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 © 2013 Elsevier Inc. All rights reserved.
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