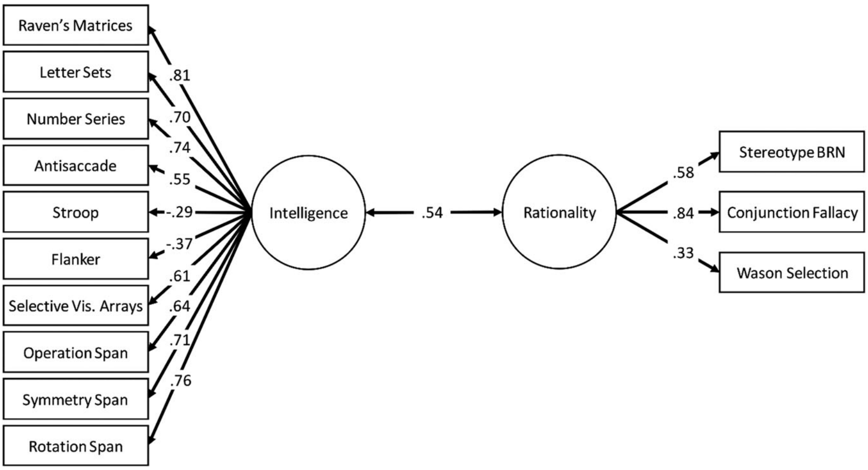

Figure 2. The relationship between latent factors representing general intelligence and rationality. $\chi^2_{SB(64)} = 147.03$, $p < .001$, *CFI =.93, *RMSEA =.06, 90% CI = [.05,.08].

intelligence and rationality latent constructs, a sizeable and practically significant amount. Importantly, however, the correlation fell well short of unity; constraining the correlation between rationality and general intelligence to 1.0 resulted in significantly worse model fit, $\Delta\chi^2_{SB(1)} = 39.68$, $p < .001$, indicating that the rationality and intelligence factors are empirically distinct.

Having established that rationality and general intelligence are correlated but distinct at the latent level, we investigated whether rationality measures still relate to one another after partialling out variance attributable to general intelligence. Years of research have established that performance on different cognitively-demanding tasks correlates positively and significantly (Jensen, 1998); the question addressed by this analysis is whether rationality items share a source of variance above and beyond the *g*-factor. If, after controlling for general intelligence, measures of rationality do not load on their own factor, this would provide evidence that rationality is little more than *g* plus task-specific variance, unshared across rationality tests. If, on the other hand, measures of rationality still load on their own factor after controlling for intelligence, this would suggest that there is

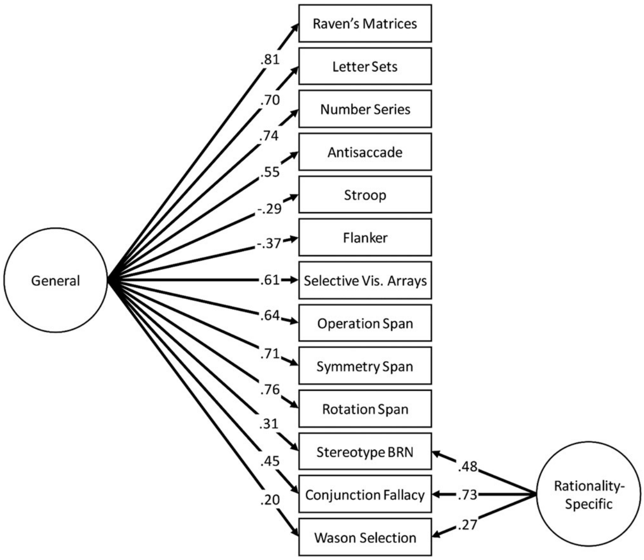

Figure 3. Bifactor model with latent factors representing variance common to all measures and specific to rationality measures. $\chi^2_{SB(64)} = 146.46$ $p < .001$, *CFI =.93, *RMSEA =.07, 90% CI = [.05,.08].

something shared among rationality tests which is "missed" by other intelligence tests.

We analyzed a bifactor model to address these possibilities. A general factor was specified to load on all measures, including the 10 cognitive ability measures and the three rationality measures. A rationality-specific factor had loadings only on the three rationality measures. The bifactor model is depicted in Figure 3. Critically, the rationality tests loaded significantly on the rationality-specific factor even after accounting for variance attributable to the general factor (Table 5). The standardized loadings of the rationality measures on the rationality-specific factor were slightly lower than those reported in the one-factor model in Figure 1 (compare .57 to .48 for stereotype base rate neglect; .86 to .73 for conjunction fallacy; and .33 to .27 for Wason selection). Nevertheless, for each of the three rationality measures, their loadings on the rationality-specific factor were numerically larger than their loadings on the general factor (avg. difference in magnitude = .17), which were also statistically significant. Taken together, the results indicate that the relationships among rationality measures cannot be attributed solely to their mutual dependence on cognitive ability.

Rationality, working memory capacity, fluid intelligence, and attention control

In our next set of analyses, we tested whether the relationship between working memory capacity and rationality could be attributed to attention

Table 5. Standardized factor loadings, significance tests, and coefficients of determination for the rationality measures in the bifactor model.

	General factor			Rationality-specific factor			
Measure	Loading	Z	p	Loading	Z	p	R^2
Stereotype base rate neglect	.31	5.43	$<$ .001	.48	4.66	$<$ .001	.33
Conjunction fallacy	.45	7.48	$<$ .001	.73	5.41	$<$ .001	.74
Wason selection	.20	2.57	.010	.27	2.82	.005	.11

control, and whether the relationship between fluid intelligence and rationality could be attributed to attention control. First, we estimated the relationships between latent factors representing rationality, working memory capacity, fluid intelligence, and attention control. There was an unusually strong correlation between fluid intelligence and attention control ($r = .93$), which was not significantly different than 1.0, $\Delta\chi^2_{SB(1)} = 1.66$, $p = .20$. This indicates that fluid intelligence and attention control were not statistically dissociable at the latent level. Multicollinearity is a common problem in individual differences research on intelligence (see e.g., Conway et al., 2002); overly strong correlations between variables sometimes necessitate modifying a statistical model. In the present case, multicollinearity posed an issue for interpreting the unique contribution of attention control or fluid intelligence to rationality, because the same conclusions could apply to either cognitive construct. Therefore, to statistically dissociate fluid intelligence from attention control, we omitted selective visual arrays as an indicator of attention control. We elected to omit selective visual arrays because of the four attention control measures, it had the highest average correlation with the fluid intelligence measures. Thus, removing visual arrays from the model should help to dissociate attention control from fluid intelligence. This attention control factor, with antisaccade, Stroop, and flanker measures as indicators, demonstrated acceptable model fit, $\chi^2_{SB(48)} = 79.27$, $p = .003$, *CFI = .97, *RMSEA = .05, 90% CI = [.03, .06]. Furthermore, attention control and fluid intelligence were statistically dissociable ($r = .83$); setting their correlation to 1.0 significantly reduced model fit, $\Delta\chi^2_{SB(1)} = 4.20$, $p = .04$.

Next, we ran a correlated-factors model with latent factors representing rationality, fluid intelligence, working memory capacity, and attention control. As shown in Figure 4, rationality correlated significantly with fluid intelligence ($r = .56$, $p < .001$), working memory capacity ($r = .44$, $p < .001$), and attention control ($r = .49$, $p < .001$). As expected, working memory capacity and attention control were strongly correlated ($r = .79$, $p < .001$), indicating that the two constructs shared 62% of their reliable variance. Fluid intelligence and attention control were also strongly correlated ($r = .85$, $p < .001$); the two constructs shared 72% of their variance.

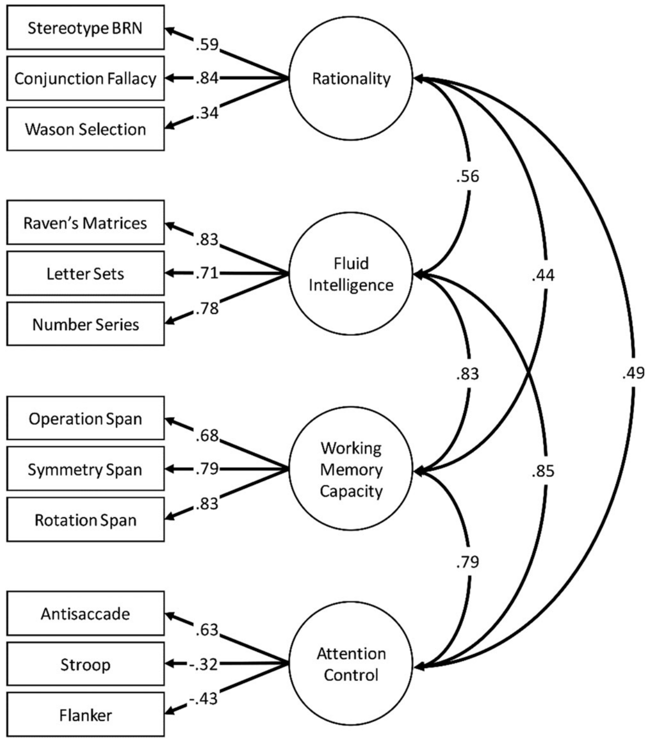

Figure 4. Relationships between rationality, fluid intelligence, working memory capacity, and attention control. $\chi^2_{SB(48)} = 79.27$, $p = .003$, *CFI =.97, *RMSEA =.05, 90% CI = [.03,.06].

We then used structural equation modeling to test whether attention control accounted for the relationship between working memory capacity and rationality. We specified a model in which attention control predicted working memory capacity and rationality. In this model, the error terms of working memory capacity and rationality represent the variance in each latent factor not accounted for by attention control. These error terms were allowed to correlate. If attention control accounts for the relationship between working memory capacity and rationality, then the residual correlation between working memory capacity and rationality should not be significantly greater than zero after accounting for attention control.

The model is depicted in Figure 5 (see Table 6). Attention control fully accounted for the relationship between working memory capacity and rationality. That is, attention control significantly predicted working memory capacity (std. coefficient = .79, $p < .001$) and rationality (std. coefficient = .48, $p < .001$). The residual correlation between working memory capacity and rationality was not statistically significant ($r = .10$, $p = .448$). Thus, after accounting for attention control, the once-strong relationship between

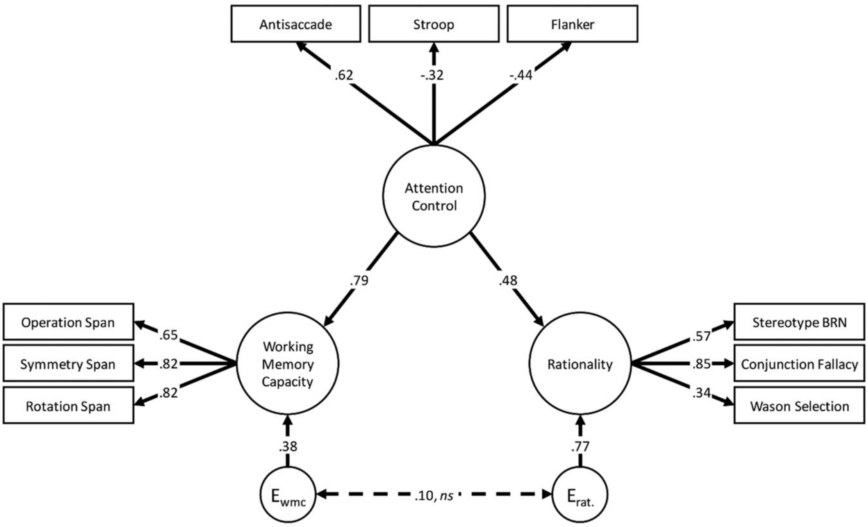

Figure 5. Attention control fully accounted for the relationship between working memory capacity and rationality. $\chi^2_{SB(24)} = 25.99$, $p = .354$, *CFI = 1.00, *RMSEA =.02, 90% CI = [.00,.05].

Table 6. Parameter estimates from the attention control, working memory capacity, and rationality model.

Effect	Std. Coefficient	Wald Z	p
Attention control to working memory capacity	.79	4.21	< .001
Attention control to rationality	.48	4.85	< .001
Residual correlation between WMC and rationality	.10	.76	.448

Note. WMC = working memory capacity.

working memory capacity and rationality ($r = .44$; see Figure 4 above) was no longer significantly different from zero. This indicates that the relationship between working memory capacity and rationality can be explained by variance in each construct attributable to attention control.

As our final analysis, we tested whether attention control accounted for the relationship between fluid intelligence and rationality. We specified a model in which attention control predicted fluid intelligence and rationality. The error terms of fluid intelligence and rationality, which represent the variance in each factor not accounted for by attention control, were allowed to correlate. As shown in Figure 6 and Table 7, attention control partially accounted for the relationship between fluid intelligence and rationality. Attention control significantly predicted fluid intelligence (std. coefficient = .84, $p < .001$) and rationality (std. coefficient = .48, $p < .001$). The residual correlation between fluid intelligence and rationality was statistically significant ($r = .32$, $p = .011$). However, partialling out the variance in each factor attributable to attention control significantly reduced the

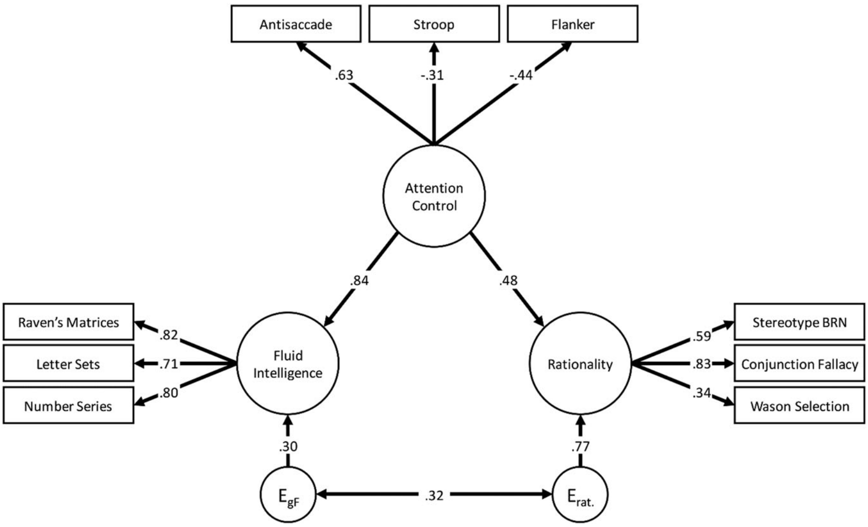

Figure 6. Attention control partially accounted for the relationship between fluid intelligence and rationality. $\chi^2_{SB(24)} = 32.31$, $p = .12$, *CFI $= .99$, *RMSEA $= .03$, 90% CI $= [.00, .06]$.

Table 7. Parameter estimates from the attention control, fluid intelligence, and rationality model.

Effect	Std. Coefficient	Wald *Z*	*p*
Attention control to fluid intelligence	.84	3.71	$< .001$
Attention control to rationality	.48	5.05	$< .001$
Residual correlation between gF and rationality	.32	2.53	.011

Note. gF = fluid intelligence.

relationship between fluid intelligence and rationality (compare $r = .56$ from Figure 4 to $r = .32$ in Figure 6). Indeed, constraining the residual correlation between fluid intelligence and rationality to .56 resulted in significantly worse model fit, $\Delta\chi^2_{SB(1)} = 5.15$, $p = .02$, suggesting that attention control accounted for a significant portion of the covariance between the two constructs, but not all of it.

Discussion

In the present study, we examined the relationship between rationality and intelligence in a sample of 331 participants representing a wide range of educational backgrounds and ability. Over multiple two-hour sessions, participants completed 14 tasks designed to measure rationality, fluid intelligence, working memory capacity, and attention control. We used confirmatory factor analysis and structural equation modeling to estimate relationships between rationality measures and broad cognitive abilities.

First, we tested whether a common ability underpinned performance on the rationality tests. We found significant, small-to-moderate correlations among the four rationality measures (avg. $r = .24$). One measure, however – diagnostic base rate neglect – correlated opposite the predicted direction, suggesting that more rational participants performed worse on this task. This was particularly surprising, because a different measure of base rate neglect (i.e., stereotype base rate neglect) correlated in the expected direction. Setting aside diagnostic base rate neglect, the remaining three measures loaded significantly on a latent rationality factor (loadings ranged from .33 to .86), which primarily reflected probabilistic reasoning. Thus, the evidence suggests that there may be a common rationality ability underpinning performance on *some* rationality tests, but that not all rationality measures equally reflect the construct of interest, or necessarily cohere in the task battery.

Next, we tested whether rationality and intelligence were empirically distinct. Latent variables representing general intelligence and rationality correlated strongly and significantly ($r = .54$), but the correlation fell significantly short of unity, indicating that the two constructs were dissociable. Indeed, more than half the variance in rationality was unaccounted for by general intelligence. We also found that measures of rationality loaded significantly on a rationality-specific factor after partialling out variance in task performance attributable to general intelligence. This result runs counter to an argument that rationality measures correlate because they tap general intelligence, but share little else in common. To the contrary, the loadings of the rationality measures on the rationality-specific factor were only slightly lower after controlling for general intelligence, suggesting that there is reliable variance shared among rationality tests that is independent of general intelligence.

Our final set of analyses assessed the contributions of latent cognitive ability factors to individual differences in rationality. Rationality correlated significantly with fluid intelligence ($r = .56$), working memory capacity ($r = .44$), and attention control ($r =. 49$). Attention control fully accounted for the relationship between working memory capacity and rationality. That is, after accounting for attention control, the residual relationship between working memory capacity and rationality was no longer statistically significant ($r = .10$, $p = .448$). This result is consistent with a growing body of evidence suggesting that the primary reason working memory capacity predicts higher-order cognitive functions and real-world outcomes is because working memory measures tap attention control (see Burgoyne & Engle, 2020; Engle, 2018). Our view is that attention supports the maintenance and manipulation of information in the service of rational thought, a role that has historically been attributed to working memory capacity

(Engle, 2018). To our knowledge, this is the first study to demonstrate that attention control underpins the working memory capacity-rationality relationship, and should be corroborated by further research.

Attention control also partially accounted for the relationship between fluid intelligence and rationality. After accounting for attention control, the residual relationship between fluid intelligence and rationality was significantly reduced ($r = .32$ as opposed to $r = .56$). In our theoretical framework linking attention control to rationality, attention control is responsible for inhibiting the first response that comes to mind, which often serves as a lure or foil in rationality tests. As disengagement from faulty hypotheses is also theoretically tapped by tests of fluid intelligence (Burgoyne & Engle, 2020), this suggests that one mechanism by which fluid intelligence contributes to rationality may be subsumed by attention control. However, fluid intelligence remained a significant predictor of rationality even after accounting for attention control. Although only speculative, it is possible that novel hypothesis generation is one cognitive process that contributes to rationality that is not subsumed by attention control.

Broader impact

The overall picture to emerge from this research is that individual differences in rationality are strongly related to general intelligence, but even after accounting for general intelligence, substantial variance in rationality remains unexplained. The correlation of $r = .54$ between latent factors representing general intelligence and rational thinking in the present study is broadly consistent with studies that examined this relationship at the observed level (i.e., without the use of latent variables). For instance, Bruine de Bruin et al. (2007) found that scores on the Decision-Making Competence inventory correlated $r = .61$ with performance on Raven's Matrices and $r = .50$ with reading comprehension test performance. Similarly, Stanovich and West (1998b) found a correlation of $r = .53$ between a rational thinking composite and SAT scores, and Stanovich et al. (2016) report correlations of around .50 between cognitive ability and the probabilistic reasoning and statistical reasoning subtests of the Comprehensive Assessment of Rational Thinking. Generally speaking, our results corroborate Stanovich and colleagues' (Stanovich, 2009; Stanovich et al., 2016) claim that rationality tests capture something that other cognitive ability tests miss.

What else might contribute to individual differences in rationality? One possibility is *thinking dispositions* – the extent to which people are inclined to deliberate carefully about problems rather than engaging in miserly cognitive processing (i.e., providing the first answer that springs to mind). Thinking dispositions may serve to dissociate rationality from fluid intelligence, because unlike rationality tests, fluid intelligence tests typically do not evoke intuitive yet incorrect answers. Research has shown that thinking dispositions can explain variance in rationality above and beyond measures of cognitive ability. For instance, West et al. (2008) found that a personality composite variable representing thinking dispositions accounted for 3% of the variance in rationality after accounting for SAT scores. Similarly, Stanovich and West (1997) found that actively open-minded thinking dispositions predicted students' evaluations of the quality of different arguments concerning real-life situations above and beyond measures of cognitive ability. In future work, we plan to measure problem-solvers' thinking dispositions in addition to their cognitive abilities, to test whether non-ability factors contribute to rationality above and beyond cognitive ability measures such as attention control. Our prediction is that attention control may account for some of the variance in rationality tests that is captured by thinking dispositions, because attention control theoretically supports the ability to inhibit automatic responses.

At a more mechanistic level, research by Kleitman and colleagues points to metacognitive factors that may contribute to individual differences in thinking dispositions and help explain why some people are more susceptible to faulty reasoning: monitoring confidence and control thresholds (Jackson et al., 2016, but see also Jackson et al., 2017 and Kleitman et al., 2019). The first factor, *monitoring confidence*, refers to 1) "monitoring," the process by which evidence accumulates in a decision-making model, and 2) "confidence," how much weight is given to evidence as it accumulates. People with a high degree of monitoring confidence weight evidence more heavily, such that, all else being equal, a person with higher monitoring confidence will reach a decision-making threshold before a person with lower monitoring confidence. Control thresholds, on the other hand, refer to how much evidence must be accumulated before a decision is made. A person with a lower control threshold requires less evidence to trigger a decision than a person with a higher control threshold.

Using structural equation modeling, Jackson et al. (2016) found that individuals with greater fluid intelligence, lower monitoring confidence, and higher control thresholds tended to perform better on reasoning tests. What's more, these three factors fully accounted for the covariance between the reasoning tests, suggesting that the reason rationality tests correlate positively with one another is because they mutually depend on fluid intelligence and these metacognitive factors. Importantly, monitoring confidence and control thresholds were only moderately correlated with one another ($r = .31$), suggesting that they are largely distinct cognitive constructs. It remains an open question worthy of future investigation whether attention control would be more closely related to monitoring confidence or control thresholds, and if so, whether attention control might account for variance in reasoning performance that is shared with one or both of these constructs.

Yet another factor that may set rationality apart from other broad cognitive abilities is acquired problem-solving strategies and rules, particularly those pertaining to probabilistic reasoning and logic. Sometimes referred to as *mindware* (Perkins, 1995), these procedures serve as tools for the reasoner that, when coupled with effortful engagement and sufficient computational ability, result in accurate responses. For example, in a probabilistic reasoning task, the likelihood of two independent events co-occurring can be computed as the product of their individual probabilities. That said, because mindware is acquired knowledge, one might expect individual differences in the acquisition or application of mindware to be related to individual differences in cognitive ability. Nevertheless, the problem-solving strategies demanded by the rationality tests in the present study were unshared with the other cognitive ability tests, providing a means by which the constructs might have dissociated. In a similar vein, it is possible that because the rationality tests shared method-specific variance, this might have biased the rationality measures to load more strongly on a common factor. This concern could be mitigated by including rationality tests with different question and response formats in future studies.

Taken together, our results shed new light on the contribution of attention control to individual differences in rational thinking, with implications for both theory and practice. From a theoretical standpoint, the observation that attention control fully explained working memory capacity's relationship with rationality has bearing on dual process theories of reasoning. According to dual process theory, the deliberate and effortful processing that is characteristic of Type 2 reasoning requires the engagement of attention control and working memory (Evans, 2008). Our results indicate that attention control may be more fundamental than working memory with respect to Type 2 processing, although this finding warrants replication. From a practical standpoint, attention control tests may be instrumental in assessing individual differences in the ability to engage in rational thinking, and, if the current results hold in future research, could serve as a theoretically defensible substitute for working memory capacity tests down the road. This would be consistent with recent research suggesting that much of working memory capacity's predictive utility can be attributed to attention control (see e.g., Burgoyne & Engle, 2020).

Limitations

It must be noted that, as is the case with all psychological research, our conclusions are limited by the tasks administered to participants. In particular, our rationality test battery did not tap all the heuristics and biases in the literature. Granted, there are more than 30 heuristics and biases that we could have studied (Stanovich, 2016), and participants completed eight hours of testing for the present data collection effort. Nevertheless, it might be argued that a more comprehensive assessment of rational thinking would yield different results. While this is possible, the rationality tests we administered to participants are frequently used and cover important aspects of the construct, including probabilistic reasoning, confirmation bias, and scientific thinking. Furthermore, Stanovich et al. (2016) found a strong correlation ($r = .695$) between scores on the Comprehensive Assessment of Rational Thinking (which contains 20 subtests) and a limited cognitive ability test comprised of antonyms, a word checklist, and analogies. Although our conclusions are broadly consistent with Stanovich et al. (2016), in future work we plan to use a more comprehensive battery of rational thinking items to test whether the relative contributions of broad cognitive abilities to rationality differ according to the heuristics or biases tapped by the tasks.

Taking a closer look at the rationality tests we administered to participants can also be instructive to researchers interested in studying individual differences in rationality. In the following paragraphs, we detail some of our observations and those made by a helpful reviewer with regard to three of the four rationality tasks that were administered to participants: the Wason Card Selection task, the Diagnostic Base Rate Neglect Task, and the Stereotype Base Rate Neglect Task.

First, our version of the Wason card selection task instructed participants to select two cards, whereas in the traditional version of the task, participants can select as many cards as they think are needed to determine a rule's veracity. Whereas the two-card selection method results in a dichotomous accuracy measure (i.e., correct or incorrect), continuous accuracy measures and strategy profiles can be computed when participants are told to select as many cards as they think are necessary (see Pollard & Evans, 1987; Stanovich & West, 1998b). Doing so would have provided more nuance to our understanding of how participants completed the Wason task and could lead to larger observed correlations by maximizing between-subjects variance.

Furthermore, the accuracy rates on the Wason selection task were fairly low, particularly for the abstract (i.e., nondeontic) items (average accuracy = 6%; see Table A1). Restriction of range due to floor affects attenuates relationships between measures, and may partly explain the relatively low loading of the Wason selection task on the rationality factor. Providing *violation instructions* (i.e., telling participants to flip cards to determine "whether or not the rule is being violated" rather than "to test whether the rule is true or false") could have helped reduce this floor effect by increasing overall accuracy rates (see e.g., Stanovich & West, 1998b). Another manipulation that may have increased accuracy rates, particularly on the less abstract (i.e., deontic) items, is the inclusion of contextualizing scenarios to frame the items. For example, Stanovich and West (1998b) found that including a scenario in which participants were asked to play the role of a police officer (e.g., "Imagine that you are a police officer on duty, walking through a local bar. It is your job to ensure that the drinking laws are in effect in this bar … .") raised overall accuracy rates on the drinking-age Wason selection item by 50%. Thus, researchers should be aware that changes to the format of the Wason selection test can result in substantial differences in accuracy rates and potential differences in their relation to cognitive abilities.

Turning to the diagnostic base rate neglect task, we found that scores correlated opposite the predicted direction with all other measures, indicating that more intelligent or rational participants did worse on this task. To reiterate, a similar pattern of results was observed by Stanovich and West (1998b). One potential explanation for this result is that the items did not always provide the false positive rate (e.g., the Face Recognition item, see the Appendix), or in one case provided the false positive rate in a manner such that it was equated to one minus the hit rate (e.g., the Skin Cancer item), which could potentially confuse participants. As these items are difficult to construct in a way that is understandable to participants, other measures of rational thinking may be preferable to the diagnostic base rate neglect task.

Finally, an analysis of the stereotype base rate neglect task suggests that due to variability in the likelihood ratio that participants assign to stereotypical vs. non-stereotypical outcomes, reasonable participants could disagree about the normatively correct response to some of the items. For example, consider the Kurt item (see the Appendix for the full description). One might estimate that there is a 30% chance that Kurt lives in a condo and a 0.01% chance that he lives in a farmhouse given that he works long hours on Wall Street, wears Armani suits to work, likes wearing shades, and is single. This likelihood ratio (300/1) is almost exactly the inverse of the base rate provided by the problem (approximately 1/332), such that the posterior odds are nearly 1 (i.e., both possibilities are almost equally likely under these conditions). From this analysis, we conclude that it is difficult to construct stereotype base rate neglect items that evoke heuristics and biases via stereotype priming without sometimes rendering the normatively correct response a toss-up between the two response options. This could explain why accuracy rates on this task were low in our study and in De Neys and Glumicic (2008) study, which used the same task stimuli.

Conclusion

We tested the relationship between rationality and general intelligence at the latent level in a sample of 331 participants. Rationality and intelligence were strongly correlated, but substantial variance in rationality remained unexplained, indicating that the two constructs were empirically distinct. Novel to the present study, we found that attention control fully accounted for the relationship between working memory capacity and rationality, and partially accounted for the relationship between fluid intelligence and rationality. Although more research is warranted on the topic, our results suggest that the abilities underlying performance on tests of rationality and cognitive ability are correlated yet distinct.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by grants from the Office of Naval Research (N00173-18-S-BA01; N00014-12-1-1011; N00014-17-1-2061) to Randall W. Engle.

ORCID

Alexander P. Burgoyne http://orcid.org/0000-0002-2651-3782

References

Baddeley, A. (1992). Working memory. *Science (New York, N.Y.)*, *255*(5044), 556–559. https://doi.org/10.1126/science.1736359

Bar-Hillel, M. (1977). *The base-rate fallacy in probability judgments*. Decisions and Designs, Inc.

Barrett, L. F., Tugade, M. M., & Engle, R. W. (2004). Individual differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin*, *130*(4), 553–573.

Basile, A. G., & Toplak, M. E. (2015). Four converging measures of temporal discounting and their relationships with intelligence, executive functions, thinking dispositions, and behavioral outcomes. *Frontiers in Psychology*, *6*, 728.

Bruine de Bruin, W., Parker, A. M., & Fischhoff, B. (2007). Individual differences in adult decision-making competence. *Journal of Personality and Social Psychology*, *92*(5), 938–956. https://doi.org/10.1037/0022-3514.92.5.938 17484614

Bruine de Bruin, W., Parker, A. M., & Fischhoff, B. (2007). Individual differences in adult decision-making competence. *Journal of Personality and Social Psychology*, *92*(5), 938–956.

Bruine de Bruin, W., Parker, A. M., & Fischhoff, B. (2020). Decision-making competence: More than intelligence? *Current Directions in Psychological Science*, *29*(2), 186–192. https://doi.org/10.1177/0963721420901592

Burgoyne, A. P., & Engle, R. W. (2020). Attention control: A cornerstone of higher-order cognition. *Current Directions in Psychological Science*, *29*(6), 624–630. https://doi.org/10.1177/0963721420969371

Byrne, B. M. (2013). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. Routledge.

Campitelli, G., & Gerrans, P. (2014). Does the cognitive reflection test measure cognitive reflection? A mathematical modeling approach. *Memory & Cognition*, *42*(3), 434–447. https://doi.org/10.3758/s13421-013-0367-9

Conway, A. R., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*(2), 163–183. https://doi.org/10.1016/S0160-2896(01)00096-4

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, *51*(1), 42–100.

De Neys, W. (2006a). Automatic-heuristic and executive-analytic processing during reasoning: Chronometric and dual-task considerations. *Quarterly Journal of Experimental Psychology (2006)*, *59*(6), 1070–1100. https://doi.org/10.1080/02724980543000123

De Neys, W. (2006b). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science*, *17*(5), 428–433. https://doi.org/10.1111/j.1467-9280.2006.01723.x

De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. *Cognition*, *106*(3), 1248–1299.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2005). Working memory and everyday conditional reasoning: Retrieval and inhibition of stored counterexamples. *Thinking & Reasoning*, *11*(4), 349–381. https://doi.org/10.1080/13546780442000222

Del Missier, F., Mäntylä, T., & Bruine de Bruin, W. (2010). Executive functions in decision making: An individual differences approach. *Thinking & Reasoning*, *16*(2), 69–97. https://doi.org/10.1080/13546781003630117

Del Missier, F., Mäntylä, T., & De Bruin, W. B. (2012). Decision-making competence, executive functioning, and general cognitive abilities. *Journal of Behavioral Decision Making*, *25*(4), 331–351. https://doi.org/10.1002/bdm.731

Draheim, C., Harrison, T. L., Embretson, S. E., & Engle, R. W. (2018). What item response theory can tell us about the complex span tasks. *Psychological Assessment*, *30*(1), 116–129. https://doi.org/10.1037/pas0000444

Draheim, C., Tsukahara, J. S., Martin, J. D., Mashburn, C. A., & Engle, R. W. (2020). *A toolbox approach to improving the measurement of attention control*. General.

Ekstrom, R. B. (1976). *Kit of factor-referenced cognitive tests*. Educational Testing Service.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*(1), 19–23. https://doi.org/10.1111/1467-8721.00160

Engle, R. W. (2018). Working memory and executive attention: A revisit. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *13*(2), 190–193. https://doi.org/10.1177/1745691617720478

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of Experimental Psychology. General*, *128*(3), 309–331.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. https://doi.org/10.3758/BF03203267

Evans, J. S. B. (2003). In two minds: dual-process accounts of reasoning. *Trends in Cognitive Sciences*, *7*(10), 454–459. https://doi.org/10.1016/j.tics.2003.08.012
Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*, 255–278. https://doi.org/10.1146/annurev.psych.59.103006.093629
Evans, J. S. B., & Over, D. (1996). Rationality in the selection task: epistemic utility versus uncertainty reduction. *Psychological Review*, *103*(2), 356–363. https://doi.org/10.1037/0033-295X.103.2.356
Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *8*(3), 223–241.
Frankish, K. (2010). Dual-process and dual-system theories of reasoning. *Philosophy Compass*, *5*(10), 914–926. https://doi.org/10.1111/j.1747-9991.2010.00330.x
Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732
Gigerenzer, G. (1996). Rationality: Why social context matters. In P. B. Baltes & U. M. Staudinger (Eds.), *Interactive minds: Life-span perspectives on the social foundation of cognition* (pp. 319–346). Cambridge University Press.
Hallett, P. E. (1978). Primary and secondary saccades to goals defined by instructions. *Vision Research*, *18*(10), 1279–1296. https://doi.org/10.1016/0042-6989(78)90218-3
Handley, S. J., Capon, A., Beveridge, M., Dennis, I., & Evans, J. S. B. (2004). Working memory, inhibitory control and the development of children's reasoning. *Thinking & Reasoning*, *10*(2), 175–195. https://doi.org/10.1080/13546780442000051
Hutchison, K. A. (2007). Attentional control and the relatedness proportion effect in semantic priming. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *33*(4), 645–662.
Jackson, S. A., Kleitman, S., Howie, P., & Stankov, L. (2016). Cognitive abilities, monitoring confidence, and control thresholds explain individual differences in heuristics and biases. *Frontiers in Psychology*, *7*, 1559.
Jackson, S. A., Kleitman, S., Stankov, L., & Howie, P. (2017). Individual differences in decision making depend on cognitive abilities, monitoring and control. *Journal of Behavioral Decision Making*, *30*(2), 209–223. https://doi.org/10.1002/bdm.1939
Jensen, A. R. (1998). *The g factor: The science of mental ability* (Vol. 648). Praeger.
Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review*, *9*(4), 637–671. https://doi.org/10.3758/bf03196323
Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology. General*, *133*(2), 189–217.
Kelly, T. (2003). Epistemic rationality as instrumental rationality: A critique. *Philosophy and Phenomenological Research*, *66*(3), 612–640. https://doi.org/10.1111/j.1933-1592.2003.tb00281.x
Kleitman, S., Hui, J. S. W., & Jiang, Y. (2019). Confidence to spare: Individual differences in cognitive and metacognitive arrogance and competence. *Metacognition and Learning*, *14*(3), 479–508. https://doi.org/10.1007/s11409-019-09210-x
Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.

Leslie, A. M. (1987). Pretense and representation: The origins of theory of mind. *Psychological Review*, *94*(4), 412–426. https://doi.org/10.1037/0033-295X.94.4.412
Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281. https://doi.org/10.1038/36846
Markovits, H., & Doyon, C. (2004). Information processing and reasoning with premises that are empirically false: Interference, working memory, and processing speed. *Memory & Cognition*, *32*(4), 592–601.
Martin, J., Tsukahara, J. S., Draheim, C., Shipstead, Z., Mashburn, C., Vogel, E. K., Engle, R. W. (2019). The visual arrays task … Visual storage capacity or attention control? https://doi.org/10.31234/osf.io/u92cm
McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*(1), 1–10. https://doi.org/10.1016/j.intell.2008.08.004
Parker, A. M., Bruine de Bruin, W., & Fischhoff, B. (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., & Fischhoff, B. (2005). Decision-making competence: External validation through an individual-differences approach. *Journal of Behavioral Decision Making*, *18*(1), 1–27. https://doi.org/10.1002/bdm.481
Pennycook, G., Cheyne, J. A., Koehler, D. J., & Fugelsang, J. A. (2020). On the belief that beliefs should change according to evidence: Implications for conspiratorial, moral, paranormal, political, religious, and science beliefs. *Judgment and Decision Making*, *15*, 476–498.
Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015). Everyday consequences of analytic thinking. *Current Directions in Psychological Science*, *24*(6), 425–432. https://doi.org/10.1177/0963721415604610
Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, *188*, 39–50. https://doi.org/10.1016/j.cognition.2018.06.011
Perkins, D. N. (1995). *Outsmarting IQ: The emerging science of learnable intelligence*. Free Press.
Plantinga, A. (1993). *Warrant and proper function*. Oxford University Press.
Pollard, P., & Evans, J. S. B. (1987). Content and context effects in reasoning. *The American Journal of Psychology*, *100*(1), 41–60. https://doi.org/10.2307/1422641
Raven, J., & Court, J. H. (1998). *Manual for Raven's progressive matrices and vocabulary scales*. Oxford: Oxford pyschologists Press.
Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Sage Publications, Inc.
Satorra, A., & Bentler, P. M. (2001). A scaled difference Chi-square test statistic for moment structure analysis. *Psychometrika*, *66*(4), 507–514. [Database] https://doi.org/10.1007/BF02296192
Shipley, W. C., Christian, T. A., Martin, A. M., & Klein, A. M. (2009). *Shipley-2*. Western Psychological Services.
Shipstead, Z., Harrison, T. L., & Engle, R. W. (2016). Working memory capacity and fluid intelligence: Maintenance and disengagement. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *11*(6), 771–799.

Shipstead, Z., Lindsey, D. R., Marshall, R. L., & Engle, R. W. (2014). The mechanisms of working memory capacity: Primary memory, secondary memory, and attention control. *Journal of Memory and Language*, *72*, 116–141. https://doi.org/10.1016/j.jml.2014.01.004

Stanovich, K. E. (2009). *What intelligence tests miss: The psychology of rational thought*. Yale University Press.

Stanovich, K. E. (2011). Normative models in psychology are here to stay. *Behavioral and Brain Sciences*, *34*(5), 268–269. https://doi.org/10.1017/S0140525X11000161

Stanovich, K. E. (2012). On the distinction between rationality and intelligence: Implications for understanding individual differences in reasoning. *The Oxford Handbook of Thinking and Reasoning*, *1*, 343–365.

Stanovich, K. E. (2016). The comprehensive assessment of rational thinking. *Educational Psychologist*, *51*(1), 23–34. https://doi.org/10.1080/00461520.2015.1125787

Stanovich, K. E., & Toplak, M. E. (2012). Defining features versus incidental correlates of Type 1 and Type 2 processing. *Mind & Society*, *11*(1), 3–13. https://doi.org/10.1007/s11299-011-0093-6

Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology*, *89*(2), 342–357. https://doi.org/10.1037/0022-0663.89.2.342

Stanovich, K. E., & West, R. F. (1998a). Cognitive ability and variation in selection task performance. *Thinking & Reasoning*, *4*(3), 193–230. https://doi.org/10.1080/135467898394139

Stanovich, K. E., & West, R. F. (1998b). Individual differences in rational thought. *Journal of Experimental Psychology: General*, *127*(2), 161–188. https://doi.org/10.1037/0096-3445.127.2.161

Stanovich, K. E., & West, R. F. (2007). Natural myside bias is independent of cognitive ability. *Thinking & Reasoning*, *13*, 225–247.

Stanovich, K. E., West, R. F., & Toplak, M. E. (2011). The complexity of developmental predictions from dual process models. *Developmental Review*, *31*(2–3), 103–118. https://doi.org/10.1016/j.dr.2011.07.003

Stanovich, K. E., West, R. F., & Toplak, M. E. (2016). *The rationality quotient: Toward a test of rational thinking*. MIT Press.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Teovanović, P., Knežević, G., & Stankov, L. (2015). Individual differences in cognitive biases: Evidence against one-factor theory of rationality. *Intelligence*, *50*, 75–86. https://doi.org/10.1016/j.intell.2015.02.008

Thompson, V. A., Pennycook, G., Trippas, D., & Evans, J. S. B. T. (2018). Do smart people have better intuitions? *Journal of Experimental Psychology. General*, *147*(7), 945–961. https://doi.org/10.1037/xge0000457

Thurstone, L. L. (1938). Primary mental abilities. *Psychometric Monographs*, *1*, ix–121.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The cognitive reflection test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, *39*(7), 1275–1289. https://doi.org/10.3758/s13421-011-0104-1

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014a). Rational thinking and cognitive sophistication: Development, cognitive abilities, and thinking dispositions. *Developmental Psychology*, *50*(4), 1037–1048. https://doi.org/10.1037/a0034910

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014b). Assessing miserly information processing: An expansion of the cognitive reflection test. *Thinking & Reasoning*, *20*(2), 147–168. https://doi.org/10.1080/13546783.2013.844729

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science (New York, N.Y.)*, *185*(4157), 1124–1131.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505.

Unsworth, N., Schrock, J. C., & Engle, R. W. (2004). Working memory capacity and the antisaccade task: Individual differences in voluntary saccade control. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *30*(6), 1302–1321.

Wason, P. C. (1968). Reasoning about a rule. *The Quarterly Journal of Experimental Psychology*, *20*(3), 273–281. https://doi.org/10.1080/14640746808400161

Weaver, E. A., & Stewart, T. R. (2012). Dimensions of judgment: Factor analysis of individual differences. *Journal of Behavioral Decision Making*, *25*(4), 402–413. https://doi.org/10.1002/bdm.748

West, R. F., Toplak, M. E., & Stanovich, K. E. (2008). Heuristics and biases as measures of critical thinking: Associations with cognitive ability and thinking dispositions. *Journal of Educational Psychology*, *100*(4), 930–941. https://doi.org/10.1037/a0012842

Appendix

Wason Selection Task

Instructions: In this task you will be shown four cards with a rule beneath them. Each card has two sides, but you will only see one of them. Your job is to click the two cards you need to turn over in order to decide whether the rule is true or false.

Item ID: A/Even
Scenario: Suppose each card below has a letter on one side and a number on the other.
Rule: If a card has an A on one side, then it has an even number on the other side.
Face Up Cards: A; D; 4; 7
Correct Answer: A; 7

Item ID: Color/Number
Scenario: Suppose each card below has a color on one side and a number on the other. Please click the two cards you need to turn over in order to verify the following rule.
Rule: If a card has a number greater than 6 on one side, it must be blue on the other side.
Face Up Cards: 5; 7; Blue; Red
Correct Answer: 7; Red

Item ID: Black/White
Scenario: Suppose each card below has a black shape on one side and a white shape on the other.
Rule: If a card has a white triangle on one side, it must have a black square on the other side.

Table A1. Descriptive statistics for the rationality tasks ($N = 331$).

Item ID	Type	% of Correct response
Wason Selection		
A / Even	Abstract	5%
Black / White	Abstract	7%
Color / Number	Abstract	8%
Face / Number	Abstract	5%
Color / Circle	Abstract	5%
Arrow / Number	Abstract	8%
Letter / Number	Abstract	4%
Drink / Age	Deontic	12%
Seal / Stamp	Deontic	15%
Decision / Height	Deontic	6%
Conjunction Fallacy		
Linda	N/A	25%
Bill	N/A	28%
Colored Die	N/A	25%
Scandinavian	N/A	46%
Grand Prix	N/A	44%
Rails	N/A	36%
Tennis Player	N/A	31%
Stereotype Base Rate Neglect		
Jo	Conflict	38%
Jack	Conflict	24%
Kurt	Conflict	17%
Paul	Conflict	36%
Jeremy	Conflict	26%
Ellen	Conflict	79%
Tara	Non-Conflict	89%
Martine	Non-Conflict	85%
Karen	Non-Conflict	94%
Erin	Non-Conflict	95%
Jay	Non-Conflict	93%
Diagnostic Base Rate Neglect		Mean Absolute Distance
Cab	N/A	29.2
Depression	N/A	23.4
Breathalyzer	N/A	56.9
AIDS	N/A	48.6
Skin Cancer	N/A	38.8
Holiday Shopping	N/A	45.7
Face Recognition	N/A	63.7
Nut Detection	N/A	25.0

Face Up Cards: White Triangle; White Square; Black Star; Black Square
Correct Answer: White Triangle; Black Star

Item ID: Face/Number
Scenario: Suppose each card below has a face on one side and a number on the other.
Rule: If there is a frowny face on one side of a card, there must be an odd number on the other side.
Face Up Cards: Smiley Face; Frowny Face; 5; 8
Correct Answer: Frowny Face; 8

Item ID: Color/Circle
Scenario: Suppose each card below has a color on one side and a circle on the other.

Rule: If a card is blue on one side, it must have a red circle on the other side.
Face Up Cards: Red; Blue; Red Circle; Blue Circle
Correct Answer: Blue; Blue Circle

Item ID: Arrow/Number
Scenario: Suppose each card below has an arrow on one side and a number on the other.
Rule: If a card has an up arrow on one side, then it has an odd number on the other side.
Face Up Cards: Up Arrow; Right Arrow; 4; 7
Correct Answer: Up Arrow; 4

Item ID: Letter/Number
Scenario: Suppose each card below has a letter on one side and a number on the other.
Rule: If a card has a blue letter on one side, it will always have a blue number on the other.
Face Up Cards: Blue C; Red C; Blue 7; Red 7
Correct Answer: Blue C; Red 7

Item ID: Drink/Age
Scenario: Suppose each card below has a drink on one side and an age on the other.
Rule: If a patron is drinking a beer, then they must be 21 years or older.
Face Up Cards: Beer; Coke; 35; 19
Correct Answer: Beer; 19

Item ID: Seal/Stamp
Scenario: Suppose each card below has a seal on one side and a postage stamp on the other.
Rule: The country's postal regulation requires that if a letter is sealed, then it must carry a 20-cent stamp. In order to check that the regulation is followed, which of the following four envelopes would you turn over?
Face Up Cards: Sealed Envelope; Unsealed Envelope; 20 Cent Stamp; 10 Cent Stamp
Correct Answer: Sealed Envelope; 10 Cent Stamp

Item ID: Decision/Height
Scenario: Suppose each card below has a decision on one side and a height on the other.
Rule: In order to ride a rollercoaster, you must be at least 5 feet tall.
Face Up Cards: Can Ride Rollercoaster; Cannot Ride Rollercoaster; 5 ft Tall; 4 ft Tall
Correct Answer: Can Ride Rollercoaster; 4 ft Tall

Conjunction Fallacy Task

Instructions: In this task you will be given a scenario and two statements. You will be asked which of the two statements, given the scenario, you think is more probable. Choose the option you think is more probable by using the mouse to 'click' on that option.

Item ID: Linda
Scenario: Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear.

Option 1: Linda is a bank teller
Option 2: Linda is a bank teller and is active in the feminist movement
Correct Answer: Linda is a bank teller

Item ID: Bill
Scenario: Bill is 34 years old. He is intelligent, but unimaginative, compulsive and generally lifeless. In school, he was strong in mathematics but weak in social studies and humanities.
Option 1: Bill plays jazz for a hobby
Option 2: Bill is an accountant who plays jazz for a hobby
Correct Answer: Bill plays jazz for a hobby

Item ID: Colored Die
Scenario: Consider a regular six-sided die with four green faces and two red faces. The die will be rolled 20 times and the sequence of greens (G) and reds (R) will be recorded. Imagine a hypothetical scenario in which you are asked to select one sequence, from a set of three, and you will win $25 if the sequence you chose appears on successive rolls of the die.
Option 1: GRGRRR
Option 2: RGRRR
Correct Answer: RGRRR

Item ID: Scandinavian
Scenario: The Scandinavian peninsula is the European area with the greatest percentage of people with blond hair and blue eyes. This is the case even though every combination of hair and eye color occurs. Suppose we choose at random 100 individuals from the Scandinavian population.
Option 1: Individuals who have blond hair and blue eyes
Option 2: Individuals who have blond hair
Correct Answer: Individuals who have blond hair

Item ID: Grand Pix
Scenario: Suppose Ivan Lendl reaches the final of a Grand Pix tournament.
Option 1: Lendl will lose the first set
Option 2: Lendl will lose the first set, but win the match
Correct Answer: Lendl will lose the first set

Item ID: Rails
Scenario: Because of the Italian Rail's new policies aimed at encouraging voyages longer than 100 km, the number of passengers will
Option 1: will decline by 5% on commuter trains and increase by 10% on long distance trains.
Option 2: will decline by 5% on commuter trains.
Correct Answer: will decline by 5% on commuter trains.

Item ID: Tennis Player
Option 1: Sandy, an alcoholic tennis player, drinks for five more days after which she joins Alcoholics Anonymous. Eight months later, she wins a tennis tournament.
Option 2: Sandy, an alcoholic tennis player, drinks for five more days after which she joins Alcoholics Anonymous. Eight months later, she wins a tennis tournament.
Correct Answer: Sandy, an alcoholic tennis player, drinks for five more days after which she joins Alcoholics Anonymous. Eight months later, she wins a tennis tournament.

Stereotype Base Rate Neglect

Instructions: In a big research project a number of studies were carried out where short personality descriptions of the participants were made. In every study there were participants from two population groups (e.g., carpenters and policemen). In each study one participant was drawn at random from the sample.
You'll get to see the personality description of this randomly chosen participant. You'll also get information about the number of participants from each of the population groups in the study. You'll be asked to indicate to which population group the participant most likely belongs.

Item ID: Bill
Conflict Type: Conflict
Scenario: In a study 1000 people were tested. Among the participants there were 4 men and 996 women. Jo is a randomly chosen participant of this study.
Description: Jo is 23 years old and is finishing a degree in engineering. On Friday nights, Jo like to go out cruising with friends while listening to loud music and drinking beer.
What is more likely?
Option 1: Jo is a man
Option 2: Jo is a woman
Correct Answer: Jo is a man

Item ID: Jack
Conflict Type: Conflict
Scenario: In a study 1000 people were tested. Among the participants there were 5 engineers and 995 lawyers. Jack is a randomly chosen participant of this study.
Description: Jack is 36 years old. He is not married and is somewhat introverted. He likes to spend his free time reading science fiction and writing computer programs.
What is more likely?
Option 1: Jack is an engineer
Option 2: Jack is a lawyer
Correct Answer: Jack is a lawyer

Item ID: Kurt
Conflict Type: Conflict
Scenario: In a study 1000 people were tested. Among the participants there were three who live in a condo and 997 who live in a farmhouse. Kurt is a randomly chosen participant of this study.
Description: Kurt works on Wall Street and is single. He works long hours and wears Armani suits to work. He likes wearing shades.
What is more likely?
Option 1: Kurt lives in a condo
Option 2: Kurt lives in a farmhouse
Correct Answer: Kurt lives in a farmhouse

Item ID: Paul
Conflict Type: Conflict
Scenario: In a study 1000 people were tested. Among the participants there were 997 nurses and 3 doctors. Paul is a randomly chosen participant of this study.

Description: Paul is 34 years old. He lives in a beautiful home in a posh suburb. He is well spoken and very interested in politics. He invests a lot of time in his career.
What is more likely?
Option 1: Paul is a nurse
Option 2: Paul is a doctor
Correct Answer: Paul is a nurse

Item ID: Jeremy
Conflict Type: Conflict
Scenario: In a study 1000 people were tested. Among the participants there were four whose favorite series is Star Trek and 996 whose favorite series is Days of Our Lives. Jeremy is a randomly chosen participant of this study.
Description: Jeremy is 26 and is doing graduate studies in physics. He stays at home most of the time and likes to play video-games.
What is more likely?
Option 1: Jeremy's favorite series is Days of Our Lives
Option 2: Jeremey's favorite series is Star Trek
Correct Answer: Jeremy's favorite series is Days of Our Lives

Item ID: Ellen
Conflict Type: Conflict
Scenario: In a study 1000 people were tested. Among the participants there were 5 sixteen-year olds and 995 fifty-year olds. Ellen is a randomly chosen participant of this study.
Description: Ellen likes to listen to hip hop and rap music. She enjoys wearing tight shirts and jeans. She's fond of dancing and has a small nose piercing.
What is more likely?
Option 1: Ellen is sixteen
Option 2: Ellen is fifty
Correct Answer: Ellen is fifty

Item ID: Tara
Conflict Type: Non-conflict
Scenario: In a study 1000 people were tested. Among the participants there were 4 Bruce Springsteen fans and 996 Britney Spears fans. Tara is a randomly chosen participant of this study.
Description: Tara is 15. She loves to go shopping at the mall and to talk with her friends about their crushes at school.
What is more likely?
Option 1: Tara is a Bruce Springsteen fan
Option 2: Tara is a Britney Spears fan
Correct Answer: Tara is a Britney Spears fan

Item ID: Martine
Conflict Type: Non-conflict
Scenario: In a study 1000 people were tested. Among the participants there were 5 Americans and 995 French people. Martine is a randomly chosen participant of this study.
Description: Martine is 26 years old. She is bilingual and reads a lot in her spare time. She is a very fashionable dresser and a great cook.
What is more likely?
Option 1: Martine is American

Option 2: Martine is French
Correct Answer: Martine is French

Item ID: Karen
Conflict Type: Non-conflict
Scenario: In a study 1000 people were tested. Among the participants there were 995 who buy their clothes at high-end retailers and five who buy their clothes at Wal-Mart. Karen is a randomly chosen participant of this study.
Description: Karen is a 33-year-old female. She works in a business office and drives a Porsche. She lives in a fancy penthouse with her boyfriend.
What is more likely?
Option 1: Karen buys her clothes at high-end retailers
Option 2: Karen buys her clothes at Wal-Mart
Correct Answer: Karen buys her clothes at high-end retailers

Item ID: Erin
Conflict Type: Non-conflict
Scenario: In a study 1000 people were tested. Among the participants there were 997 girls and 3 boys. Erin is a randomly chosen participant of this study.
Description: Erin is 13 years old. Erin's favorite subject is art. Erin's favorite things to do are shopping and having sleepovers with friends to gossip about other kids at school.
What is more likely?
Option 1: Erin is a girl
Option 2: Erin is a boy
Correct Answer: Erin is a girl

Item ID: Jay
Conflict Type: Non-conflict
Scenario: In a study 1000 people were tested. Among the participants there were 997 who have a tattoo and three without tattoo. Jay is a randomly chosen participant of this study.
Description: Jay is a 29-year-old male. He has served a short time in prison. He has been living on his own for 2 years now. He has an older car and listens to punk music.
What is more likely?
Option 1: Jay has a tattoo
Option 2: Jay has no tattoo
Correct Answer: Jay has a tattoo

Diagnostic Base Rate Neglect

Instructions: You will be presented with a scenario in which you must assess the likelihood of an event occurring given certain information. In assessing the likelihood you may you use all, some or none of the information given to you in the scenario, as you see appropriate. Your likelihood assessments should be numbers between 0 and 100. 100 means "I think there is a 100% chance of this event occurring" 0 means "I think there is a 0% chance of this event occurring." 65 means "I think there is a 65% chance of this event occurring" and so forth. You can use any whole number between 0 and 100.

Item ID: Cab
Scenario: Two cab companies operate in a given city, the Blue and the Green (according to the color of cab they run). 85% of the cabs in the city are Blue, and

the remaining 15% are Green. A cab was involved in a hit -and-run accident at night. A witness later identified the cab as a Green cab. The court tested the witnesses' ability to distinguish between Blue and Green cabs under nighttime visibility conditions. It found that the witnesses were able to identify each color correctly about 80% of the time, but confused it with the other color about 20% of the time.
Question: What do you think are the chances that the errant cab was indeed Green, as the witness claimed?
Base Rate Neglect: 80%
Correct Answer: 41%

Item ID: Depression
Scenario: A study was done on the causes of depression among young adults (aged 25 to 35). It was found that the percentage of depression is three times larger among single people than among married people. In this age group, 80% are married and 20% are single.
Question: Of 100 cases of depression among people aged 25 to 35, what percent of those people afflicted would you estimate were single?
Base Rate Neglect: 75%
Correct Answer: 43%

Item ID: Breathalyzer
Scenario: A group of policemen have breathalyzers displaying false drunkenness in 5% of the cases in which the driver is actually sober. However, the breathalyzers never fail to detect a truly drunk person. One out of 1000 of drivers in the population are driving drunk at any given moment. Suppose the policemen then stop a driver at random and force the driver to take a breathalyzer test. It indicates that the driver is drunk.
Question: We assume you don't know anything else about the driver. How high is the probability he or she really is drunk?
Base Rate Neglect: 95%
Correct Answer: 2%

Item ID: AIDS
Scenario: Imagine that AIDS occurs in five in every 1,000 people. Imagine also there is a test to diagnose the disease that always gives a positive result when a person has AIDS. Finally, imagine that the test has a false positive rate of 15 percent. This means that the test wrongly indicates that AIDS is present in 15 percent of the cases where the person does not have AIDS. Imagine that we choose a person to randomly administer the test to, and that it yields a positive result (indicates that the person has AIDS).
Question: What is the probability that the individual actually has AIDS, assuming that we know nothing else about the individual's personal or medical history?
Base Rate Neglect: 85%
Correct Answer: 3%

Item ID: Skin Cancer
Scenario: Imagine that skin cancer in the Atlanta population affects 1% of the population. Maria Sanchez's doctor tells her that he has an inexpensive and accurate test for skin cancer. If a patient has skin cancer the test will detect it 99.5% of the time. However, this means that .5% of the time the test will come back

positive for skin cancer even though the patient does not have it. Maria decides to go ahead with the test and it comes back positive for skin cancer.
Question: What do you think is the probability that Maria has skin cancer?
Base Rate Neglect: 99%
Correct Answer: 67%

Item ID: Holiday Shopping
Scenario: Emily Bean is selling facial products and makeup in a shopping mall during the busy holiday season. Based on sales from previous years it was found that 1% of mall shoppers would stop and buy her product and that 99% would not buy. Emily has developed a strategy to detect which shoppers are most likely to buy her product. Emily has been in this business a long time and is able to correctly detect a potential buyer 70% of the time and incorrectly detects a buyer 30% of the time. Randomly, Emily begins talking to a mall shopper and right away Emily believes the shopper is a potential buyer.
Question: What do you believe are the chances that this person will buy Emily's product
Base Rate Neglect: 70%
Correct Answer: 2%

Item ID: Face Recognition
Scenario: A casino has decided to install a facial recognition software to catch known card counters in their casino. The owners estimate that .1% of the people who come to their casino are card counters and the other 99.9% are not card counters. If the recognition software scans a known card counter, 99% of the time it will correctly identify them and an alarm will notify the owner of the casino immediately. The other 1% of the time the recognitions software will fail to detect the known card counter.
Question: If the owner receives an alarm from the facial recognition software, then what is the probability that the person is in fact a known card counter?
Base Rate Neglect: 99%
Correct Answer: 9%

Item ID: Nut Detection
Scenario: A company is testing a new technology that is able to detect if a food item contains nuts. However, detecting nut content in food is proving difficult. To test this product the company obtained a random sample of food items. 20 of those food items contained nuts while 80 of them contained no nuts at all. The latest results showed that when a food sample contained nuts, the product was able to detect nut content 70% of the time, but 30% of the time it detected nut content even when there was none. Suppose an experimenter at the company uses the product to randomly scan one of the 100 food items.
Question: The nut detector beeps, indicating that it has detected nut content. What is the probability that the food item scanned actually does contain nuts?
Base Rate Neglect: 70%
Correct Answer: 37%