Intelligence 42 (2014) 1–4

Contents lists available at ScienceDirect

Intelligence

Smarter but slower? A comment on Woodley, te Nijenhuis & Murphy (2013)

Ted Nettelbeck

School of Psychology, University of Adelaide, South Australia 5005 Australia

ARTICLE INFO

Article history:
Received 17 June 2013
Received in revised form 24 September 2013
Accepted 24 September 2013
Available online 7 November 2013

Keywords:
Simple reaction times
General intelligence
Dysgenic fertility
Flynn effect
Inspection time

ABSTRACT

Woodley, te Nijenhuis, and Murphy (2013) have concluded that markedly slower mean simple reaction times (RT) across the past century are consistent with dysgenic fertility being responsible for a mean loss of 13 IQ points in the general population. They have recognised that the capacity to engage in abstract problem solving, as tapped by tests like the Wechsler scales and Raven's Progressive Matrices (RPM), has improved substantially throughout the same time but have concluded that this trend has masked the dysgenic effect. I suggest that there are reasonable grounds to challenge these conclusions. For them to be sound requires, first, accepting that reliable, absolute measures of simple RT are not influenced by different apparatus and procedures. This is inconsistent with current knowledge. Second, the observed slowing in mean simple RT would need to be entirely attributable to genetically caused decline in general intelligence. This has not been established. Furthermore, although it is possible in principle that different cognitive abilities could simultaneously diverge because of counter acting influences, decoupling such trends in performance on a single measure of general intelligence is not possible.

1. Smarter but slower? A comment on Woodley, te Nijenhuis, and Murphy (2013)

Woodley et al. (2013) have argued that (i) mean simple reaction time (RT) has progressively and steadily slowed across about 100 years by a massive 68%; (ii) individual differences in general intelligence (*g*) are substantially determined by individual differences in simple RT because simple RT taps a genetically determined capacity that is fundamental to *g* and slower simple RT is therefore consistent with a dysgenic fertility effect for general intelligence equivalent to about 13 IQ points; and (iii) diminishing average general intelligence in the population has not been observed because the trend has been masked by the Flynn effect, whereby average IQ has steadily risen over the same period, presumably as a consequence of favourable environmental influences. Flynn's (2007) position is that improved strategies for thinking have driven up IQ scores, while at the same time unspecified fundamental capacities have remained unchanged for very long periods of time that have straddled the past 100 years in question. Both Flynn's and Woodley et al.'s position therefore allow the hypothetical possibility of distinguishing between a fundamental intelligent capacity and aspects of intelligence subject to environmental differences. However, contrary to Flynn, Woodley et al. have argued that fundamental capacities could have declined, unrelated to environmental circumstances.

One problem for accepting Woodley et al.'s version is that, although speed of reaction is partially heritable, they provide no evidence that the observed change in mean simple RT has been entirely the consequence of genetic influences. Moreover, the validity for simple RT as a genetically determined marker for *g* rests on correlation with the same tests that have provided the case for rising IQ. However, as things stand currently, there is no way of directly establishing that variance shared between simple RT and *g* extracted from IQ tests is fundamentally different from variance between *g* and whatever drives rising IQ. In what follows I expand further on these challenges to Woodley et al.'s conclusion, proposing that

E-mail address: ted.nettelbeck@adelaide.edu.au.

0160-2896/$ – see front matter
http://dx.doi.org/10.1016/j.intell.2013.09.006

it cannot be established that the apparent trend to slower simple RT is not attributable to nongenetic influences. I further suggest that, plausibly, the apparent change reflects differences in measurement practices, something that is directly testable.

2. Differences in simple RT substantially account for differences in general intelligence

Speed of reaction has long been a candidate for a non-psychometric measure to tap the essential core of intelligence (see Vernon, 1987 for wide ranging discussion) although, ultimately, the only means for verifying this has been to rely on IQ-type tests as the means for validating a particular speed measure as an ability measure. To summarise 100 years of RT research, there is reliable correlation between measures of RT and measures of IQ but the order of such correlations is far short of what would be required to use the former to explain the latter. Jensen's (2006) suggestion that RT may index some fundamental characteristic of the brain that reflects the efficiency of complex cognitive systems of information processing is plausible, particularly in relation to cognitive changes associated with normal ageing (Salthouse, 1996). This possibility is at least consistent with the general observation from RT studies of higher correlations with g, particularly with individual variability in RT, if the RT task makes more complex demands (Jensen, 1998, p.237). Individual differences on RT tasks and other mental chronometric tasks are partially heritable, with more complex tasks yielding higher estimates of heritability (Beaujean, 2005). In short, a more complex measure of speed of responding might account for something like 20–25% of variance on IQ measures; but this leaves a lot unaccounted for — and the outcome for simple RT has generally been found to be a lot less promising. Although Woodley et al. (2013) have used results from Deary, Der, and Ford (2001) "for obtaining benchmark estimates of the simple RT/IQ correlation" (Woodley et al., p. 3), this correlation of −.31 was obtained from a sample of 56 year-olds and is considerably higher than estimates that typically have been found with younger adults. For example, the overall unweighted mean correlation between the intercept of the Hick function (simple RT) and sundry intelligence measures from Jensen's (1987) considerable data base ($N = 774$), derived from the decision component of his RT apparatus, was −.12. Applying corrections for range and attenuation did not increase the estimate beyond −.19. Jensen's conclusion was that the association between simple RT and intelligence was weak. Nonetheless, if we allow that simple RT has slowed, even if the association with g is substantially weaker than claimed, this still permits a claim for a dysgenic fertility effect, if it can be established that the change in simple reaction time is entirely the consequence of genetic factors. However, this is not so; and a serious challenge to claims by Woodley et al. is that the apparent slowing observed in simple RT reflects confounds that are unrelated to intelligence.

3. Simple RT has slowed appreciably

The evidence provided for the claim about a dysgenic fertility effect is not convincing because the simple RT data have been derived by widely different methods. Woodley et al. (2013) considered whether methodological differences could confound outcomes but concluded that this was unlikely. I do not agree; and raise three different objections that challenge their broad conclusion. First, the statistically significant meta-regression summarized in their Fig. 1 is heavily reliant on the two 19th century studies, which are markedly distant in the overall time line from those that follow. Silverman (2010), whose article has provided most of the simple RT data used by Woodley et al., clearly recognised this: "Accurately describing change over time requires that both ends of the temporal dimension be well represented" (p. 45). The assumption of a general linear trend to slower simple RT throughout this long period of time is therefore questionable. The result probably represents in part markedly different earlier methods of measurement from those employed more recently, as I discuss below.

Second, although the level of detail provided by the various papers cited by Woodley et al. (2013) has often been insufficient to permit firm conclusions, it is obvious that there is no common method among them and, instead, a range of different technologies has been used to measure simple RT. This alone might explain different mean outcomes. Woodley et al. have referred to "methodological artefacts and sample peculiarities" (p. 7) but this does not capture the extent of differences among the studies listed. These technologies have included Galton's mechanical procedure for translating the rate of a swinging pendulum into a response latency and different kinds of lights as target stimuli (light reflected from a mirror, electric filaments, liquid crystal displays, stimuli generated on a computer screen). These alternatives have involved apparatus layout in different configurations with timing controlled by different chronometric methods capable of accuracy ranging from only hundredths of a second for earlier equipment to millisecond accuracy more recently. Different kinds of apparatus have been driven by different electric circuitry and computer programs. Moreover, there have been procedural differences in the extent of prior practice afforded participants and in the numbers of trials from which means have been derived, which can influence the reliability of measurement. Notably, few of these studies have provided reliability data although Johnson et al. (1985) estimated the test–retest reliability of Galton's method at about .2. This result, for the earliest and quickest estimate, compares unfavourably with estimates of between .7 and .8 that Jensen commonly registered during the 1980s and beyond (Jensen, 2006). I submit that measurements with very low reliability should not be included if one is interested in obtaining absolute estimates of simple RT. The main point to note from this is that it is not possible to aggregate data generated by different methods. Jensen early recognised that nonstandardised methods across different laboratories would preclude comparability of data sets when using RT to explore its relevance to a better understanding of intelligence (see, for example, 2006. p. 76). From the outset of his RT program in the 1970s he maintained uniform method and by doing so he accumulated a valuable data set that permitted comparisons across time. Such considerations apply no less importantly to other Elementary Cognitive Tasks. My experience with inspection time measurement has demonstrated marked effects on absolute values that procedural changes can cause to results from a single laboratory. That correlation between IT and various IQ measures has been demonstrated by so many methodological variations to the basic idea has tended to confirm the reliability of the relationship but, that aside, I have long recognised that standardisation of method would be essential if the measure was to find wide application that permitted normative comparisons (Nettelbeck, 1987).

The third problem relates to the instructions given to respondents or, more accurately, how different samples may have interpreted instructions. Few details about this are available from the studies under consideration but any RT measure depends on how closely a respondent conforms to an assumption that every response be as fast as possible without making errors. It is long established that different forms of instruction can bias respondents to adopt responding strategies closer to or further from their accuracy thresholds (Pachella, 1974). It is not my intention to imply that instructions to participants could have varied systematically over time but, rather, that earlier circumstances may have resulted in faster estimates if implausibly quick reactions were not excluded from analysis. Achieving a reaction that is as fast as possible requires that every response be at the individual's threshold for maintaining accuracy. This assumption can never be met literally but it has been widely accepted that a sufficient approximation to the ideal is achieved if the respondent registers a small number of occasional errors. However, whereas errors for choice reactions can be reliably identified as incorrect stimulus–response pairings, an error for simple RT can only be observed by the occurrence of an anticipated response, registered before the signal occurs.

Modern practice has often accepted that latencies faster than any possible human decision capacity will occur because of anticipation but that these should be excluded (Jensen, 2006). Jensen routinely omitted latencies of 150 ms or less, a practice that inevitably increased mean and median latencies. However, because historically there has been little detailed attention in the literature to these matters, I have not been able to determine if something approaching a standard practice has ever been adopted, whether there has been an increasing trend to such practice in more recent times, or to what extent any such practices were followed by the authors included by Woodley et al. (2013). However, Johnson et al.'s (1895) account of Galton's method suggests that this did not exclude implausibly fast responses if these occurred. Particularly when only few trials are used to estimate individual measures of simple RT, I predict that higher proportions of anticipatory responses will be generated, assuming that participants are genuinely attempting to "react as quickly as possible". To my knowledge, this has not been tested but it certainly could be. I do not know whether, compared to earlier studies, those more recent have tended to discard very fast responses. I also acknowledge that any such tendency is unlikely to fully account for the slower more recent measurements; but it could be another contributory factor.

The only sure way of reliably testing whether simple RT has slowed over time is to replicate the earlier studies using exactly the same methods and procedures followed formerly. Nettelbeck and Wilson (2004) did precisely this so as to compare the inspection times (ITs) of school children aged 6–13 years with ITs of children measured 20 years previously, at the same ages and from the same school, which was still located in the same socio-economic area. The IT paradigm measures processing speed from accuracy of responding under time constraints, without relying on speed of movement (Nettelbeck, 2001). IT has consistently been found to correlate with IQ to an extent as strong as or stronger than reports for both simple and choice RT (Grudnik & Kranzler, 2001). Nettelbeck and Wilson replicated in every detail the apparatus, computer technology, instructions and procedure that they had used on the earlier occasion. They found no change in processing speed across a span of two decades, even though word knowledge had improved, consistent with a Flynn effect. To the best of my knowledge this study remains the only example of a direct, reliable test as to whether processing speed is constant across generations even though performance on an IQ-type measure improves; the results suggested that it is. Furthermore, Woodley et al.'s conclusions from Silverman's (2010) data notwithstanding, Silverman carefully emphasised that he was not reporting "change according to a specific function over time" (p. 45) but, instead, slower RTs in the modern era compared to Galton's time. His opinion was that slower RT was unlikely to be the consequence of methodological differences; but he did recognise that confirmation would require the application of standard methods "in order to enable more refined analyses" (p. 39). I disagree with his first assertion but applaud his second. Until those advocating a decline in RT have demonstrated this directly by the means suggested here, I submit that it is reasonable to doubt that any true change in average reaction capacity has occurred during the past 100 years.

4. The Flynn effect has masked a dysgenic fertility effect for general intelligence

Nunnally and Bernstein (1994) advised against correcting correlations for measurement error where there could be doubt about the reliability of underlying assumptions. I submit that such is the case here. Deary et al.'s (2001) correlation between simple RT and IQ of −.31 is among the highest reported, something that these authors acknowledged. Before accepting this value it would be preferable to see it replicated. However, by correcting this coefficient for unreliability, range restriction in IQ and imperfect measure of *g*, Woodley et al. (2013) have estimated the true correlation to be −.54, a considerable increase. It is reasonable to doubt an estimate for IQ loss over 100 years derived solely from Deary et al.'s correlation.

Furthermore, to accept that a dysgenic effect has occurred requires accepting that we can decouple two different hypothetical capacities that would underpin performance on the kinds of tests developed to estimate intelligence. It is not clear to me how this would be possible. The Flynn effect (Flynn, 2007) has been demonstrated for different kinds of ability tests like some of those included in Wechsler scales, but also for Raven's Progressive Matrices (RPM), a test widely accepted as a measure of *g* (Jensen, 1998). The argument for a dysgenic fertility effect is therefore necessarily derived from the same tests used to demonstrate rising IQ. Flynn has not argued that intelligence has increased. Instead, his position is that successive generations have become "smarter" because they have learned new ways of thinking that have increasingly relied on more abstract strategies and that tests like RPM are particularly sensitive to this trend. However, accepting the dysgenic hypothesis requires that those tests most sensitive to more abstract ways of problem solving can simultaneously register both an increase in those capacities and a decrease in something else that is the essence of the "education of relations and correlates". Although in theory diverging trends are possible, the RPM can only yield a net outcome, namely the number of items correct; and on current evidence this is improving.

5. Conclusion

Throughout the history of research into the nature of human intelligence there has been occasional expression of concern in some quarters about dysgenic fertility decline influencing intelligence; i.e. a trend to lower average IQ because the genetic potential for IQ within the population is becoming less favourable (Cattell, 1940, cited in Hunt, 2011, p. 292; Herrnstein & Murray, 1994; Lynn, 1996). The basic argument appears plausible; intelligence is substantially inherited and therefore, if smarter people have fewer children than less smart people, intelligence will decline. As Woodley et al. (2013) and others have pointed out, the trend in Western European–North American cultures has for some time been that those presumed to be most able have smaller numbers of children. However, a critical test of this argument has proven to be excessively difficult, principally because of continuing debate about limits to the validity of IQ tests, uncertainty about the extent to which IQ is heritable, the difficulties of unscrambling heritable and environmental influences on familial variables and little knowledge to date about what genes are responsible for intelligence. Despite speculation by advocates for a dysgenic fertility effect, many remain sceptical (see, for example, Hunt's (2011) conclusion that he does not find the dysgenic argument "completely compelling" (p. 269; pp. 292–4)).

It is not my intention to challenge the validity or heritability of IQ. I remain a strong advocate of testing as an extremely useful tool for assessing a general intelligence capacity as well as a wide range of relatively independent but related abilities, I regard the availability of these instruments as a major achievement of 20th century psychology and I acknowledge the valuable contribution of behavioural genetics in establishing that genetic inheritance sets real limits on the development of intelligence. There are, however, strong reasons as outlined above for doubting the conclusions that Woodley et al. have advanced.

Acknowledgments

I am grateful to Carlene Wilson and Nick Burns for comments on drafts of this paper and to an anonymous reviewer for encouraging me to articulate my suggestions more clearly.

References

Beaujean, A. A. (2005). Heritability of cognitive abilities as measured by mental chronometric tasks: A meta-analysis. *Intelligence*, *33*, 187–201.

Deary, I. J., Der, G., & Ford, G. (2001). Reaction times and intelligence differences: A population-based cohort study. *Intelligence*, *29*, 389–399.

Flynn, J. R. (2007). *What is intelligence?* Cambridge: Cambridge University Press.

Grudnik, J. L., & Kranzler, J. H. (2001). Meta-analysis of the relationship between intelligence and inspection time. *Intelligence*, *29*, 523–535.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life.* New York: The Free Press.

Hunt, E. (2011). *Human intelligence.* Cambridge: Cambridge University Press.

Jensen, A. R. (1987). Individual differences in the Hick paradigm. In P. A. Vernon (Ed.), *Speed of information-processing and intelligence* (pp. 101–175). Norwood NJ: Ablex.

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport: Praeger.

Jensen, A. R. (2006). *Clocking the mind: Mental chronometry and individual differences.* Oxford: Elsevier.

Johnson, R. C., McClearn, G. E., Yuen, S., Nagoshi, C. T., Ahern, F. M., & Cole, R. E. (1985). Galton's data a century later. *American Psychologist*, *40*, 875–892.

Lynn, R. (1996). *Dysgenics: Genetic deterioration in modern populations.* Westport: Praeger.

Nettelbeck, T. (1987). Inspection time and intelligence. In P. A. Vernon (Ed.), *Speed of information-processing and intelligence* (pp. 295–346). Norwood NJ: Ablex.

Nettelbeck, T. (2001). Correlation between inspection time and psychometric abilities: A personal perspective. *Intelligence*, *29*, 459–474.

Nettelbeck, T., & Wilson, C. (2004). The Flynn effect: Smarter not faster. *Intelligence*, *32*, 85–93, http://dx.doi.org/10.1016/S0160-2896(03)00060-6.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.)New York: McGraw-Hill.

Pachella, R. G. (1974). The interpretation of reaction time in information processing research. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognitive*. Hillsdale NJ: Erlbaum.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, *103*, 403–428.

Silverman, I. W. (2010). Simple reaction time: It is not what it used to be. *American Journal of Psychology*, *123*, 39–50 (http://www.jstor.org/stable/10.5406/amerjpsyc.123.1.0039)

Vernon, P. A. (Ed.). (1987). *Speed of information-processing and intelligence* (pp. 101–175). Norwood NJ: Ablex.

Woodley, M. A., te Nijenhuis, J., & Murphy, R. (2013). Were the Victorians cleverer than us? The decline of general intelligence estimated from a meta-analysis of the slowing of simple reaction time. *Intelligence*, http://dx.doi.org/10.1016/j.intell.2013.04.006.