

# Individual differences in decision-making and confidence: capturing decision tendencies in a fictitious medical test

Simon A. Jackson · Sabina Kleitman

Received: 31 October 2012 / Accepted: 5 December 2013 /  
Published online: 22 December 2013  
© Springer Science+Business Media New York 2013

**Abstract** Decision-making is a complex process that is largely studied from an experimental perspective or in specific organizational contexts. As such, no generalizable framework exists with which to study decision-making from an individual differences perspective for predictive/selection purposes. By generalising a context-specific decision model proposed by Koriat and Goldsmith (1996), the focus of this research was to therefore test a novel framework for studying individual differences in decision-making tendencies. Utilising this framework within a fictitious Medical Decision-Making Test (MDMT) yielded five novel variables that provided unique insight into individuals' decision tendencies: Optimal, Realistic, Incompetent, Hesitant and Congruent. Metacognitive confidence and its calibration (bias and CAQ) were used as predictor variables to validate this framework. One hundred ninety-three undergraduate students completed the MDMT and three cognitive ability tests with confidence ratings, a personality questionnaire, and the Need for Closure questionnaire. All decision tendency variables demonstrated excellent internal consistency and were predicted by the metacognitive variables incrementally to the remaining variables as hypothesized. Additionally, the metacognitive indices were found to generalize across the decision-making and cognitive tests. The results imply that this novel framework and MDMT reliably capture individuals' decision behaviour that shares a meaningful relationship with their general confidence and calibration.

**Keywords** Decision-making · Individual differences · Confidence · Calibration · Metacognition

## Introduction

Decision-making is a complex process of making choices in order to achieve our goals, and identifying the psychological constructs that contribute to this process can be of great importance. Research in this area has, by and large, been conducted from an experimental

---

S. A. Jackson (✉) · S. Kleitman  
School of Psychology, The University of Sydney, Sydney, NSW 2006, Australia  
e-mail: sjac9712@uni.sydney.edu.au

S. Kleitman  
e-mail: sabinak@psych.usyd.edu.au

perspective or in specific organizational contexts. That is, decision-making has been largely studied via group differences, where individual deviation from the group mean is treated as error. While contributing to our understanding immensely, the lack of research being conducted from an individual differences perspective limits the applied use of relevant measures for predictive, selection/profiling purposes. Individual differences, or differential psychology is concerned with how individuals differ from each other and treats stable differences relative to other people as meaningful variance. In this paper, we will propose and test a broadly applicable framework to study decision-making from an individual differences perspective in order to address this limitation. Judgement confidence, an important metacognitive experience that drives decision-making and is known to reliably differ between individuals, will be used as a predictor variable to validate this framework.

The decision-making process involves making judgements that inform decisions (Edwards 1954, 1961; Harvey 2001; Mellers et al. 1998). Judgements are our beliefs or predictions and they vary in the extent to which they are an accurate reflection of reality (accuracy). By themselves, judgements do not have any direct consequences in the world. Rather, they have indirect consequences via the decisions that they inform (Harvey 2001). Decisions are choices that can influence the world around us, and vary in the extent to which the outcomes of that influence, or lack of influence, meet our goals (i.e. their optimality; see Harvey 2001 for a review). Importantly, the optimality of a decision varies as a function of the accuracy of the judgements on which it is based. As an intentionally simplified example, a doctor's decision to conduct surgery rather than acquire more information is optimal following a correct diagnosis (judgement), but incompetent following an incorrect one. Our framework for the study of individual differences in decision-making will be based on this concept that decision optimality can vary as a function of judgement accuracy.

### Decision tendencies

This same concept was used by Koriat and Goldsmith (1996) to classify task-specific decisions given preceding judgement accuracy. In their study, participants completed a general knowledge test with the ability to provide or withhold answers. Participants received monetary rewards for questions answered correctly but lost money for questions answered incorrectly. No gains or losses were incurred for unanswered questions. They classified decisions into four categories as follows: a decision to answer a question when the answer was correct was a *Hit*, whereas this same decision following an incorrect answer was called a *False alarm*; Similarly, the decision to withhold a response when the answer was incorrect was a *Correct rejection*, alternatively labelled a *Miss* following a correct judgement. To extend Koriat & Goldsmith's work, we will (a) define and distinguish congruent and incongruent decisions to extend this task-specific categorisation to a more general framework; and (b) transfer this framework to an individual differences approach that captures decision-making tendencies.

*Congruent/incongruent decisions* A congruent decision involves making a choice that is most optimal if the judgements upon which it is based are true. In contrast, an incongruent decision describes making any other available choice: That is, any choice that would *not* be most optimal were the judgements true. For example, a clinician conducting an appendectomy is congruent with a diagnosis of appendicitis; investing in gold is congruent with the prediction that the price of gold will rise; or an eyewitness volunteering information is congruent with their memory of an event. Note that this does not mean that, in reality, the judgements have been accurate and/or the decision has been optimal. Rather, congruent decisions are in line with the *belief* that the judgements are accurate. the clinician deciding to acquire a blood test,

the investor keeping their cash, or the eyewitness withholding information are all incongruent with their respective judgements. Thus, Table 1 shows our general extension of Koriat & Goldsmith's model.

*Individual differences model* Of great interest to psychologists and applied practitioners alike would be the ability to reliably profile an individual's tendencies to make decisions that result in each of these categories. For example, it would be of great benefit to profile undergraduate medical student tendencies to incompetently treat misdiagnosed patients (False alarms) to better address their weaknesses; Companies may want to select individuals who tend to maximise their Hits and/or Correct rejections when making financial decisions. Yet, to our knowledge, no such system exists for measuring such decision tendencies. To create such a system, an individual must repeatedly make similar decisions in a given scenario. The frequency of their decisions that result in each general category can then be used to compute different variables that capture that individual's decision-making tendencies within that scenario. Many computations with these raw values can be conducted. However, the five variables that we propose capture meaningful and important individual differences in the tendency to make decisions that lead to particular outcomes are outlined in Table 2 and described below.

*Optimal decision tendencies* is the primary index of optimal decision-making and computed as the total number of Hits divided by the number of decisions. For example, an optimal clinician tends to correctly diagnose and immediately treat their patients.

*Realistic decision tendencies* is the second optimal variable, measuring an individual's tendency to make decisions that accurately reflect reality and reduce decision errors overall.<sup>1</sup> It is computed as the sum of Hits and Correct rejections, divided by the number of decisions. For example, a realistic clinician tends to treat their patients following correct diagnoses or request tests following incorrect diagnoses.

*Incompetent decision tendencies*, reflecting erroneous decisions, is associated with congruent tendencies following incorrect judgments. It is computed as the frequency of False alarms divided by the number of *incorrect* judgements. For example, an incompetent clinician tends to treat their patients despite incorrect diagnoses.

*Hesitant decision tendencies* is another error variable and is associated with incongruent tendencies following correct diagnostic judgments. It is computed as the frequency of Misses divided by the frequency of *correct* judgements. For example, a hesitant clinician tends to request tests despite correct diagnoses.

*Congruent decision tendencies* is computed as the sum of Hits and False alarms, divided by the total number of decisions. It indicates an individual's tendency to make congruent decisions overall, such as a congruent clinician who tends to treat their patients in general.

To our knowledge, the study of individual differences in the tendency to make decisions that fall into these categories has never been done. For example, the question of whether individuals reliably vary in their tendencies to be optimal decision-makers by maximising their Hits, or be Realistic decision-makers by further maximising their Correct rejections, has never been asked. This is likely to be best studied with controlled and repeated stimuli. We will test this framework by computing these variables within a novel decision-making test which utilises a fictitious medical scenario: the Medical Decision-making Test (MDMT). We will then assess their reliability and determine whether they are predicted by variables assumed to underpin decision-making: metacognitive confidence and its calibration.

<sup>1</sup> This variable can also be thought of as the inverse proportion of erroneous decisions, as it is equal to  $1 - (B + C)/(A + B + C + D)$ .

**Table 1** General extension of Koriat and Goldsmith's (1996) model for classifying decisions

Decision	Judgement accuracy	
	Correct	Incorrect
Congruent	<i>A. Hits</i>	<i>B. False alarms</i>
Incongruent	<i>C. Misses</i>	<i>D. Correct rejections</i>

## Cognition and metacognition

Cognitive processes inform our judgements. For example, a clinician employs pattern recognition and inference to form a preliminary patient diagnosis. Broadly speaking, an individual's cognitive abilities underpin the formation of accurate judgements, including correct answers, leading to greater optimal decision tendencies. The study of cognitive abilities has a long history, and valid and reliable measures for capturing individual differences are well established. Most prominent are measures of fluid and crystallised intelligence, which define two broad ability domains related to levels of reasoning ability and knowledge respectively (Cattell 1987). Hence, a battery of standardised fluid and crystallised intelligence tests will be administered with the Medical Decision-Making Test. While cognitive abilities are not the focus of this research, these tests will be included to control for these well-known abilities and for reasons discussed later.

The focus of this research will be on the metacognitive processes that monitor and control these cognitive processes to direct decisions (see Azevedo 2009; Efkliides 2008; Nelson 1996; Stankov 1999, for reviews). In this research, we focus on one of the key metacognitive experiences, judgement confidence, which reflect one's certainty that their judgement is accurate (Allwood et al. 2006; DeMarree and Petty 2007; Efkliides 2008; Stankov 1999) thus guiding decision behaviour (Cowley 2004). Using this approach, people are asked to provide their on-task confidence estimates while being engaged with cognitive stimuli. These confidence levels are used on their own to gauge the levels of certainty that the individual holds in decisions/decisions/opinions (see Harvey 1997; Kleitman 2008; Stankov 1999 for reviews). They can also be used to derive calibration measures which reflect the 'goodness of fit' of confidence to the accuracy of the judgements from which it is derived (see Schraw 2009 for a review). In order to validate our decision framework, the following sections will outline the way in which confidence and some of its calibration indices should differentially predict our five decision tendency variables.

**Table 2** Decision tendency variables

Decision tendency	Indexes	Calculation <sup>a</sup>	Desired score
Optimal decision tendencies	Optimal decision-making	$(A)/(A + B + C + D)$	1
Realistic decision tendencies	Overall error reduction	$(A + D)/(A + B + C + D)$	1
Incompetent decision tendencies	Congruent errors	$(B)/(B + D)$	0
Hesitant decision tendencies	Incongruent errors	$(C)/(A + C)$	0
Congruent decision tendencies	Overall congruent decisions	$(A + B)/(A + B + C + D)$	n/a

Each variable is calculated as a proportion ranging from 0 to 1

<sup>a</sup> A, B, C and D represent the four cells in Table 1

## The confidence assumption

Confidence levels hold a strong position in judgement and decision-making research. For example, Knight (1921, p. 100), claimed “the action which follows an opinion depends as much upon the amount of confidence in that opinion as it does upon the favourableness of the opinion itself.” Some of the biggest names in contemporary decision-making research have asserted “confidence controls action” (Gilovich et al. 2002, p. 248). Although these claims were made in reference to different theories, each shares the assumption that as confidence in one’s judgement—and thus subjective certainty—increases, so too does the likelihood of translating that judgment into a congruent decision (DeMarree and Petty 2007; Slovic et al. 1977). This assumption will hereafter be referred to as the *Confidence Assumption*.

Currently, the primary methodological approach to the study of the Confidence Assumption has been experimental and focused on the frequency of congruent decisions only. For example, McKenzie (1998) utilised a fictitious medical scenario test to assess changes in confidence following various learning procedures. In this test, participants adopted the role of a physician learning to use symptoms to diagnose fictitious patients with one of two illnesses: puneria and zymosis. During learning, participants were shown 40 patient profiles each listing the same set of symptoms, whether each symptom was present or not, and which illness the patient had. Symptoms were assigned prior probabilities of 0.85 or 0.15 for puneria and/or zymosis. That is, each symptom occurred in 85 % or 15 % of the 40 profiles shown during learning. Participants were told that the prevalence of a particular symptom with each illness during learning was indicative of its association with that illness. For example, if headaches occurred in 85 % of the puneria profiles but only 15 % of the zymosis profiles during learning, then the most accurate diagnosis for a novel patient with headaches was puneria. Participants then saw two novel patient profiles: a ‘both’ profile which, based on the present symptoms, could be diagnosed as either puneria or zymosis; and a ‘neither’ profile that could not be diagnosed with either illness. Participants then indicated their confidence that each new patient had a focal illness—half puneria, half zymosis—and were then asked if they would administer that patient the focal illness treatment (a congruent decision). Participants were significantly more confident and more likely to treat the patient when judging the ‘both’ profile rather than the ‘neither’ profile. While this supports the Confidence Assumption, it does not answer whether any single individual tends to initiate congruent decisions. Nor does it address whether the degree to which an individual’s confidence departs from the group mean can be treated as meaningful variance, or simply error (as is the case in experimental research).

Koriat and Goldsmith (1996) found similar results. Before administering participants the general knowledge test in which they could withhold responses, participants received an identical test in which responses and confidence regarding the accuracy of those responses were required. The results clearly demonstrated that questions answered in the latter test were assigned confidence ratings, from the former test, almost four times greater than questions for which answers were withheld. That is, in support of the Confidence Assumption, congruent decisions followed judgements made with higher confidence than judgements followed by incongruent decisions. Again, however, individual differences in confidence and decision-making tendencies were not assessed. In this study we combine both of these approaches to capture individual differences in decision-making tendencies and to determine their vital psychometric properties: reliability and predictive validity.

*Individual differences in decision-making and confidence* Confidence is undoubtedly an important psychological experience that guides decision-making, but this relationship has never been examined from an individual differences perspective. We should, for example,

expect more confident individuals to demonstrate greater congruent, optimal and incompetent tendencies, and lower hesitant tendencies. That is, more confident individuals should tend to make more congruent decisions, leading to more hits while holding accuracy constant, more false alarms following incorrect judgements, and fewer misses following correct judgements. Hence, the robust and reliable individual differences observed in confidence measures—described shortly—should predict our decision tendency variables as outlined above.

### Measuring confidence

Inline with McKenzie and Koriat & Goldsmith, this research focuses on the measurement of judgement confidence that immediately follows a cognitive act, and reflects the assessment of one's performance. The methods for measuring confidence in this way can vary considerably across domains (see Moore and Healy 2008, for a review). One popular approach is to ask an individual how confident they are in the accuracy of their judgement as a percentage (e.g., Allwood et al. 2006; Costermans et al. 1992; Efklides 2008; Flavell 1979; Schraw and Dennison 1994; Stankov 1999, 2000). Specifically, individuals indicate confidence in their judgement from guessing (0 % for open-ended questions, 20 % for 5-option multiple choice, etc.) to 100 % being absolutely sure they are correct.<sup>2</sup> Figure 1 shows a typical cognitive knowledge question with the respective confidence rating, which comes from the Vocabulary test (Stankov 1997) used in the present research.

Upon completion of a test, the values of all such confidence ratings are averaged to give a continuous measure of overall confidence in that test. This method of assessment has been demonstrated to be well understood by adults (Williams and Gilovich 2008) and children (Kleitman and Gibson 2011; Kleitman et al. 2011), and to possess excellent psychometric properties (see Stankov and Kleitman 2008, for a review). Hence, the present study adopts the percentage scale approach for measuring confidence.

### Calibration

Measuring confidence in this way allows for a multitude of calibration indices to be calculated (Boekaerts and Rozendaal 2010; Harvey 1997; Schraw 2009; Yates 1990). Calibration is broadly defined as a metacognitive phenomenon relating to the adaptiveness and effectiveness of the monitoring process (Nelson 1996; Stankov 1999), and such measures indicate the “goodness of fit” of confidence ratings to the accuracy of the judgements from which it is derived (Schraw 2009). Hence, good calibration is assumed to be necessary in order to execute optimal decision-making behaviour. The two measures for assessing calibration that are of the utmost theoretical importance, and therefore best used to validate the decision tendency variables, are bias and CAQ.

*Bias* The most widely used and investigated calibration index is bias, also referred to as over/underconfidence. The bias score indicates whether, on average, an individual has been able to match or calibrate their confidence levels with their actual levels of accuracy (Stankov

<sup>2</sup> The lower end of the confidence scale is defined by the minimum probability of being correct. For example, in a five option multiple-choice test the probability of being correct is at least 20 %. In this case, the scale will range from 20 to 100 %. The scale can request any value, or make categories available such as an 11-point scale ranging from 0 to 100 %.

What means the same as FEIGN?										
Pretend				Prefer				Wear		
<input type="checkbox"/>				<input type="checkbox"/>				<input type="checkbox"/>	Be cautious	<input type="checkbox"/>
How confident are you that your answer is correct?										
20%	30%	40%	50%	60%	70%	80%	90%	100%		
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

**Fig. 1** Example vocabulary test question and accompanying confidence rating

et al. 2012). Bias is calculated across a number of judgements as the difference between average subjective confidence estimates and objective accuracy:

$$\text{Bias} = \frac{\sum c_i}{n} - \frac{\sum a_i}{n}$$

Where  $n$  is the total number of items;  $c_i$  is the confidence assigned to the  $i^{\text{th}}$  item;  $a_i$  is accuracy of the  $i^{\text{th}}$  item, scored 1 for correct and 0 for incorrect. Bias scores can range from +1 to -1. High (above zero) and low (below zero) scores, indicative of poor confidence calibration, are described as over and underconfidence respectively (Lichtenstein and Fischhoff 1977). Furthermore, superior test-retest and split-half reliability estimates for bias have been demonstrated relative to other calibration indices (Stankov and Crawford 1996).

Differential predictions, with respect to decision-making, are made regarding the *direction* of the bias score. For example, in the area of self-regulated learning, Efklides (2009, p. 81) postulates that overconfidence can make an individual “less perceptive of situational demands,” and underconfidence may increase anxiety resulting in task avoidance. Others have suggested that underconfident students might devote an unnecessary amount of time to studying learned material (Hacker et al. 2008). Likewise, financial decision models incorporating the confidence calibration model predict that overconfident investors will trade more stock than well-calibrated investors (Glaser and Weber 2007). While only limited support is available for these predictions, across domains, they can be largely described in terms of our decision framework: Overconfident individuals tend to be incompetent, incorrectly making congruent decisions following incorrect judgements, while underconfident individuals tend to be hesitant, incorrectly making incongruent decisions following correct judgements.

Further predictions are made on the basis of the *magnitude* of the bias score—increasing deviation from zero—in either direction, reflecting the raw degree of miscalibration. The larger the magnitude, the larger the discrepancy between one’s confidence in their performance and their actual performance. Normatively, as *either* over or underconfidence increases, self-monitoring skills are more impaired and the tendency to reduce errors overall should decrease at an increasing rate (see Schraw 2009; Stankov 1999 for reviews). Because of this property, a quadratic relationship is expected between bias and any measure of optimal behaviour as, theoretically, it should increase when bias approaches zero. A linear relationship would simply imply that as the bias score is increasing/decreasing, optimal behaviour will increase/decrease: A relationship that does not address the complex nature of the bias score. In contrast, a quadratic function would imply that optimal behaviour is expected at a particular level of bias, here zero. That is, optimal and realistic tendencies should decline—at an increasing rate—as bias deviates from zero. Thus, the hypothesised relationship between the bias score and these decision tendencies can be best represented by a quadratic function—illustrated later in Fig. 3—where scores

on these tendencies increase to their most optimal levels as the bias score approaches zero. As research typically considers only the linear relationship between bias and different outcomes (e.g., Yang and Thompson 2010; Glaser and Weber 2007), our investigation of this quadratic component, although clearly predicted from a theoretical perspective, will be largely novel. In-line with the traditional approach, we will also investigate a linear relationship between bias and the remaining decision variables: Increasing bias—from under to overconfidence—should predict an increase in congruent and incompetent decision tendencies, and a decrease in hesitant tendencies.

*CAQ* The second calibration index is a measure of discrimination, referring to the ability to distinguish correct and incorrect judgements. Discrimination is typically computed simply as the difference between average confidence assigned to correct and incorrect items. However, this score is unduly dependent on the distance measured by the confidence scale. For example, there is no reason to suspect that an individual who assigns 40 % to all incorrect and 60 % to all correct answers is a poorer discriminator than one who assigns 30 % and 70 % respectively, but the traditional measure would say they are. There are a number of indices designed to capture this aspect of metacognitive monitoring (e.g., slope, resolution), but these have not demonstrated acceptable internal reliability estimates in the past to justify their use in individual differences research (Stankov and Crawford 1996). Hence, to account for such individual variation in confidence ratings, we will use a comparable measure that standardises these distances: the confidence-judgement accuracy quotient (CAQ; Shaughnessy 1979; Schraw 2009), calculated as the difference between average confidence assigned to correct and incorrect items, divided by the standard deviation of all confidence ratings. Formally:

$$CAQ = \frac{\left( \frac{\sum c_{i\text{correct}}}{p} - \frac{\sum c_{i\text{incorrect}}}{q} \right)}{\sigma}$$

Where  $c_{i\text{correct}}$  is confidence assigned to the  $i^{\text{th}}$  correct item;  $p$  is the number of correct items;  $c_{i\text{incorrect}}$  is confidence assigned to the  $i^{\text{th}}$  incorrect item;  $q$  is the number of incorrect items;  $\sigma$  is the standard deviation of all the confidence ratings, and adjusts for how tightly an individual uses the confidence scale. CAQ scores can range from negative to positive values, with positive scores indicating higher confidence for correct rather than incorrect judgements. Negative values also indicate better, albeit faulty, discrimination. Thus, increasingly extreme CAQ indicates better discrimination with positive values being desirable.

Discrimination was highlighted by Koriat & Goldsmith as the most important metacognitive index for decision-making performance. This is because individuals who discriminate well should make congruent decisions when their judgements are correct, as a result of higher certainty/confidence. Furthermore, they should tend to make incongruent decisions when their judgements are incorrect, as a result of lower certainty/confidence. We therefore expect that better discrimination (indexed by higher CAQ) will predict increasing optimal and realistic decision tendencies and decreasing incompetent and hesitant tendencies. Given the nature of this hypothesis, only a linear trend is expected.

#### Individual differences in confidence

Aside from its clear theoretical and empirical importance, our use of confidence is motivated by research that has found stable individual differences in confidence levels within the



cognitive domain (e.g., Kleitman and Mascrop 2010; Kleitman and Stankov 2001, 2007; Kleitman et al. 2011; Mengelkamp and Bannert 2010; Pallier et al. 2002; Schraw et al. 1995; Stankov 1999; Stankov and Crawford 1996, 1997; Stankov and Lee 2008). That is, regardless of changes in overall confidence levels, individuals more or less confident than others in one cognitive test tend to be respectively more or less confident than others in any other cognitive test. The degree to which an individual's confidence departs from a group mean cannot, therefore, simply be error. In particular, these studies found that confidence levels acquired across a diverse battery of cognitive tests demonstrated higher intercorrelations than the correlations between them and the relevant test accuracy scores. A robust Confidence factor emerged when both exploratory and confirmatory factor analytic models were employed. This factor was positively related to, yet distinct from, the relevant Accuracy factors (see Kleitman 2008; Kleitman et al. 2011; Pallier et al. 2002; Stankov 1999 for reviews). Recently replicated in a large cross-cultural study (Stankov et al. 2012), these findings are in support of a broad Confidence factor present in the cognitive domain that is distinct, yet related to, ability. This finding is most consistently derived from tests of fluid and crystallised Intelligence (Gf and Gc) such as Raven's Progressive matrices and Vocabulary Tests respectively. In addition to controlling for general Intelligence, the use of Gf and Gc tests here will therefore help to best capture this confidence trait.

Similarly, scores obtained on different cognitive tests converged on a broad Bias factor (Kelemen et al. 2000; Kleitman 2008; Schraw et al. 1995; Stankov and Crawford 1996; West and Stanovich 1997). For example, Schraw et al. (1995) found that the bias scores of 143 undergraduate students, obtained from seven cognitive ability tests—spanning domains of general knowledge, mathematics, spatial judgement and reading comprehension—converged on a single factor when submitted to Principal-Component Analysis. That is, irrespective of the nature of the cognitive tasks, and their difficulty, people who tend to be overconfident on one type of task, tend to be overconfident on other types of tasks relative to the others. Likewise, people who tend to be underconfident on one type of task, tend to be underconfident on other types of tasks. These findings are in support of a broad Bias factor.

Limited investigation into individual differences in discrimination has resulted in mixed findings. For example, Schraw et al. (1995) correlated discrimination scores across eight cognitive tests spanning different intelligence domains. They only found two significant positive correlations. However, in a second experiment, they found a strong pattern of positive intercorrelations—eight out of ten significant—utilising five general knowledge tests. Similarly, Kelemen et al. (2000) found four significant correlations among discrimination scores obtained from three learning and general knowledge based tests administered together at two time points, 1 week apart. However, they found no significant correlations in a second experiment using similar measures. Furthermore, metacognitive skills related to discrimination consistently demonstrate domain general properties (e.g., Veenman et al. 2004). As such, there is limited support that individuals who are better able to discriminate in one domain, tend to better discriminate in others.

Thus, confidence, bias and CAQ acquired in our novel decision-making test—described next—from which our decision tendency variables will be derived, should reflect these general metacognitive factors. We would therefore expect them to respectively converge with confidence, bias and CAQ acquired from the battery of cognitive tests when submitted to exploratory factor analysis. Given the Confidence Assumption, this expectation holds important theoretical and applied implications. If confidence levels preceding decision-making and acquired in cognitive tests reflect the same trait, this would highlight the generality of metacognitive confidence and advocate the acquisition of confidence ratings in applied settings for predictive purposes. However,

while the empirical evidence is in strong support that this is the case, no study to date has examined it. The present study will therefore also aim to bridge this gap by examining whether confidence, bias, and CAQ scores, within our novel decision-making test will converge with these scores acquired from a battery of standard cognitive ability tests.

#### Additional control variables

In addition to controlling for general Intelligence, other relevant stable individual differences variables will be collected as important control variables. Firstly, we will collect a broad measure of Personality. This is because various personality traits may be influential in the decision-making processes: such as Conscientious individuals being more hesitant decision-makers. Similarly, we will collect a cognitive style measure: the Need for Closure. This scale targets an individual's aversion to uncertainty. It is important to control for this variable as individuals scoring higher on this cognitive style might be more incompetent decision-makers, or might generally indicate higher confidence. Finally, gender and age will be collected as each has demonstrated relationships with the confidence trait. While not a consistent finding (e.g., Bavolar 2013), males and older people are sometimes found to be more confident, and overconfident, than females and younger people respectively (see Boekaerts and Rozendaal 2010; Crawford and Stankov 1996). These variables will therefore be controlled for in our analyses.

#### Medical decision-making test

To compute the decision-making tendency variables and address the related hypotheses, we constructed a test to provide a variety of stimuli, holding decision options constant, within a context suitable for naïve participants that does not require specialised or prolonged training. McKenzie's (1998) original medical test provided an excellent framework for this. Mentioned earlier, McKenzie's participants learned about symptoms associated with two fictitious illnesses, made subsequent diagnoses of new patient profiles, and decided whether to send patients to treatment. This process, including the labels and use of fictitious illnesses, was adopted for our purposes. However, rather than 2 ambiguous profiles, 42 different profiles, each with a single correct answer, were completed after learning. Furthermore, a background story was constructed to make participants aware of the outcomes that their decisions might lead to. Here, they were also informed that the illnesses were fictitious. The learning phase of McKenzie's (1998) task, where participants learned until they could accurately identify which illness each symptom was associated with, were also altered for the present study. Rather, participants in the present study were given a limited amount of time to learn as much as they could. This was done to ensure that a level of uncertainty would be present when diagnosing and treating new patients, even though those patients could be accurately diagnosed. Ultimately, our novel design required participants to diagnose and treat 42 patients based on the novel symptom combinations they had learned, as well as indicate their confidence in each diagnosis. Each patient profile could be accurately diagnosed and, for each, participants could immediately treat (congruent decision) or administer a blood test (incongruent). Thus, like Koriat & Goldsmith, decisions can be categorised based on their preceding judgement accuracy and the frequency of each category used to compute the decision tendency variables. Hereafter, this test will be referred to as the Medical Decision-Making Test (MDMT).

## Aims and hypotheses

Utilising the MDMT and the set of novel and traditional variables, the present study tests a broadly applicable framework for the study of individual differences in decision-making. The first aim is to determine the generality of confidence and its calibration across decision and cognitive domains using an individual differences methodology. It is hypothesised that confidence, bias, and CAQ scores, derived from the MDMT and three cognitive tests, will converge on single factors when submitted to respective exploratory factor analyses (metacognitive generality hypothesis). This will be to ensure that these metacognitive indices acquired within the MDMT are an accurate reflection of stable metacognitive factors observed elsewhere. Our following aim was to establish the internal consistency and validity of the novel decision-making tendencies (optimal, realistic, incompetent, hesitant and congruent) within the decision-making task. It is hypothesised that the decision tendency variables will demonstrate acceptable internal reliability estimates, and that they will be differentially predicted by confidence, bias and CAQ as described by the Confidence Assumption and outlined in the “[Introduction](#)”. The latter hypothesis will be investigated utilising multiple regression to control for diagnostic accuracy, intelligence, personality, cognitive styles, gender, and age.

## Method

### Participants

In total, 193 first year psychology students at the University of Sydney participated in return for partial course credit (114 female, 79 male,  $M_{\text{age}}=19.41$  years, age range: 17–39 years).

### Materials

All tasks were programmed using the online questionnaire program ‘Surveygizmo 3.0’ and completed in an Internet Explorer 8 browser. In addition to the tasks described below, the battery also included the Big-6 Personality Inventory (Saucier 2008), chosen to remain consistent with previous research on individual differences in confidence (see Kleitman 2008), and the Need for Closure (Roets and Van Hiel 2011). However, along with gender and age, neither contributed significantly to the findings and all of the presented results were demonstrated incrementally over and above these measures (results available on request from the first author).

*Medical decision-making test (MDMT)* In this test, participants adopt the fictitious role of a specialist in deadly paralympnal illnesses of which there are two kinds: puneria and zymosis. Each illness has a unique but potent treatment such that correct administration of a treatment will save patients, but incorrect treatment will kill them. After learning about the scenario, participants were given 10 min<sup>3</sup> to learn from three tables, presented on a single A4 page, how eight symptoms were associated with three illness states: puneria, zymosis, and paralympnal free. Each table presented the symptoms experienced by 20 patients who had one of the three illnesses (See Appendix). As in McKenzie’s task, symptoms occurred in 85 % or 15 % of each illness state. Symptoms that occurred in 85 % of the learning profiles could be used to

<sup>3</sup> Ten minutes learning time was deemed appropriate and tested in a pilot study ( $N=19$ ). Feedback received indicated that 10 min was a sufficient amount of time to learn and that the instructions had been understood. 10 min was therefore retained as the learning time in the present study.

diagnose patients with that illness state. Once 10 min had elapsed, tables were removed and participants progressed to the test phase. The learning phase and diagnoses that follow is basically a multiple cue association learning, memory and reasoning test.

In the test phase, participants completed 42 patient profiles in a randomised order. Each patient profile was different to those used by McKenzie as they were constructed such that a definitive diagnosis could be made. For each patient, participants had to make a diagnosis, indicate their confidence in each diagnosis from 30 to 100 % in 10 % increments,<sup>4</sup> and make a treatment decision: a congruent option to immediately treat the patient with the diagnosis cure/release if diagnosed as paralympic free, or an incongruent option to request a blood test. Participants were informed that a blood test would accurately diagnose and treat the patient, but about 50 % of infected patients die waiting for the results and frequent testing strains available resources and slows the procedure. Semi-structured interviews conducted during a pilot study ( $N=19$ ) confirmed that this scenario was not distressing for students. No feedback was provided, and each novel profile was completed in the same way depicted as the example shown in Fig. 2.

The following variables were calculated from the diagnosis and diagnostic confidence: diagnostic accuracy, confidence, bias, bias squared and CAQ. The five decision tendency variables were calculated from diagnostic accuracy and the final decision as described in the “Introduction”. For example, incompetent decision tendencies were computed as the frequency of patients incorrectly treated/released (False alarms) divided by the number of patients diagnosed incorrectly.

*Raven’s advanced progressive matrices (APM; Raven 1938-65)* This test included 20 items each presenting a  $3 \times 3$  display of abstract figures following a pattern both horizontally and vertically. The bottom right figure is left blank and participants choose which of 8 alternative figures will complete the display. Participants indicated their confidence in each answer by selecting a value from 10 to 100 %, in 10 % increments. APM is a gold-standard measure of fluid intelligence (Gf), and has been shown to possess good internal alpha reliability estimates, typically greater than 0.80 (Raven 1938-65).

*Esoteric analogies test (EAT; Stankov 1997)* This test required participants to complete 24 verbal analogies by selecting one of four alternative words that share the same relationship with a target word as that of an original pair. For example, FIRE is to HOT as ICE is to: POLE, COLD\*, CREAM, or WHITE. Confidence ratings for each answer were obtained by typing some value between 25 and 100 %. This test requires both fluid and crystallised intelligence (Gf and Gc), and has been shown to possess internal alpha reliability estimates acceptable for research purposes of 0.66 to 0.76 (Kleitman 2008; Kleitman and Stankov 2007; Want and Kleitman 2006).

*Vocabulary test (VT; Stankov 1997)* This test involved completing 18 items in which participants select which of five words or short phrases has the same meaning as a target word. For example, FEIGN is to: PRETEND\*, PREFER, WEAR, BE CAUTIOUS, SURRENDER. Confidence ratings were obtained by typing some value between 20 and 100 %. This test is a distinct marker of Gc, and has been shown to possess internal alpha reliability estimates acceptable for research purposes, ranging between 0.67 and 0.81 (Kleitman 2008; Stankov and Crawford 1997).

<sup>4</sup> Given that the confidence scales for the MDMT and the APM were presented as categorical responses representing 10 % increments (e.g., 50 %, 60 % etc.), we chose to represent the lower bound of the scale with the value that was closest to the actual probability associated with guessing. That is, 30 % was used as the closest estimate to the 33 % chance of being correct in the MDMT, and 10 % as the closest estimate to the 12.5 % chance in the APM. Given the small deviations here (3 % and 2.5 %) and previous results in our laboratory, there was no reason to suspect that this would alter the results.

	Symptom	Present?
Coughing		
Headache		<input type="checkbox"/>
Rash		
Nausea		<input type="checkbox"/>
Swollen glands		
Sneezing		<input type="checkbox"/>
Diarrhoea		
Congestion		

Most likely diagnosis?

Puneria	Zymosis	Paralymphal Free
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

How confident are you that your diagnosis is correct?

30%	40%	50%	60%	70%	80%	90%	100%
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Your final decision?

Direct diagnosis decision	Request blood test
<input type="checkbox"/>	<input type="checkbox"/>

**Fig. 2** MDMT example question. Direct diagnosis decision would initiate treatment or release as paralymphal free based on the diagnosis

Each participant had the following variables computed for all three cognitive tests<sup>5</sup>: test accuracy, confidence, bias, bias squared and CAQ.<sup>6</sup>

**Procedure**

In groups of up to 10, participants received instructions upon arrival and completed basic demographic and English proficiency questions. The MDMT was then completed first to ensure that confidence ratings in this test were not influenced by performance or exposure to these ratings in the cognitive tests. The remaining tasks were counterbalanced. Testing was self-paced and completed in approximately 60 to 90 min.

**Results**

**Preliminary analysis**

*Missing values analysis* Other than the APM, which had 10 % of its data missing, no more than 5 % of data was missing for any other variable of interest. All missing values were the

<sup>5</sup> Marker tests were selected to be consistent with the Horn-Cattell Gf/Gc theory of intelligence (Horn and Cattell 1982): a hierarchical model defining intelligence in terms of two broad independent abilities (Carroll 1993). Fluid intelligence (Gf) reflects basic reasoning abilities, whilst crystallized intelligence (Gc) reflects the ability to learn and use information acculturated through education and experience.

<sup>6</sup> CAQ cannot be calculated when overall accuracy equals 0 % or 100 % because average confidence for correct or incorrect items is not assigned by the participants, thus is not available (Shaughnessy 1979). CAQ was therefore calculated for any individual with at least one correct and incorrect answer. CAQ scores were not computed for five participants in the MDMT, one in Raven’s Advanced Progressive Matrices, and two in Esoteric Analogies due to 100 % accuracy (no participants scored 0 %).

result of software errors. Exceptions were missing values for CAQ, Incompetent and Hesitant tendencies, which were the result of computational requirements. All analyses were therefore conducted in a pairwise fashion.

*Descriptive statistics and reliabilities* Descriptive statistics and Cronbach's alpha reliability estimates (where applicable) for all variables are presented in Table 3.

The pattern of results for the decision variables suggested that participants generally demonstrated reasonable decision-making ability. The means of optimal and realistic decision tendencies were 0.51 and 0.62 (out of 1) respectively. On average, participants were therefore able to accurately diagnose and treat over half of the patients and appropriately test a further 11 % of misdiagnosed patients. Moreover, participants were generally more likely to treat than test with a mean of 0.65 for congruent decision tendencies. This was also evident in the higher mean of incompetent than hesitant decision tendencies. Indeed, individuals incompetently killed 50 % of their misdiagnosed patients by incorrectly treating them, on average. As hypothesised, high reliability estimates were obtained for all decision-making variables or their key components with respect to incompetent and hesitant decision tendencies. These results suggested that there were strong within-test consistencies in decision-making tendencies.

Mean accuracy scores were highest for the MDMT, followed by the EAT, APM, and the VT, and their Cronbach's alpha reliability estimates were in an acceptable range. These results were similar to those found in previous research with Australian undergraduates using these same cognitive tests (Kleitman 2008; Kleitman and Stankov 2007). Supporting its future use, the reliability estimate for MDMT diagnostic accuracy was greater than those of the ability tests.

Mean confidence ratings for the cognitive tests were comparable with those in the studies cited above. Furthermore, mean MDMT diagnostic confidence fell within this range. Consistent with previous research, reliability estimates for all confidence ratings were high (see Stankov and Kleitman 2008 for a review).

Also consistent with this research, a slight overconfidence bias was evident in the ability tests. Mean MDMT diagnostic bias was close to perfect calibration, but in the underconfidence region. However, as was the case for MDMT diagnostic accuracy, this is to be anticipated when test accuracy approaches 80 % (Lichtenstein and Fischhoff 1977). With the exception of the VT, internal reliability estimates of the bias scores were satisfactory. The low VT estimate is undoubtedly linked to its low accuracy reliability estimate (Kaplan and Saccuzzo 2005). Again in support of its future utility, the MDMT yielded the greatest bias reliability estimate.

All mean CAQ scores were greater than 0, indicating that, on average, participants appropriately adjusted their confidence between correct and incorrect answers. CAQ for the cognitive tests converged around 1, but was lower for the MDMT, indicating that participants found it somewhat more difficult to discriminate between correct and incorrect answers in this test.

*Additional analyses* To ensure task order did not have an effect, an analysis of variance was conducted on each variable comparing differences across counterbalanced conditions. No significant differences emerged. Furthermore, to allow for an analysis of the MDMT by combining the results across all test profiles—puneria, zymosis, and paralympnal free—diagnosis and decision preferences were examined. Consistent with the high reliability estimates for the derived scores, nothing of concern emerged and a combined analysis was considered appropriate.

**Table 3** Descriptive statistics and reliabilities

	<i>n</i>	<i>M</i>	<i>SD</i>	$\alpha^a$	Range
Decision tendencies					
Optimal	186	0.51	0.17	0.84	0.07–0.86
Realistic	186	0.62	0.13	0.74	0.24–0.90
Incompetent	181	0.50	0.29	0.83	0.00–1.00
Hesitant	186	0.32	0.18	0.81	0.03–0.88
Congruent	186	0.65	0.17	0.84	0.12–0.95
Accuracy (% correct)					
MDMT	186	75.67	15.90	0.86	26–100
APM	173	59.71	20.36	0.81	5–100
EAT	192	70.65	14.96	0.70	17–100
VT	193	49.28	14.90	0.65	11–89
Confidence (average %)					
MDMT	186	70.18	8.87	0.93	43–92
APM	173	72.03	14.50	0.93	29–99
EAT	192	77.08	10.63	0.87	50–100
VT	193	63.02	13.74	0.89	20–98
Bias					
MDMT	186	−0.05	0.16	0.86	(−0.43)–0.49
APM	173	0.12	0.15	0.58	(−0.28)–0.54
EAT	192	0.06	0.15	0.65	(−0.23)–0.59
VT	193	0.14	0.12	0.38	(−0.16)–0.53
Discrimination (CAQ)					
MDMT	181	0.52	0.62	n/a	(−1.12)–2.95
APM	172	1.22	0.62	n/a	(−0.43)–4.38
EAT	190	0.99	0.60	n/a	(−0.55)–3.77
VT	193	1.13	0.39	n/a	(−0.46)–1.82

*MDMT* medical decision-making test, *APM* Raven's advanced progressive matrices, *EAT* esoteric analogies test, *VT* vocabulary test

<sup>a</sup>A set of 42 binary items (1 or 0), representing each MDMT item, were constructed for each decision tendency variables in order to assess their reliability. Each item in a set was scored 1 if the outcome of the participant's decision on that MDMT item was in the numerator of the variable's equation, and 0 otherwise. For example, a Hit or Correct Rejection would be scored 1 in the set of Realistic tendency items, whereas a Miss or False alarm would be scored zero

### Exploratory factor analyses

*Confidence* Table 4 summarises the correlation coefficients and the results of an Exploratory Factor Analysis (EFA; Principal Components [PC] with PROMAX-rotation), constrained to two factors, performed on accuracy and confidence scores.<sup>7</sup>

<sup>7</sup>An initial exploratory factor analysis (EFA; Principal Components with PROMAX-rotation) was run without constraints. The latent root criterion suggested 3 factors, but the scree plot clearly suggested the presence of two distinct factors only. Considering the moderate to high intercorrelations among many variables, a single factor solution was also run. However, a single factor only accounted for 44.158 % of the variance (a drop of almost 16 %), and one communality was below 0.20. Hence, the two-factor solution was retained and used in the final analysis.

These results were in support of broad Ability and Confidence factors. Accuracy scores were significantly and positively correlated with each other, as were confidence scores. Despite particularly strong correlations between the APM and VT accuracy and confidence scores, the remaining correlations between the accuracy and confidence scores were positive, but weaker in general. In support of the MDMT's divergence from typical cognitive tests, diagnostic accuracy was not significantly correlated with any cognitive test confidence scores.

The EFA (PC) provided further support, with the two factors explaining 60.04 % of the common variance. Communalities for all scores were high, except for a low MDMT accuracy communality, again indicative of the test's divergence from the typical cognitive tests. Notably, MDMT diagnostic confidence did not demonstrate such divergence. Using 0.30 as the cut-off criterion for a meaningful factor loading, Factor 1 was defined by all of the accuracy scores, as well as a meaningful loading from VT confidence. In support of the postulated hypothesis, Factor 2 was defined by each of the confidence scores, but also a considerable loading from APM accuracy. As expected, the MDMT diagnostic confidence had a high loading on this factor, and the two factors were positively correlated with each other ( $r = .47$ ). Thus, even with the two cross-loadings, these results were in support of two broad and distinct, but related, factors: General Cognitive and Diagnostic Ability, and Confidence.

*Calibration* Table 5 summarises the relevant correlation coefficients and results of two EFAs (PC) performed on (1) the bias scores, and (2) the CAQ scores.

The results were in support of broad Bias and CAQ (Discrimination) factors. All correlations between bias scores (above the diagonal) were positive and significant. The same was evident for all CAQ intercorrelations but one, between APM and VT (below the diagonal). Submitting each set of scores to EFA (PC) clearly revealed single factors explaining 54.01 % and 37.87 % of the common variance for bias and CAQ respectively. Factor loadings and communalities were all moderately high for both calibration indices.

Overall, these results supported the hypothesis that metacognitive variables within the MDMT would reflect general metacognitive factors consistently observed in individual differences research and that these factors extend beyond typical cognitive tests.

#### Predictive validity: regression analyses

Predictive validity was investigated via relevant sets of multiple regression analyses. Set one regressed each decision tendency variable in a hierarchical fashion on (1) diagnostic accuracy, (2) general cognitive ability—Intelligence; computed as the mean of the three cognitive ability tests—and (3) diagnostic confidence. Set two regressed each decision tendency variable on diagnostic bias. Optimal and Realistic decision tendencies were then regressed on the square of bias in a subsequent step to investigate the non-linear component: whether the tendency to minimise decision errors would diminish with increasing bias in the form of a quadratic trend. Set three regressed each decision tendency variable on diagnostic CAQ. It was not possible to include diagnostic accuracy or Intelligence in the calibration sets due to these indices being derived from accuracy scores. To reiterate, all analyses were conducted controlling for personality, need for closure, gender, and age, all of which did not contribute statistically significantly to the results and have been omitted here. The results of these analyses can be seen below (Table 6).

*Confidence* As expected an increase in diagnostic confidence predicted a statistically significant incremental increase in congruent, optimal and incompetent decision tendencies (19 %,



**Table 4** Confidence and accuracy intercorrelations and EFA results

		Pearson <i>r</i> correlations						Factor loadings			
		2	3	4	5	6	7	8	1	2	<i>h</i> <sup>2</sup>
Accuracy											
1	MDMT	0.31**	0.39**	0.25**	0.23**	0.13	0.08	0.13	<b>0.51</b>	-0.04	0.24
2	APM		0.51**	0.36**	0.26**	0.68**	0.35**	0.30**	<b>0.31</b>	<b>0.53</b>	0.54
3	EA			0.58**	0.15*	0.29**	0.39**	0.46**	<b>0.84</b>	-0.00	0.70
4	VT				0.05	0.16*	0.37**	0.63**	<b>0.96</b>	-0.20	0.78
Confidence											
5	MDMT					0.43**	0.36**	0.28**	-0.25	<b>0.82</b>	0.55
6	APM						0.57**	0.43**	-0.08	<b>0.93</b>	0.80
7	EA							0.65**	0.23	<b>0.64</b>	0.60
8	VT								<b>0.58</b>	<b>0.31</b>	0.60

Factor loadings >.30 are in boldface. *h*<sup>2</sup> communality, *MDMT* medical decision-making test, *APM* Raven’s advanced progressive matrices, *EA* esoteric analogies test, *VT* vocabulary test

\**p*<0.05. \*\* *p*<0.01. \*\*\* *p*<0.001

10 % and 9 % respectively) and a decrease in hesitant tendencies (17 %), over and above diagnostic accuracy and Intelligence. Somewhat unexpectedly, diagnostic confidence also predicted a statistically significant incremental increase in realistic decision tendencies (8 %).

*Bias* As predicted, increasing bias predicted a statistically significant increase in congruent and incompetent decision tendencies, and a decrease in hesitant decision tendencies (3 %, 9 % and 4 % respectively). Increasing bias also predicted a significant decrease in optimal tendencies (12 %).

Also as predicted, the square of diagnostic bias incrementally and negatively predicted, optimal and realistic decision tendencies (3 % and 6 % respectively). These regression equations were further examined to assess our hypothesis these relationships were described by functions turning on perfect calibration (bias=0) (see Cohen et al. 2003, for more details). While realistic decision tendencies were greatest only slightly below perfect calibration (bias=-0.07), optimal decision tendencies were greatest considerably below perfect calibration

**Table 5** Bias and CAQ intercorrelations and EFA results

Score	Pearson <i>r</i> correlations				Bias factor loadings		CAQ factor loadings		
	1	2	3	4	1	<i>h</i> <sup>2</sup>	1	<i>h</i> <sup>2</sup>	
1	MDMT	-	0.34**	0.39**	0.29**	<b>0.68</b>	0.44	<b>0.64</b>	0.40
2	APM	0.16*	-	0.48**	0.41**	<b>0.77</b>	0.59	<b>0.54</b>	0.31
3	EA	0.17*	0.18*	-	0.40**	<b>0.78</b>	0.62	<b>0.64</b>	0.42
4	VT	0.20**	0.11	0.20**	-	<b>0.69</b>	0.51	<b>0.63</b>	0.39

Correlations above the diagonal are between bias scores. Correlations below the diagonal are between CAQ scores. Factor loadings >.30 are in boldface. *h*<sup>2</sup> communality, *MDMT* medical decision-making test, *APM* Raven’s advanced progressive matrices, *EA* esoteric analogies test, *VT* vocabulary test

\**p*<0.05. \*\* *p*<0.01. \*\*\* *p*<0.001

**Table 6** Multiple regression analyses of decision tendencies on metacognitive indices

Predictor set	MDMT decision tendencies									
	Optimal		Realistic		Incompetent		Hesitant		Congruent	
	$\Delta R^2$	$\beta$	$\Delta R^2$	$\beta$	$\Delta R^2$	$\beta$	$\Delta R^2$	$\beta$	$\Delta R^2$	$\beta$
Confidence										
Diagnostic accuracy	0.38***	0.62	0.09***	0.31	0.03*	-0.17	0.00	-0.02	0.01	0.09
Intelligence	0.02*	0.14	0.03*	0.17	0.00	0.00	0.03*	-0.19	0.02*	0.16
Diagnostic confidence	0.10***	0.33	0.08***	0.28	0.09***	0.32	0.17***	-0.43	0.19***	0.44
Bias										
Diagnostic bias	0.12***	-0.35	0.01	-0.11	0.09***	0.30	0.04***	-0.21	0.03*	0.16
Diagnostic bias <sup>2</sup>	0.03*	-0.18	0.06**	-0.26	-	-	-	-	-	-
CAQ										
Diagnostic CAQ	0.06**	0.25	0.12***	0.35	0.43***	-0.66	0.02*	-0.15	0.00	-0.01

Predictors in each model set are entered as steps in the same order as displayed

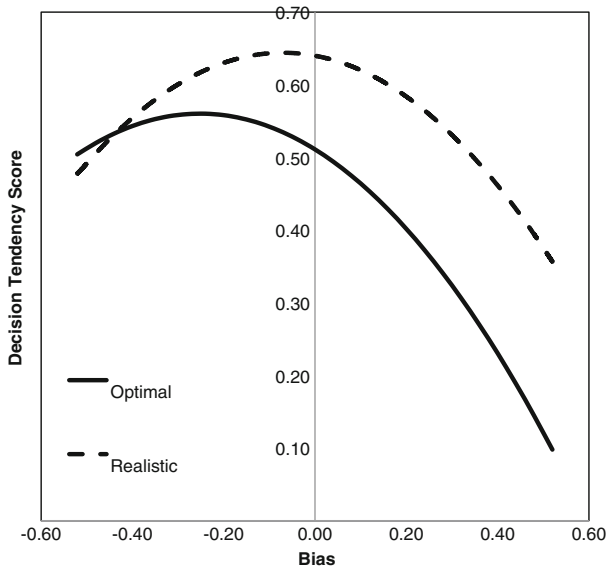
\* $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$

(bias = -0.25). A plot of the curves described by these regression equations, within the range of observed bias scores, can be seen below (Fig. 3).

*CAQ* As hypothesised, an increase in diagnostic CAQ predicted a statistically significant increase in optimal and realistic decision tendencies (6 % and 12 %) and a decrease in incompetent and hesitant decision tendencies (2 % and 43 %).

## Discussion

Decision-making is a complex, albeit important, process that is typically studied under experimental conditions. Despite much being learned from this perspective, a lack of individual differences research limits the identification of tendencies, hampering future research and limiting the possibility of using these findings for predictive and selection/profiling purposes. We therefore outlined a general framework to capture individuals' decision tendencies, albeit within a given scenario, and developed the Medical Decision-making Test (MDMT) to test it. That is, by extending Koriat and Goldsmith's (1996) model and utilising an individual differences approach, we used a modified McKenzie's (1998) medical test to examine a set of novel decision-making variables. Finally, we addressed a number of novel questions: Can the Confidence Assumption be confirmed from an individual differences perspective; Do confidence and calibration factors generalise to decision-making scenarios; and does a quadratic function best represent the relationship between bias and optimal decision-making. This was done controlling for general Intelligence, personality, cognitive styles, gender and age. Hence, utilising the novel MDMT, we sought to determine evidence for the existence and consistency of five decision tendencies and examine the predictive validity of metacognitive confidence and its calibration on them.



**Fig. 3** Each line in the figure represents the relationship between optimal (*solid line*) and realistic (*dotted line*) tendencies with bias defined by the regression coefficients of these variables on MDMT bias and bias squared. The *vertical axis* represents decision tendency value. The *horizontal axis* represents bias and has been shortened to include the range of bias scores observed in the present study

### Metacognitive generality

To examine whether the cognitive and metacognitive variables derived from the MDMT reflect general factors, we used several Exploratory Factor Analyses to determine whether these indices would converge across the MDMT and cognitive tests. Strong to moderate evidence for the metacognitive generality hypothesis (Veenman et al. 2004) was found for each index: Confidence, bias, and CAQ scores from all tests clearly converged on three respective factors. Regarding confidence, despite two cross-loadings—undoubtedly resulting from the use of only one ‘pure’ marker for each cognitive domain (see Carroll 1993 for a review)—the results were in support of two distinct albeit related factors: general Confidence and Ability. Bias and CAQ scores clearly converged on a single Bias and Discrimination factor, respectively. However, aligning with the results of Kelemen et al. (2000), and Schraw et al. (1995), support for a broad Discrimination factor was noticeably weaker. Importantly, all three indices derived from the MDMT clearly converged on their intended factors.

It is worth noting that the accuracy of the MDMT diverged from the accuracy of the cognitive tests employed in this study, providing evidence for its divergent validity. Thus, the MDMT did not simply measure Gf/Gc cognitive abilities (Cattell 1987). However, the relative confidence and calibration scores from the MDMT clearly converged together with the confidence and calibration measures across domains. Thus, other than supporting the validity of these measures within the MDMT, these results provide support for some important implications: (i) Individuals who are more confident in the accuracy of decisions in typical cognitive tests tend to be more confident in their judgments in decision-making contexts; (ii) Relative to others, individuals more over/underconfident in the accuracy of decisions in cognitive tests tend to be respectively over/underconfident in the accuracy of their judgements prior to making a decision; (iii) Individuals who better discriminate between correct and

incorrect answers in cognitive tests tend to better discriminate between correct and incorrect judgments prior to making decisions. However, the extent of this third overlap clearly requires further scrutiny. Collectively, these results further support the hypothesis that stable individual differences in metacognitive confidence and its calibration generalise beyond typical cognitive tests to decision-making conditions.

### Decision tendencies

By extending Koriat and Goldsmith's (1996) model we proposed a general framework to classify decisions and, for the first time, capture decision-making tendencies within a given scenario. Specifically, optimal tendencies captured the likelihood of individuals making accurate judgements and congruent decisions following them, and realistic tendencies of appropriately making congruent and incongruent decisions following correct and incorrect judgements. Incompetent and hesitant decision tendencies captured the likelihood of making congruent or incongruent decision errors following incorrect and correct judgements respectively. Finally, congruent tendencies captured individual tendencies to make decisions aligned with their judgements in general. All decision tendency variables demonstrated excellent internal consistency. These variables capture individual differences in aspects of decision-making which did not share a meaningful relationship with personality and cognitive styles. The present results therefore suggest that various personality traits, such as Conscientiousness, Extraversion or Neuroticism, have little to do with the nature of an individual's decision-making tendencies.

### Predictive validity

Our subsequent aim was to investigate the differential predictive validity of the metacognitive measures on the decision tendencies in the manner predicted by the Confidence Assumption. Support existed for each hypothesis.

*Confidence* The Confidence Assumption predicts that the greater a person's confidence in the appraisal of their judgement accuracy, the more likely they are to engage in the congruent rather than incongruent decision-making act (e.g., DeMarree and Petty 2007; Slovic et al. 1977). The results of the hierarchical regression analyses strongly supported this hypothesis. As expected, individuals more confident in their diagnoses scored higher on congruent and incompetent tendencies, and lower on hesitant tendencies. That is, they tended to make more congruent decisions overall, following correct and incorrect judgements (diagnoses). Additionally, more confident individuals scored higher on optimal and realistic decision tendencies. However, mean diagnostic accuracy in the MDMT was high (70 %). Decisions therefore followed correct diagnoses more often than not: on average, 70 % of the time. When judgement accuracy is high, making congruent decisions, more often than not, would generally lead to Hits, inflating optimal and realistic decision tendencies. Increasing the difficulty level of the scenario will therefore be important for future research. Regardless, these results were demonstrated incrementally over diagnostic accuracy, personality, cognitive styles, gender, age, and even general cognitive ability, which itself is a powerful predictor of real-life outcomes (e.g. Hunter 1986). This indicates that metacognitive confidence—a construct typically ignored in preference of Intelligence and/or other popular measures—is an important psychological construct to be included in the study of decision-making processes. This provides additional support for recent findings that confidence is an important incremental predictor of real-world outcomes, such as academic achievement (Stankov et al. 2012) and retirement planning (Parker et al. 2012).

*Bias* It was hypothesised that overconfidence would lead to incompetent decision tendencies, underconfidence to hesitant tendencies, and deviation from perfect calibration (bias=0) to poorer optimal and realistic tendencies in the form of a quadratic trend. The bias regression analyses supported this prediction. Increasingly overconfident—or decreasingly underconfident—individuals tended to score higher on incompetent and lower on hesitant decision tendencies. Furthermore, individual decision-making was most optimal and realistic as bias respectively approached slightly below, and near perfect calibration in a quadratic fashion. Optimal tendencies were additionally and negatively associated with linearly increasing overconfidence. No doubt this was the result of strong correlations between diagnostic accuracy and optimal tendencies ( $r=.62; p<.01$ ), and diagnostic accuracy and diagnostic bias ( $r=-.85; p<.01$ ). Both relationships were expected: the former the result of optimal tendencies being contingent on the frequency of correct diagnoses; the latter describing a persistent finding known as the hard–easy effect (Juslin et al. 2000; Lichtenstein and Fischhoff 1977). This then explains why optimal tendencies were greatest when bias was negative: below perfect calibration of zero. All other results were in support of the postulated hypotheses, supporting our decision framework and advocating the use of the trend analysis, and quadratic function in particular, when examining relationships between bias and measures of optimal tendencies. It simply helps to address the complex nature of this calibration index.

However, these results also call into question the utility of the bias score as an individual differences variable relative to its constituent components of confidence and accuracy. The proportion of variance in the decision variables accounted for by bias was considerably lower than that accounted for by accuracy and confidence. Even the novel predictions described by the quadratic trends had very little to add in this respect. This point has been voiced by Stankov et al. (2013), who suggest that the bias score is best suited as a convenience measure to show groups differences. We are therefore of the opinion that future researchers should always consider interpreting accuracy and confidence as separate variables, controlling for one another, in conjunction with their use of the bias score.

*CAQ* In relation to our final metacognitive predictor, we hypothesised that better discrimination between correct and incorrect judgements should lead to more optimal decision-making and a reduction in decision errors. The regression analyses provided support for this hypothesis. Better discriminators scored higher on optimal and realistic tendencies, and lower on incompetent and hesitant tendencies. Additionally, discrimination accounted for a far greater proportion of the variance in the primary error tendency, incompetent tendencies (43 %), than any other variable. This high percentage may be due to the fact that incompetent tendencies are based on incorrect diagnoses only, and people low on the discrimination index had obvious difficulties discriminating between their correct and incorrect answers. Overall, these results again support that the decision tendency variables are a valid representation of their intended constructs and that discrimination, indexed by CAQ, is an important measure for studying individual differences in metacognitive calibration.

### Limitations and future directions

While the present findings are encouraging, certain limitations of the current design need to be considered. Firstly, the nature of the MDMT restricts the generalisability of this novel framework and the tendency variables. It is unclear whether these findings will extend to conditions of objective uncertainty or ambiguity when outcome probabilities are clearly defined or unknown, respectively. Furthermore, despite the high internal reliability

estimates obtained for the decision tendencies, other psychometric properties, especially temporal stability, predictive validity, and the stability of their ground frequencies under varying conditions, will require investigation. Furthermore, real-world decision-making often requires identifying and gathering information about decision options. It would therefore also be of interest to examine whether these results extend to broader decision-making contexts, including actual medical students and other real-world contexts, such as military, business, political, or nuclear emergency scenarios. This may be possible to achieve utilising the MDMT design in similar or more advanced formats. For example, fire-fighter decision tendencies could be assessed in scenarios requiring judgements about where and what the source of a building fire might be, followed by a standard set of decisions about how to tackle that scenario: Such as proceeding to the judged source with extinguishers suitable for ordinary combustible or metal fires, or requesting advice from more experienced personnel. Similarly, assessing doctors with *real* medical scenarios, such as evidence-based constructed patient vignettes, would determine whether these tendencies remain stable when based on content-related knowledge and experience. While the MDMT appears useful for generating novel decision-making variables in a general, non-medical, population, future research should aim to assess a more diverse range of decision contexts.

Future research may wish to address the limited utilisation of only three distinct cognitive markers. Other than those already discussed, a further consequence of this was that the role of general cognitive ability (Intelligence)—undeniably crucial to optimal decision-making in general—might not have been appropriately represented. Administering a broader selection of tests, such as working memory or/and visual intelligence, would be necessary to appropriately consider a fuller range of cognitive abilities. This will also allow for the use of Confirmatory factor analytic techniques to more stringently examine the expected structure of the metacognitive factors across tests. In line with previous research, we recommend that future studies utilise at least three markers for each cognitive domain.

Finally, additional studies into a variety of psychometrically sound individual tendencies related to decision-making might yield great benefit. For example, MDMT like tests could help establish the point between and within individuals at which the switch from incongruent to congruent decision-making occurs. Furthermore, a large collection of variables purported to underlie individual differences in decision-making presently exist and continue to emerge, such as measures of rationality and normative responding (see Appelt et al. 2011 for a review). It may be of future benefit to examine how these measures interrelate as well as relate to our decision tendencies and other metacognitive variables. For example, it may be the case that more rational individuals demonstrate more realistic decision tendencies. Such endeavours offer new avenues of research and can provide a clearer and more unified approach to the study of individual differences in decision-making.

## Implications and conclusions

To this end, the present study provides a number of important theoretical and applied implications for metacognitive and decision-making research. On a theoretical level, this research extended knowledge about the generality and predictive validity of metacognitive judgement confidence and its calibration. Our results provide strong support that individual differences in metacognitive confidence and its calibration, so consistently observed within the cognitive domain, generalise to decision-making under conditions of subjective uncertainty. Furthermore, the present results revealed reliable decision-making tendencies, and that these tendencies share meaningful predictive relations with confidence and its calibration.

Three applications immediately present themselves. Firstly, interventions designed to develop and improve confidence and its calibration might have significant impacts on decision-making. Secondly, given the generality observed here, confidence in any domain might be indicative of decision-making tendencies elsewhere. For example, including confidence ratings in cognitive tests used for pre-employment selection might improve the predictive validity of the selection process. Finally, keeping the limitations in mind, our framework might be used in applied settings, such as training. For example, decision tendency feedback can be provided to clinicians who have diagnosed and treated patient vignettes described earlier. Pilots could be given feedback after watching videos of a plane landing in various weather conditions and being asked to provide judgments about safety and the subsequent decision to either continue or abort the landing. Many other examples mentioned throughout present similar opportunities. Each of these applications will require considerable investigation, but the potential utility they present is certainly worthy of future consideration.

## References

- Allwood, C. M., Granhag, P. A., & Jonsson, A.-C. (2006). Child witnesses' metamemory realism. *Scandinavian Journal of Psychology*, *47*(6), 461–470. doi:10.1111/j.1467-9450.2006.00530.x.
- Appelt, K. C., Milch, K. F., Handgraaf, M. J., & Weber, E. U. (2011). The decision making individual differences inventory and guidelines for the study of individual differences in judgment and decision-making research. *Judgment and Decision Making*, *6*(3), 252–262.
- Azevedo, R. (2009). Theoretical, conceptual, methodological, and instructional issues in research on metacognition and self-regulated learning: a discussion. *Metacognition Learning*, *4*, 87–95.
- Bavolar, J. (2013). Validation of the adult decision-making competence in Slovak students. *Judgment and Decision Making*, *8*(3), 386–392.
- Boekaerts, M., & Rozendaal, J. S. (2010). Using multiple calibration indices in order to capture the complex picture of what affects students' accuracy of feeling of confidence. *Learning and Instruction*, *20*(5), 372–382.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.
- Cattell, R. B. (1987). *Intelligence: Its structure, growth and action*. Amsterdam: North Holland.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioural sciences* (3rd ed.). Mahwah: Lawrence Erlbaum Associates.
- Costermans, J., Lories, G., & Ansay, C. (1992). Confidence level and feeling of knowing in question answering: the weight of inferential processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(1), 142–150. doi:10.1037/0278-7393.18.1.142.
- Cowley, E. (2004). Recognition confidence, recognition accuracy and choice. *Journal of Business Research*, *57*(6), 641–646. doi:10.1016/s0148-2963(02)00307.
- Crawford, J. D., & Stankov, L. (1996). Age differences in the realism of confidence judgements: a calibration study using tests of fluid and crystallized intelligence. *Learning and Individual Differences*, *8*(2), 83–103.
- DeMarree, K. G., & Petty, R. E. (2007). Self-certainty: parallels to attitude certainty. *International Journal of Psychology and Psychological Therapy*, *7*(2), 159–188.
- Edwards, W. (1954). The theory of decision making. *Psychological Bulletin*, *51*(4), 380–417.
- Edwards, W. (1961). Behavioral decision theory. *Annual Review of Psychology*, *12*, 473–498.
- Efkldes, A. (2008). Metacognition: defining its facets and levels of functioning in relation to self-regulation and co-regulation. *European Psychologist*, *13*(4), 277–287. doi:10.1027/1016-9040.13.4.277.
- Efkldes, A. (2009). The role of metacognitive experiences in the learning process. *Psicothema*, *21*(1), 76–82.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: a new area of cognitive-developmental inquiry. *American Psychologist*, *34*(10), 906–911.
- Gilovich, T., Griffin, D. W., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgement*. New York: Cambridge University Press.
- Glaser, M., & Weber, M. (2007). Overconfidence and trading volume. *The Geneva Risk and Insurance Review*, *32*(1), 1–36.
- Hacker, D. J., Bol, L., & Keener, M. C. (2008). Metacognition in education: A focus on calibration. *Handbook of metamemory and memory*, 429–455.
- Harvey, N. (1997). Confidence in judgment. *Trends in Cognitive Sciences*, *1*(2), 78–82.

- Harvey, N. (2001). Studying judgement: general issues. *Thinking & Reasoning*, 7(1), 103–118. doi:10.1080/13546780042000064.
- Horn, J. L., & Cattell, R. B. (1982). Whimsy and misunderstanding of gf-gc theory: a comment on Guilford. *Psychological Bulletin*, 91(3), 623–633.
- Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior*, 29(3), 340–362. doi:10.1016/0001-8791(86)90013-8.
- Juslin, P., Winman, A., & Olsson, H. (2000). Naive empiricism and dogmatism in confidence research: a critical examination of the hard-easy effect. *Psychological Review*, 107(2), 384–396.
- Kaplan, R. M., & Saccuzzo, D. P. (2005). *Psychological testing: principles, applications, and issues* (6th ed.). Belmont: Thomson Wadsworth.
- Kelemen, W. L., Frost, P. J., & Weaver, C. A. (2000). Individual differences in metacognition: Evidence against a general metacognitive ability. *Memory & Cognition*, 28(1), 92–107.
- Kleitman, S. (2008). *Metacognition in the rationality debate*. Saarbrücken: VDM Verlag Dr. Müller Aktiengesellschaft & Co. KG.
- Kleitman, S., & Gibson, J. (2011). Metacognitive beliefs, self-confidence and primary learning environment of sixth grade students. *Learning and Individual Differences*, 21, 728–735.
- Kleitman, S., & Mascrop, T. (2010). Self-confidence and academic achievements in primary-school children: Their relationships and links to parental bonds, intelligence, age, and gender. In A. Efklides & P. Misailidi (Eds.), *Trends and prospects in metacognition research* (pp. 293–326). US: Springer.
- Kleitman, S., & Stankov, L. (2001). Ecological and person-oriented aspects of metacognitive processes in test-taking. *Applied Cognitive Psychology*, 15(3), 321–341. doi:10.1002/acp.705.
- Kleitman, S., & Stankov, L. (2007). Self-confidence and metacognitive processes. *Learning and Individual Differences*, 17, 161–173.
- Kleitman, S., Stankov, L., Allwood, C. M., Young, S., & Mak, K. (2011). *Metacognitive self-confidence in school-aged children educators*. Unpublished Book Chapter.
- Knight, F. H. (1921). *Risk, uncertainty and profit*. Boston: Houghton Mifflin.
- Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, 103(3), 490.
- Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance*, 20(2), 159–183. doi:10.1016/0030-5073(77)90001-0.
- McKenzie, C. R. (1998). Taking into account the strength of an alternative hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(3), 771–792.
- Mellers, B., Schwartz, A., & Cooke, A. (1998). Judgment and decision making. *Annual Review of Psychology*, 49, 447–477.
- Mengelkamp, C., & Bannert, M. (2010). Accuracy of confidence judgments: stability and generality in the learning process and predictive validity for learning outcome. *Memory & Cognition*, 38(4), 441–451.
- Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, 115(2), 502–517.
- Nelson, T. O. (1996). Consciousness and metacognition. *American Psychologist*, 51(2), 102–116.
- Pallier, G., Wilkinson, R., Danthir, V., Kleitman, S., Knezevic, G., Stankov, L., & Roberts, R. D. (2002). The role of individual differences in the accuracy of confidence judgments. *Journal of General Psychology*, 129(3), 257–299. doi:10.1080/00221300209602099.
- Parker, A. M., Bruine de Bruin, W., Yoong, J., & Willis, R. (2012). Inappropriate confidence and retirement planning: four studies with a national sample. *Journal of Behavioral Decision Making*, 25(4), 382–389.
- Raven, J. C. (1938-65). *Progressive matrices*. New York: The Psychological Corporation.
- Roets, A., & Van Hiel, A. (2011). Item selection and validation of a brief, 15-item version of the need for closure scale. *Personality and Individual Differences*, 50(1), 90–94. doi:10.1016/j.paid.2010.09.004.
- Saucier, G. (2008). Measures of personality factors found recurrently in human lexicons. In G. J. Boyle, G. Matthews, & D. Sakloske (Eds.), *Handbook of personality theory and testing* (Vol. 2). London: Sage.
- Schraw, G. (2009). *Measuring metacognitive judgments. Handbook of metacognition in education* (pp. 415–429). New York: Routledge/Taylor & Francis Group.
- Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology*, 19(4), 460–475. doi:10.1006/ceps.1994.1033.
- Schraw, G., Dunkle, M. E., Bendixen, L. D., & Roedel, T. D. (1995). Does a general monitoring skill exist? *Journal of Educational Psychology*, 87(3), 433–444. doi:10.1037/0022-0663.87.3.433.
- Shaughnessy, J. J. (1979). Confidence-judgment accuracy as a predictor of test performance. *Journal of Research in Personality*, 13(4), 505–514. doi:10.1016/0092-6566(79)90012-6.
- Slovic, P., Fischhoff, B., & Lichtenstein, S. (1977). Behavioral decision theory. *Annual Review of Psychology*, 28, 1–39.
- Stankov, L. (1997). *The Gf/Gc quickie test battery*. School of Psychology: The University of Sydney. Unpublished test battery.



- Stankov, L. (1999). Mining on the “no man’s land” between intelligence and personality. In P. L. Ackerman, P. C. Kyllonen, & R. D. Roberts (Eds.), *Learning and individual differences: Process, trait, and content determinants* (pp. 315–337). Washington, DC: American Psychological Association.
- Stankov, L. (2000). Complexity, metacognition, and fluid intelligence. *Intelligence*, *28*, 121–143.
- Stankov, L., & Crawford, J. D. (1996). Confidence judgments in studies of individual differences. *Personality and Individual Differences*, *21*(6), 971–986.
- Stankov, L., & Crawford, J. D. (1997). Self-confidence and performance on tests of cognitive abilities. *Intelligence*, *25*(2), 93–109.
- Stankov, L., & Kleitman, S. (2008). Processes on the borderline between cognitive abilities and personality: Confidence and its realism. In G. J. Boyle, G. Matthews & D. Saklofske (Eds.), *The SAGE handbook of personality theory and assessment* (pp. 541–555). London, England: SAGE Publications.
- Stankov, L., & Lee, J. (2008). Confidence and cognitive test performance. *Journal of Educational Psychology*, *100*(4), 961–976. doi:10.1037/a0012546.
- Stankov, L., Lee, J., Luo, W., & Hogan, D. J. (2012). Confidence: a better predictor of academic achievement than self-efficacy, self-concept and anxiety? *Learning and Individual Differences*, *22*(6), 747–758.
- Stankov, L., Kleitman, S., & Jackson, S. A. (2013). Measures of the trait of confidence: In G. Boyle, & D. Saklofske (Eds.), *Measures of personality and social psychological attitudes*. Amsterdam, Netherlands: Elsevier (in press).
- Veenman, M. V., Wilhelm, P., & Beishuizen, J. J. (2004). The relation between intellectual and metacognitive skills from a developmental perspective. *Learning and Instruction*, *14*(1), 89–109.
- Want, J., & Kleitman, S. (2006). Imposter phenomenon and self-handicapping: links with parenting styles and self-confidence. *Personality and Individual Differences*, *40*(5), 961–971.
- West, R. F., & Stanovich, K. E. (1997). The domain specificity and generality of overconfidence: individual differences in performance estimation bias. *Psychonomic Bulletin & Review*, *4*(3), 387–392.
- Williams, E. F., & Gilovich, T. (2008). Do people really believe they are above average? *Journal of Experimental Social Psychology*, *44*(4), 1121–1128.
- Yang, H., & Thompson, C. (2010). Nurses’ risk assessment judgements: a confidence calibration study. *Journal of Advanced Nursing*, *66*(12), 2751–2760. doi:10.1111/j.1365-2648.2010.05437.x.
- Yates, J. F. (1990). *Judgment and decision making*. Englewood Cliffs: Prentice-Hall.