The Cognitive Reflection Test: how much more than Numerical Ability?

Matthew B. Welsh, Nicholas R. Burns & Paul H. Delfabbro
[{matthew.welsh}; {nicholas.burns}; {paul.delfabbro} @adelaide.edu.au]
University of Adelaide, North Terrace
Adelaide, SA 5005 Australia

Abstract

Frederick's (2005) Cognitive Reflection Test (CRT) is a 3-item task shown to predict susceptibility to decision-making biases better than intelligence measures. It is described as measuring 'cognitive reflection' - a metacognitive trait capturing the degree to which people prefer to reflect on answers rather than giving intuitive responses. Herein, we ask how much of the CRT's success can be explained by assuming it is a test of numerical (rather than general) intelligence. Our results show CRT is closely related to numerical ability and that its predictive power is limited to biases with a numerical basis. Although it may also capture some aspect of a rational cognition decision style, it is unrelated to a metacognitive, error-checking and inhibition measure. We conclude that the predictive power of the CRT can, largely, be explained via numerical ability without the need to posit a separate 'cognitive reflection' trait.

Keywords: cognitive reflection; heuristics and biases; individual differences; numerical ability; intelligence.

Introduction

Frederick's (2005) Cognitive Reflection Task (CRT) asks people to solve three, mathematically-simple problems on which intuitive answers are wrong. Frederick explains CRT performance as reflecting a person's preference for using either System 1 (intuitive) or System 2 (rational) processes (Stanovich & West, 2000). Given the ease with which one can check whether intuitive answers are incorrect, the score on CRT shows how likely a person is to reflect on their answer rather than respond intuitively. Frederick's (2005) data shows that CRT is superior to intelligence measures in predicting susceptibility to various cognitive biases or errors made due to inherent, cognitive processes (see, e.g., Tversky & Kaheman, 1974); a conclusion supported by Toplak, West and Stanovich's (2011) recent work.

Given the surprising finding – that a 3-item test better predicts decision-making ability than intelligence tests, Frederick's work has been influential (cited over 600 times). Its results, however, are in line with previous findings which show that, while intelligence is useful in predicting some decision-making biases, in other cases intelligence and bias susceptibility seem independent (Stanovich & West, 2008).

These findings have led to suggestions that decision style (or a person's preference for thinking rationally or intuitively) may be more important than intelligence for predicting bias susceptibility. CRT shares variance with a number of decision style measures (Frederick, 2005) and 'cognitive reflection' is thought to be central to the meta-cognitive processes underlying the relationship between System 1 and System 2 thinking. The latter, System 2 processes, inhibit the automatic and frequently incorrect answers generated by System 1 thinking. It is reasonable, then, that intelligence might determine how efficiently a person uses System 2 reasoning but *whether* they use it may be determined by a separate, metacognitive process, thereby weakening the observed relationship between intelligence and bias susceptibility.

A potential criticism of Frederick's (2005) paper – and other work in this area – however, lies in the choice of intelligence measures. For example, a commonly used intelligence measure is self-reported SAT scores (see, e.g.: Frederick 2005; Stanovich & West, 1998). Another is the Wonderlic Personnel Test (Wonderlic, 1973 – used in Frederick, 2005; and Furnham, Boo & McClelland, 2012). Finally, Toplak et al. (2011), use the Vocabulary and Matrix Reasoning scales from the Wechlser Abbreviated Scale of Intelligence (WASI, Wechsler, 1999).

While all of these do measure 'intelligence' - and WASI divides this into Verbal and Non-verbal ability - none take into account the current understanding of the hierarchical nature of intelligence described by the Cattell-Horn-Carroll model (see, e.g., McGrew, 2005), which recognizes at least ten, related, cognitive abilities. By focusing on the relationship between *general* intelligence and bias susceptibility, it is, therefore, possible to underestimate the relevance of specific intelligences to specific biases.

A key omission is of numerical ability – *Gq* or quantitative ability in CHC terms. Given that the CRT, and many decision-making problems, rely on numerical calculation to determine the correct response, it seems strange to report correlations between biases and general intelligence rather than the type of intelligence most likely to influence such tasks. Thus, it seems possible that the low predictive power of intelligence on bias susceptibility results from poor measure selection.

The way forward, then, is to incorporate measures of the specific abilities most likely to relate to the biases under consideration – thereby establishing an accurate baseline for the strength of the relationship before positing additional constructs like cognitive reflection. Concerning metacognition, this work has already begun, with Toplak et al. (2011) including measures of metacognitive abilities (e.g., working memory; Baddeley & Hitch, 1974) that seem likely to be implicated in recognizing errors in intuition and thus switching from System 1 to System 2 reasoning.

CRT, Heuristics and Biases

Given the numerical basis of the CRT questions, a key question is whether it predicts numerical biases better than less numerical ones. For example, a between-subjects framing task such as the Asian Disease Problem (Tversky & Kahneman, 1981) is structured so that a person can calculate the expected value of the options and recognize that the values of the options do not change with the frame reversal.

By comparison, the conjunction fallacy (Tversky & Kahneman, 1983) requires an understanding of the logical rule of conjunction - and numerical ability *per se* may not assist in avoiding the bias. Similarly, while the anchoring bias (Tversky & Kahneman, 1974) seems numerical – with a seen number affecting a subsequent estimate – numerical ability can not help a person calculate the correct response.

Other tasks are even less clear cut in this aspect. For example, delay discounting tasks like that used by Frederick (2005) can be regarded as a bias measure indicating the extent to which people misjudge the time value of money. This calculation, however, requires the inclusion of non-numerical factors such as immediate need for money and degree of trust in the person offering the delayed reward. Recent work, however, has suggested that these actually measure a distinct personality trait - impulsivity (Odum, 2011) – and, thus, one might expect less covariance between numerical ability and delay discounting. Similarly, a base rate neglect task (see, e.g. Bar-Hillel, 1982) can be answered using a variety of distinct response strategies (Welsh & Navarro, 2012) and, for this reason, it is not necessarily the case that estimates closer to the Bayesian solution actually reflect better numerical skills (Welsh, Burns, Delfabbro & Begg, 2013) as has traditionally been assumed.

Method

Participants

Participants were 58 university students and 44 non-students (22 graduates and 22 who had never attended university), recruited via posters and research participation lists, aged between 18 and 46 (M =22.5, SD = 4.9); sixty-eight were female and all received $50 for participating.

Materials & Procedure

Participants completed an online questionnaire, including demographic details, and the decision style measures described below prior to attending the lab for cognitive and metacognitive tests. The bias measures were included in the online questionnaire – excepting the anchoring task.

Cognitive Reflection Task

Frederick's (1995) CRT was used to measure cognitive reflection. This test asks three questions requiring numerical responses with CRT score being the number answered correctly. For example, the first question asks:

> *A bat and a ball together cost $1.10. The bat costs $1.00 more than the ball. How much does the ball cost?*

Bias Measures

Anchoring. Anchoring bias refers to the unwarranted effect that presented numbers have on subsequent estimates (Tversky & Kahneman, 1974). The measure used here was derived from a computerized card game in which participants estimated the probability that they would win, given the hand they had been dealt (for details, see, Welsh, Delfabbro, Burns & Begg, in press). Prior to this, they were asked whether their chance of winning was greater or less than a randomly generated number (the anchor) between 0 and 100%. The anchoring measure was the partial correlation (controlling for the true chance of winning) between the anchor and the person's estimate - measured across 140 hands. Higher values thus reflect greater influence of the anchor on estimates (i.e., more bias).

Base Rate Neglect. The Taxi Cab problem (Bar-Hillel, 1982) requires people to integrate base rate and reliability information to determine the probability of a taxi involved in an accident actually being the color a witness describes. As previously noted (Welsh, Burns, Delfabbro & Begg, 2013), responses to such problems form distinct categories. We scored responses as Mathematical, Non-Mathematical and Unclassifiable according to whether the person: mathematically combined the probabilities given in the task (i.e., either the Bayesian solution or an incorrect calculation); selected one probability as their response; or did something other than either of these.

Conjunction Fallacy. The Linda Problem (Tversky & Kahneman, 1983) asks participants to judge whether Linda, a woman described as politically active, is more likely to be a "feminist bank teller" or a "bank teller", with the former indicating the conjunction fallacy – as the conjunction can never be more likely than the simple probability of her being a "bank teller".

Delay Discounting. A series of questions asked how long a person would delay taking a smaller amount of money in order to receive a larger amount. The smaller amount varied from $500 to $900 while the delayed amount was always $1000. The maximum delay a participant would tolerate was indicated on an 8 point scale: 1) 6 hours; 2) 1 day; 3) 1 week; 4) 2 months; 5) 6 months; 6) 1 year; 7) 5 years; 8) 25 years). The average of a person's responses from five such questions was used as their overall score.

Framing. The Asian Flu problem (Tversky & Kahneman, 1981) asks people to select a treatment schedule for dealing with a disease outbreak – with either certain (200 alive, 400 dead) or uncertain (1/3 chance of all alive, 2/3 chance of all dead) outcomes. The manipulation lies in the framing of the options. Positive framing describes the treatments in terms of the number of people who live, with the result that more people select the certain option. In contrast, negative framing describes the treatments in terms of the number of people who die, with the result that more people select the uncertain option. Our task included both versions and we categorized people according to whether their responses changed with the frame (displaying the framing bias) or were invariant to framing (unbiased).

Cognitive and Metacognitive Measures

Numerical Abilities Test (NAT). A computerized, 12-item version of the 48-item Numerical Abilities scale from the Differential Aptitudes Test (Bennett, Seashore & Wesman, 1989), measuring quantitative ability (*Gq*).

Symbol-Digit Test (SD). A computerized measure of cognitive processing speed (*Gs*) similar to the Wechlser IQ test's Digit-Symbol (see McPherson & Burns, 2005).

Dot Matrix Task (DM). A computerized version of the Dot Matrix working memory measure (Law, Morrin & Pelligrino, 1995).

Sustained Attention to Response (SART). A computerized test of executive function – requiring the identification and inhibition of a habituated response (Robertson, Manly, Andrade, Baddeley & Yiend, 1997).

Decision Style Measures

Need for Cognition (NfC). The 10-item International Personality Inventory Pool (IPIP; Goldberg et al, 2006) version of Cacioppo & Petty's (1982) scale measuring people's engagement and enjoyment of cognitive activities.

Decision Outcomes Inventory. A 20-item version of Bruine de Bruin, Parker and Fischoff's (2007) test examining whether people have made various, poor decisions (e.g., bought things they did not use, et cetera). The version we used removed US-specific questions.

Rational Experiential Inventory. A 30-item test of risk style (Epstein, Pacinin, Denes-Raj & Heier, 1996) yielding four measures distinguishing between 'Ability' and 'Engagement' for two different cognitive styles – Rational (conscious, analytical) and Experiential (intuitive, holistic).

Intellect. A 20-item inventory from the IPIP (Goldberg et al, 2006) combining Cattell's (1973) and Costa and McCrae's (1992) approaches. This measures openness to new ideas in an intellectual context - a facet of 'openness-to-experience' from Costa and McCrae's (1992) NEO-PI-R.

Rationality. Measured by a 14-item test from the IPIP (Goldberg et al, 2006), high Rationality reflects high Conscientiousness and low Agreeableness in NEO-PI-R (Costa & McCrae, 1992) terms.

Stimulating Instrumental Risk Inventory. A 28-item test yielding two risk attitude measures (Zaleskiewicz, 2001) – Stimulating (positive arousal, short-term and impulsive) and Instrumental (negative arousal, long term and reflective).

Results

CRT and Demographic Measures

Age did not co-vary significantly with CRT (or with any measures other than the SRT and IRT scales from the Stimulating Instrumental Risk Inventory). An independent samples *t*-test, however, showed that males (M = 1.53, SD = 1.11) scored significantly higher than females (M = 1.01, SD = 0.98), on CRT, $t(57)$ = 2.18, p = .033, Cohen's d = .51 - in line with Frederick's (2005) observation.

CRT also varied with level of education, with participants who had never attended university scoring lowest (M = 0.73, SD = 0.94), then current university students (M = 1.24, SD = 1.05) and graduates the highest (M = 1.50, SD = 1.06). A one-way ANOVA confirmed these differences were statistically significant, $F(2,99)$ = 3.31, p = .041, with post-hoc Bonferroni testing indicating that the no-university group differed significantly from both others.

CRT and Biases

Table 1 summarizes the relationships between the CRT and the five bias measures – noting those that have numerically calculable correct responses and which showed significant relationships with the CRT.

Looking at the table, one sees an interesting pattern of responses. While CRT has relationships in the expected directions with the non-calculable biases (Anchoring and the Conjunction Fallacy) these are very weak. By comparison, its relationships with Framing and Discount Delay measures, where the unbiased answer can be calculated, are statistically significant if moderate and weak, respectively. The more complex, near significant relationship between CRT and Base Rate Neglect is discussed more fully below.

CRT and Numerical Ability

Scores on the Numerical Ability Test (NAT) correlated significantly with CRT, $r(102)$ = 0.44, p < .001 – comparable to the correlations observed between CRT and cognitive ability measures in Fredrick (2005) and Toplak et al. (2011). This correlation is the strongest that CRT has with any measure in our analyses.

The relationships between NAT and the demographic variables were also calculated – to determine whether the pattern of responses matches those of the CRT. Welch's *t*-test revealed a non-significant relationship between NAT and Sex in the same direction as the significant relationship shown by the CRT measure, $t(57)$ = 1.28, p = .204.

The relationship between NAT and Education, by contrast, was significant, as indicated by a one-way ANOVA, $F(2,99)$ = 9.41, p <.001. As with CRT, a Bonferroni post hoc test indicated that the no-university group's lower scores drove the significant result and the groups were ordered in the expected manner: no-university; current student; and, then graduates.

Factor Analysis

To assess relationships between CRT and the individual differences measures, an exploratory factor analysis (minres extraction with geomin oblique rotation) was run, revealing the 4-factor solution seen in Table 2. (NB – 2- and 3-factor solutions were also considered; these did not appreciably alter the loadings of the CRT on the first two factors.)

Table 1. Summary of Bias Task Characteristics and Results

Task	Calculable correct response?	Significant relationship with CRT?	Results
Framing	Yes	Yes	***t*(64) = 2.97, *p* = .004, Cohen's *d* = .62.** People whose responses varied with the frame scored lower than those whose responses were invariant to the frame (CRT = 0.76 vs 1.39).
Discount Delay	Yes	Yes	***r*(102) = 0.25, *p* = .010.** Higher CRT accompanied a greater willingness to wait for the larger reward.
Base Rate Neglect	Yes	No	***F*(2,98) = 2.79, *p* = .07.** People whose responses were classified as Mathematical (CRT=1.28) did not score better than the Non-Mathematical group (1.33) but both scored better than the Unclassified group (0.71).
Anchoring	No	No	***r*(102) = -0.11, *p* = .255** People more susceptible to anchoring bias scored slightly lower on CRT
Conjunction Fallacy	No	No	***t*(56) = 0.37, *p* = 0.71** People committing fallacy scored slightly lower on CRT (1.30 vs 1.21).

Looking at Table 2, one can see that a sensible structure emerges. The first factor captures the decision style measures relating to people's tendencies toward 'rational cognition'. The second seems to be an intelligence factor. The third has only the experiential measures from the Rational-Experiential Inventory (REI) loading on it – reflecting a tendency toward intuitive thinking. Finally, the fourth factor reflects attitudes to risk as captured by both measures form the Stimulating-Instrumental Risk Inventory.

Only one variable, the Rational Ability measure from the REI, loads on more than one factor at the conventional 0.3 level and only the Sustained Attention to Response Task fails to load on any factor – indicating the metacognitive measure differs from both decision style and intelligence.

Table 2. Factor loadings of CRT, cognitive and decision-style measures.

	Factors				
Variable	1	2	3	4	h^2
Intellect	**.92**	-.06	.13	.00	*.94*
Need for Cognition	**.91**	.02	.03	.05	*.85*
Rational Engagement	**.75**	-.06	-.02	.01	*.56*
Rationality	**.74**	.04	.00	-.05	*.55*
Rational Ability	**.62**	.32	-.12	-.01	*.48*
Dot Matrix	-.08	**.77**	-.03	.12	*.61*
Symbol-Digit	.00	**.68**	.16	-.06	*.44*
Numerical Ability	.00	**.66**	.02	.01	*.43*
Cognitive Reflection	.26	**.50**	.01	-.05	*.32*
Exper. Engagement	-.01	.02	**1.00**	.00	*1.00*
Experiential Ability	.16	-.02	**.68**	.02	*.56*
Stimulating Risk	-.02	-.02	.04	**.99**	*1.00*
Instrumental Risk	.04	.24	-.17	**.45**	*.26*
Sustained Attention RT	-.01	-.01	.05	.18	*.05*

Primary factor loadings are in bold. h^2 = communality, the variance in each variable captured by the four factors.

Discussion

The above results suggest that 'cognitive reflection', as measured by CRT, shares much in common with numerical ability – although there remains additional, unshared variance to account for. Key, individual results are discussed below, along with caveats and potential future research.

Cognitive Reflection and Sex

An interesting result is the relationship between CRT and Sex - and the lack of similar relationships between Sex and the other measures loading on the 'intelligence' factor in Table 2. The sex difference on CRT was observed by Frederick (2005), who noted that it was unrelated to differences in intelligence and suggested that it might be related to differences in mathematical ability. This was not, however, supported by our data where no significant relationship was seen between numerical ability and sex.

The only variable with which both Sex and CRT shared a relationship was the Rational Ability scale of the Rational-Experiential Inventory. CRT correlated with RA significantly, $r(102) = 0.33$, $p < .001$ and men's scores (22.7) were higher than women's (20.9) – significantly according to Welch's *t*-test, $t(83) = 2.34$, $p = .022$ - suggesting that the sex difference in CRT may partly reflect a difference in Rational Ability – a person's self-reported ability to think analytically (Epstein et al, 1996).

Cognitive Reflection and Numerical Biases

The pattern of bias results described above fit with a conception of the CRT as a primarily numerical measure.

On those bias tasks where numerical skill has no obvious role in arriving at the correct response – anchoring and the conjunction fallacy, the CRT has no predictive value.

By comparison, in the framing problem, where the irrelevance of the frame can be demonstrated numerically, CRT proved a good predictor of performance. Similarly, there is a significant effect for the delay discounting problem. Despite the complexity of the problem (in terms of potential, contextual factors) it appears that numerical ability pushes participants towards the economically rational choice. This is an interesting addition to Baumann and Odum's (2012) finding that delay behavior relates to temporal perception – potentially arguing for a relationship between numerical and temporal skills under the broad *Gq* 'quantitative ability' umbrella.

Complexity is added by the base rate neglect task, where the results were somewhat unexpected – although not significant. As noted above, the groups using mathematical and non-mathematical strategies did not differ statistically from one another on the CRT. Instead, both groups outscored participants whose responses were unclassifiable. As noted by Welsh et al. (2013), however, the base rate neglect task differs from many numerical bias tasks in that the calculation of the correct solution is dependent on knowing how to undertake Bayesian updating. That is, while a person with high cognitive reflection or numeracy may realize that their intuitive response is wrong and activate System 2 thinking, they may not have the knowledge required to calculate the correct answer having done so. Given that CRT only requires very simple mathematical skills – as do numerical ability tests – this task's failure to predict response types on a base rate neglect task is less surprising than it first seems.

Cognitive Reflection and Intelligence

The factor analysis shown in Table 2 indicates that the CRT is, primarily, an intelligence measure – loading on the second factor along with the three cognitive variables. It does, however, have the weakest loading of the four on this factor at 0.50. Numerical ability is, however, the variable with the most similar loading – reflecting the strength of the relationship between these two measures. This is unsurprising as the Dot Matrix and Symbol-Digit tasks require learning a novel task, whereas the NAT and CRT require prior knowledge - of how to undertake mathematical operations. (CRT scores could also be affected by prior experience of questions similar to those used in the task – making people wary of too-easy answers.)

CRT shows virtually no relationship (loadings of .01 and -.05) with the third and fourth factors ('intuition' and 'risk attitude') but its relationship to the first factor bears some scrutiny. While not reaching the 0.3 level conventionally required to be included amongst the variables loading on a factor, its loading of 0.26 on the 'rational cognition' decision style factor approaches this level and is the second highest secondary loading in Table 2 – after Rational Ability's 0.32 secondary loading on the 'intelligence' factor.

This could be taken as offering some support for Frederick (2005) and Toplak et al.'s (2011) conclusions that CRT measures something more than cognitive ability – although the factor loadings suggest that the cognitive aspect is more central.

Cognitive Reflection and Metacognition

A final observation from the above results is the lack of any relationship between the CRT and the Sustained Attention to Response Task (SART), which measures executive functioning – specifically, a person's ability to monitor their performance for errors and to inhibit incorrect responses.

Given the description of the CRT as a measure of a person's preference for activating rational thinking and thus recognizing errors in intuitive responses, its failure to correlate with the SART seems strange. In light of our results, it thus seems possible that the CRT is measuring only a person's ability to recognize errors in intuitive, *numerical* results rather than the more general metacognitive function.

Caveats and Future Research

While including measures not previously used in studies of cognitive reflection, the present analyses remain limited in their scope. The Cattell-Horn-Carroll model includes ten specific types of intelligence (acknowledging the possibility of more; McGrew, 2005). Of these, only two (plus the non-CHC working memory) were measured herein – *Gs* (processing speed) and *Gq* (quantitative or numerical ability). Similarly, while five biases were included here, further effects from the biases literature could improve understanding of what CRT does and does not predict.

A further concern is the sample size. While 102 participants is sufficient to find most large or moderate effects, small effects may still be missed. Frederick's (2005) study, for example, involved more than 3000 participants, allowing statistical significance for even very weak relationships. Given this, the obvious direction for future research is a larger study of participants from a wide range of educational backgrounds, utilizing the widest possible range of biases and cognitive abilities in order to pin down exactly what the CRT is. Including a number of tasks measuring quantitative (numerical) ability would also allow further factor analyses to decide conclusively whether CRT is, as suggested here, primarily a numerical task.

Another key direction is to determine what role metacognitive abilities do play in the divide between System 1 (intuitive) and System 2 (analytic) reasoning and whether CRT is capturing any of these. Specific measures of impulsivity, as discussed by Baumann and Odum (2012), could inform this – as this seems likely to relate to the likelihood of a person relying on System 1.

Finally, additional work could address whether CRT scores are affected by prior experience of similar questions.

Conclusions

The above results support the idea that CRT is, at heart, a numerical task, correlating with quantitative ability and predicting bias only on tasks with a calculable, correct answer. It may, however, measure some aspect of a 'rational cognition' decision style. The CRT does not, however, relate to the executive functioning measure included here, suggesting that 'cognitive reflection' may not be metacognitive as Frederick (2005) describes but, rather, measure a person's ability to quickly recognize bad math.

Acknowledgments

MBW and SHB thank ExxonMobil and Santos for their support of this research through the CIBP at the Australian School of Petroleum. The authors thank Karina Burns for her assistance with data collection and Reidar Bratvold and Steve Begg for discussions re the CRT's nature and use.

References

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. *The psychology of learning and motivation*, *8*, 47-89.

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica*, 44, 211-233.

Baumann, A. A., & Odum, A. L. (2012). Impulsivity, risk taking, and timing. *Behavioural Processes*, *90*, 408-414.

Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1989). *Differential Aptitude Test*. Marrickville, NSW: Psychological Corporation.

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

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*(1), 116-131.

Cattell, R.B. (1973). *Personality and mood by questionnaire*. San Francisco: Jossey-Bass.

Costa, P. T. J., & McCrae, R. R. (1992). *NEO PI-R Professional Manual*. Odessa, FL: Psychological Assessment Resources.

Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in Intuitive-Experiential and Analytical-Rational Thinking Styles. *Journal of Personality and Social Psychology, 71*(2), 390-405.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4), 25-42.

Furnham, A., Boo, H.C. & McClelland, A. (2012). Individual differences and the susceptibility to the influence of anchoring cues. *Journal of Individual Differences, 33*(2), 89-93.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality, 40*, 84-96.

Law, D.J., Morrin, K.A., Pellegrino, J.W. (1995). Training effects and working memory contributions to skill acquisition in a complex coordination task. *Learning and Individual Differences, 7,* 207–234.

McGrew, K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities: Past, present, and future. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp.136–182). New York: Guilford.

McPherson, J.L & Burns, N.R. (2005). A speeded coding task using a computer based mouse response. *Behavior Research Methods, Instruments & Computers, 37,* 538-544.

Odum, A.L. (2011). Delay discounting: trait variable? *Behavioural Processes*, *87*, 1-9.

Robertson, I.H., Manly, T., Andrade, J., Baddeley, B.T., Yiend, J. (1997). 'Oops!': performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia, 35,* 747–758.

Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General,* 127(2), 161-188.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: implications for the rationality debate. *Behavioral and Brain Sciences*, *23*, 645-726.

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology*, 94(4), 672-695.

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 and Cognition, 39*, 1275-1289.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124-1131.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453-458.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological review*, *90*(4), 293.

Wechsler, D. (1999). *Wechsler abbreviated scale of intelligence (WASI).* San Antonio: Harcourt Brace, Psychological Corp.

Welsh, M.B., Burns, N.R., Delfabbro, P.H. & Begg, S.H. (2013). *Individual differences in base rate neglect responses and susceptibility*. Manuscript in preparation. www.adelaide.edu.au/directory/matthew.welsh.

Welsh, M.B., Delfabbro, P.H., Burns, N.R., Begg, S.H. (in press). Individual differences in anchoring: traits and experience. *Learning and Individual Differences: Special Issue on Metacognition*. Accepted 9th January 2013.

Welsh, M. B., & Navarro, D. J. (2012). Seeing is believing: priors, trust and base rate neglect. *Organizational Behavior and Human Decision Processes*, *119(1),* 1-14.

Wonderlic, E. F. (1973). *Wonderlic Personnel Test.* Los Angeles: Western Psychological Services.

Zaleskiewicz, T. (2001). Beyond risk seeking and risk aversion: Personality and the dual nature of economic risk taking. *European Journal of Personality*, *15*, S105-S122.