Correspondence

Systematic Overestimation of Reflection Impulsivity in the Information Sampling Task

To the Editor:

Impulsivity, a psychological construct comprising both motor and cognitive factors, is thought to underlie important inter-individual differences in health and disease (1). In particular, reflection impulsivity, which refers to the tendency to gather and evaluate information before decision making (2), has been implicated in many psychiatric and neurological disorders (3–5). One of the standard tasks for measuring reflection impulsivity in healthy and clinical populations is the Information Sampling Task (IST), designed by Clark *et al.* (3) and included in the widely used Cambridge Neuropsychological Test Automated Battery (CANTAB) (6). In this CANTAB version of the IST, participants sample a variable amount of information about an uncertain outcome before making a decision. The amount of information sampled before the decision gives a measure of participants' reflection impulsivity. In this correspondence, we show that the calculation of the IST's main outcome measure, *P(correct)*, is based on incorrect statistical inference, resulting in systematic overestimation of participants' reflection impulsivity and potentially inflated type II error rates. This might affect the results of numerous recent psychopharmacological, neuropsychological, and psychiatric publications that have used the IST (4,5,7).

In the IST, participants are presented with a 5 × 5 grid of closed boxes, each of which is of one of two colors. Participants indicate on each trial which of the two colors they believe to be in the majority, and can sample information before making a decision by opening boxes one at a time to reveal their colors. The task's two outcome measures are 1) the number of boxes opened before the decision is made and 2) *P(correct)*, or the probability that the participant's chosen color was in the majority, based on the boxes opened at the time of response. Many studies (3–5,7) tend to focus on *P(correct)* as the outcome measure because it is thought to provide the more accurate estimate of reflection impulsivity, given that a given number of open boxes can be associated with very different levels of decision uncertainty, depending on the colors of the open boxes.

Let n_1 be the number of open boxes of the color the participant chose and n_2 the number of open boxes of the unchosen color. The IST calculates *P(correct)* as follows:

$$P(correct) = \frac{\sum_{k=A}^{z} \binom{z}{k}}{2^z} \tag{1}$$

where $Z = 25 - (n_1 + n_2)$, the number of boxes still to be opened, and $A = 13 - n_1$, the number of additional boxes of the chosen color required for a majority. Intuitively, Equation 1 enumerates combinations of boxes that would provide a majority for the chosen color (numerator) as a fraction of possible configurations of colors in the unopened boxes, assuming that each unopened box has an equal probability of being either of the two colors (denominator). Although prima facie reasonable, this assumption of equal probability mistakenly assumes that unopened box colors are conditionally independent from the boxes already opened. In fact, however, the colors of the opened boxes provide important information to the participant regarding the underlying proportions of the two colors. For instance, if nine boxes have been opened and all nine are blue, it is extremely unlikely that the sixteen unopened boxes are half blue and half yellow. Because of this conditional dependence, the colors of open boxes provide information about the majority color that is not incorporated into Equation 1. As a result, Equation 1 underestimates *P(correct)*, and therefore overestimates participants' reflection impulsivity.

An accurate calculation of *P(correct)* must take into account the colors of the open boxes. Formally, this can be expressed as a Bayesian inference problem over θ, the true number of boxes of the chosen color in the entire grid. Following Bayes' rule, the probability mass function (PMF) for θ given the colors of open boxes is given by Equation 2:

$$\Pr(\theta = M|n_1, n_2) = \frac{\Pr(\theta = M)\Pr(n_1, n_2|\theta = M)}{\Pr(n_1, n_2)} \tag{2}$$

where $\Pr(\theta = M)$ is the prior probability that θ takes a given value, $\Pr(n_1, n_2|\theta = M)$ is the likelihood of the opened boxes if the true number of boxes of the chosen color were M, and $\Pr(n_1, n_2)$ is a normalization constant. Under a naïve prior (i.e., because participants are given no advance knowledge regarding the relative proportions of colors), this PMF is given by Equation 3 (8):

$$\Pr(\theta = M|n_1, n_2) = \frac{\binom{M}{n_1}\binom{25-M}{n_2}}{\sum_{j=n_1}^{25-n_2}\binom{j}{n_1}\binom{25-j}{n_2}} \tag{3}$$

A more accurate *P(correct)* measure can therefore be calculated as follows (MATLAB code available at https://github.com/danielbrianbennett/ist/):

$$\begin{aligned} P(correct) &= \Pr(\theta \geq 13|n_1, n_2) \\ &= \sum_{M=13}^{25} \Pr(\theta = M|n_1, n_2) \end{aligned} \tag{4}$$

Although the error in Equation 1 is problematic in itself, a greater concern is that the size of the error is likely to vary systematically across participants. The discrepancy between the two *P(correct)* measures is largest for small numbers of boxes opened, and approaches zero with greater numbers of boxes opened. Consequently, the original *P(correct)* measure overestimates reflection impulsivity more for participants high in reflection impulsivity than for those low in reflection impulsivity. Results of previous studies comparing *P(correct)* between impulsive patient groups and healthy controls are therefore unreliable. This error will tend in most cases to underestimate the size of differences between groups, and therefore the error in *P(correct)* is likely to result in an inflated type II error rate (i.e., failure to reject a false null hypothesis).

http://dx.doi.org/10.1016/j.biopsych.2016.05.027
ISSN: 0006-3223

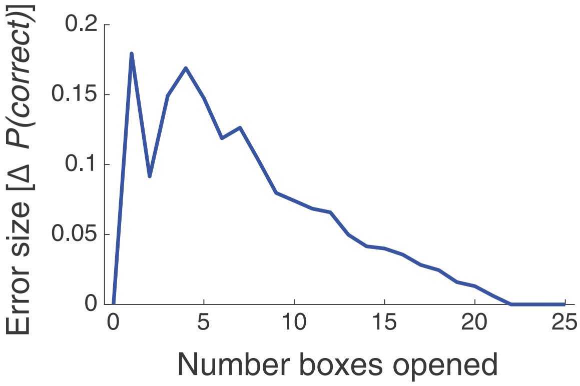

Figure 1. Mean error size (vertical axis) as a function of number of boxes opened (horizontal axis). Error size was calculated as the difference between the revised *P(correct)* measure and the original *P(correct)* measure. Larger error sizes indicate greater underestimation of *P(correct)* by the original measure, and therefore greater overestimation of reflection impulsivity.

Empirical findings support this conclusion. Reanalysis of data from a separate sample of 109 participants (9) revealed that the size of the *P(correct)* error varied as a function of number of boxes opened (Figure 1). The average amount of underestimation of *P(correct)* using the IST formula was 7.7% (SD = 3.56%). Significant group-mean differences of approximately this size have been previously reported (3,5), and therefore this error size is clearly of practical significance.

In summary, the computation of the major outcome measure of the IST is based on a flawed statistical assumption that results in systematic overestimation of reflection impulsivity. The size of this error differs for different levels of reflection impulsivity, and therefore the overestimation may compromise statistical inferences made using this measure. In particular, past research reporting group differences in reflection impulsivity (4,5) may have significantly underestimated effect sizes, whereas studies reporting null findings (7) might have failed to detect true group differences in reflection impulsivity.

Daniel Bennett
Stuart Oldham
Andrew Dawson
Linden Parkes
Carsten Murawski
Murat Yücel

Acknowledgments and Disclosures

This work was supported by a Strategic Research Initiative Grant (to CM) from the University of Melbourne, Australia, and the National Health and Medical Research Council of Australia (Grant No. APP1021973 to MY).

We thank Marni Kras for providing a dataset for reanalysis.

The authors report no biomedical financial interests or potential conflicts of interest.

Article Information

From the School of Psychological Sciences (DB) and the Department of Finance (DB, CM), The University of Melbourne, Melbourne; and the School of Psychological Sciences and the Monash Institute of Cognitive and Clinical Neurosciences (SO, AD, LP, MY), Monash University, Victoria, Australia.

CM and MY contributed equally to this work.

Address correspondence to Carsten Murawski, Ph.D., Decision Neuroscience Laboratory, The University of Melbourne, Victoria 3010, Australia; E-mail: carstenm@unimelb.edu.au.

See also associated correspondence: http://dx.doi.org/10.1016/j.biopsych.2016.06.018.

Received May 10, 2016; accepted May 10, 2016.

References

1. Evenden J (1999): Impulsivity: A discussion of clinical and experimental findings. J Psychopharmacol 13:180–192.
2. Kagan J (1966): Reflection-impulsivity: The generality and dynamics of conceptual tempo. J Abnormal Psychol 71:17–24.
3. Clark L, Robbins TW, Ersche KD, Sahakian BJ (2006): Reflection impulsivity in current and former substance users. Biol Psychiatry 60: 515–522.
4. Tavares JVT, Clark L, Cannon DM, Erickson K, Drevets WC, Sahakian BJ (2007): Distinct profiles of neurocognitive function in unmedicated unipolar depression and bipolar II depression. Biol Psychiatry 62:917–924.
5. Zamarian L, Benke T, Brand M, Djamshidian A, Delazer M (2015): Impaired information sampling in mild dementia of Alzheimer's type but not in healthy aging. Neuropsychology 29:353–367.
6. Fray PJ, Robbins TW, Sahakian BJ (1996): Neuorpsychiatyric applications of CANTAB. Int J Geriatr Psychiatry 11:329–336.
7. DeVito EE, Blackwell AD, Clark L, Kent L, Dezsery AM, Turner DC, *et al.* (2009): Methylphenidate improves response inhibition but not reflection–impulsivity in children with attention deficit hyperactivity disorder (ADHD). Psychopharmacology 202:531–539.
8. Link WA, Barker RJ (2009): Bayesian Inference: With Ecological Applications. Boston, Academic Press.
9. Kras M (2015): Impulsive traits and impulsive behavior associated with opioid dependence: Relationship to abstinence and heroin use. Doctoral Thesis, Monash University: Victoria, Austalia.