Commentary

Jensen's Reaction-Time Studies: A Reply to Longstreth

ARTHUR R. JENSEN

University of California, Berkeley

PHILIP A. VERNON

University of Western Ontario

Longstreth's (1984) critique of Jensen's research on the relationship of IQ to individual differences in visual reaction time (RT), measured in the Hick paradigm, is found to have numerous errors of fact and interpretation, some trivial and some of theoretical importance. Longstreth's narrowly focused and conjectural style of criticism, which peculiarly strains to favor the null hypothesis, unfortunately obscures the essential findings of Jensen's (and others') studies of the RT–IQ relationship. The two main negative verdicts of Longstreth's critique concerning the RT–IQ relationship are refuted by meta-analyses of presently available data. First, not only do individual differences in RT show a significant negative correlation with IQ, but individual differences in the *slope* of the regression of RT on stimulus set size scaled in bits (i.e., the binary logarithm of the number of potential reaction stimuli) also show a fully significant, albeit low, negative correlation with IQ. Contrary to Longstreth's second negative surmise, meta-analysis also shows that the magnitude of the RT–IQ correlation itself is a linearly increasing (negative) function of stimulus set-size scaled in bits.

A critique of research is peculiarly privileged, often riding with scarcely questioned license, because readers ordinarily suppose that pains are taken to ensure that a critique is less liable to faultiness than the target of its criticism. A clear object lesson proving that this common supposition is not always dependable is provided by Longstreth's (1984) critique of Jensen's reaction-time (RT) investigations of intelligence. Our examination of Longstreth's critique finds that its few valid critical points, usually on minor issues, are admixed with a number of factual and interpretive errors, some trivial and some of theoretical importance. Also, his surmise and conjecture on certain major points are actually contradicted by a preponderance of the present evidence. Not all of the relevant evidence was available at the time Longstreth's critique was written. He was correct in pointing out that certain information pertinent to some of the questions he has raised was not reported in earlier publications. Now that the information is at hand, however, it indicates that some of Longstreth's criticisms are either unfounded or do not apply to Jensen's findings. If one is not to miss seeing the forest for the

---

trees, in viewing a program of research on a given subject, a narrowly focused style of criticism aimed at particular findings in single studies must be balanced by meta-analyses of the quantitative findings assembled from all of the directly relevant and strictly comparable studies that exist. Otherwise, criticism can be biased to favor the null hypothesis prematurely, falling into Type II error.

Before taking up Longstreth's specific criticisms, some rather general re-marks are in order concerning the research under discussion. Although Jensen's studies of the chronometric correlates of *g* have used a number of different procedures for measuring individual differences in speed of information process-ing in various elementary cognitive tasks, Longstreth's critique focuses ex-clusively on only one of these, probably the simplest, which we have termed the Hick paradigm, named after Hick's law (Hick, 1952). Hick's law states that RT increases as a linear function of the logarithm of the number (*n*) of alternatives (or *stimulus set size*) in the array of reaction stimuli. Set-size (*n*) is usually scaled in *bits*. (A *bit* is the binary logarithm of *n*.) Longstreth has pictured and described Jensen's apparatus for measuring a subject's RT and movement time (MT) in response to varying values of *n*, which in most studies are 1, 2, 4, and 8, corresponding to 0, 1, 2, and 3 bits of *information* in the technical sense of that term. Longstreth has renamed RT as *initiation time*, or IT. But this new label seems pointless, and it could also be confusing, because in the literature on mental chronometry IT already stands for *inspection time* (i.e., the shortest interval between a test stimulus and a masking stimulus at which a subject can identify the test stimulus at some specified level of accuracy). (RT has also been termed *decision time*, or DT.) Because Jensen has explicitly defined RT and MT in his articles, we will be consistent here and retain his terminology. Readers simply must remain aware that what Jensen refers to as RT in the context of the Hick paradigm is labeled IT by Longstreth. (Jensen defines RT as the interval, measured in milliseconds, between the onset of the reaction stimulus and the subject's removing his index finger from the "home" button; MT is the interval between the subject's releasing the home button and pressing another button adjacent to the reaction stimulus [a light], which terminates it.)

Jensen's research on the Hick paradigm, so far, has been primarily explorato-ry, aimed at determining the correlation, if any, between IQ, or psychometric *g*, and the various parameters derived from the RT (or MT) measurements, such as their *central tendency* (mean or median), *intraindividual variability* (the standard deviation of the subject's RT across trials), and the *intercept* and *slope* of the regression of RT on bits. It was thought important to establish that a correlation between RT and *g* actually exists, before proceeding further to test hypotheses about the cause of the correlation. If no significant and systematic relationships could be found, the paradigm would, of course, be abandoned in favor of other chronometric paradigms. In fact, in recent years, investigations in Jensen's labo-ratory have turned to a number of other elementary cognitive tasks which yield chronometric measures that are more reliably and substantially correlated with *g*

than are the measurements yielded by the Hick paradigm. But the Hick paradigm is still of considerable theoretical interest. It seems the simplest chronometric task in terms of its minimal cognitive demands and the absence of any content that could be called "intellectual." And it yields the shortest RTs of any of the chronometric tasks that have been used, including the Sternberg (1966) memory scan, the Posner (1978) physical and name identity recognition task, and simple sentence verification tasks. A point of theoretical interest in finding a significant correlation, even if small, between RT parameters derived from the Hick paradigm and g as measured by conventional complex psychometric tests, is that it broadens the construct of g beyond even Spearman's (1927) original conception of it as the "eduction of relations and correlates." The existence of a correlation between speed of response, or RT, in such a simple task as the Hick paradigm on the one hand, and scores on an unspeeded standard test of intelligence on the other, is a surprising and amazing phenomenon. It must ultimately be explained by a comprehensive theory of intelligence. Is it true, as Eysenck (1986) has suggested, that "there is a central core to IQ tests which is quite independent of reasoning, judgment, problem solving, learning, comprehension, memory, etc."? Our findings with the Hick paradigm appear consistent with this radical hypothesis.

Concerning the correlation between g and RT, we hold no special brief for any particular parameter of RT, as Longstreth seems to imply. We are trying only to discover which parameters of the Hick paradigm are in fact correlated with g. Naturally, we have been interested in looking first at the slope parameter, which an earlier investigation of the Hick paradigm had found to be correlated with IQ (Roth, 1964). Surely, a finding as theoretically intriguing as Roth's, that individual differences in the slope of the regression of RT on bits is correlated with IQ, warrants further attempts at replication. As our research on RT was never intended for the purpose of devising a new "IQ test" to replace conventional tests, but as a technique for further exploring the nature of g, we are less concerned with the absolute magnitude of any correlations we might find than with the existence of correlations between psychometric tests and various parameters of the Hick paradigm. The pattern of such correlations could provide clues to the nature of the g factor of conventional tests. We are not concerned, at present, with the *practical* validity of RT as a measure of general intelligence.

Longstreth places greater emphasis on the experimental manipulation of procedural features in the Hick paradigm than we have done in our studies so far. There already exists an extensive literature on the experimental psychology of reaction time (e.g., Welford, 1980) that affords grist for almost unlimited speculation concerning the experimental variables that might possibly affect performance in the Hick paradigm. One could easily increase Longstreth's many speculations on this score by an order of magnitude, but at this stage it would be an idle exercise. Investigation could be unproductively sidetracked by the almost unlimited possibility for manipulating the many experimental variables and all

their interactions, permutations, and combinations. In the study of individual differences in RT and their correlation with psychometric tests, this kind of experimental approach would be an inefficient strategy, certainly at least in the early stage of this research. We first want to find out for sure which RT and MT parameters of the Hick paradigm are reliably related to psychometric abilities under even *one* set of procedures. Once that is firmly established, experimentalists can then manipulate procedures to determine their effects on the correlation. To show that there is a zero or nonsignificant correlation between RT and *g* under some particular set of conditions would be of real interest only if it were also clearly demonstrated that a significant correlation does appear under *some* set of conditions. But one cannot manipulate experimental variables in a single correlational study, which requires the measurement of individual differences under uniform conditions for all subjects; interactions between individual differences and experimental variables could possibly so attenuate any potential correlations as to reduce the chances of finding *any* significant relationship. Also, we have generally maintained the same procedures across different study samples to enable absolute comparisons of the Hick variables obtained from criterion groups that differ in age or scholastic aptitude. With certain relationships between RT and psychometric *g* quite well established recently by means of meta-analyses of all our studies, there is now some basis for investigating how these relationships might behave under experimental manipulations of certain procedural variables. At present, however, it is clearly evident only that, with the set of procedures we have used, individual differences in measurements of RT and MT derived from performance on a particular reaction-time task—the Hick paradigm—show a significant and systematic relationship to individual differences in performance on standard *g*-loaded psychometric tests. This is of major theoretical interest, because the Hick paradigm involves no knowledge content, no reasoning, no problem solving, no "higher mental processes," in the generally accepted meaning of these terms, and it has about as little resemblance to conventional unspeeded psychometric tests as one could possibly imagine.

Now we can take up the specifics of Longstreth's critique.

ERRORS OF FACT, CITATION, AND INTERPRETATION

Before discussing the more important substantive and theoretical issues raised by Longstreth, we must correct some of the simple inaccuracies in his article.

Vernon's Studies
In two places, Longstreth reports correlations which he says come from Vernon (1983). The first, a correlation of −.22 between slope of RT and Wechsler IQ, appears nowhere in the cited reference but rather in Vernon's (1981a) doctoral dissertation on which the 1983 article was based.

A more serious citation error is Longstreth's report of correlations of −.25,

$-.27$, $-.31$, and $-.06$ between IQ and RTs obtained under four conditions of increasing complexity (0, 1, 2, and 3 bits of information). These, again, are ascribed to Vernon (1983), but not only do these correlations not appear in this paper, they were actually reported in a completely different study (Vernon, 1981b) based on a very different sample—mentally retarded subjects instead of university students. The above misidentified correlations were cited by Longstreth in contradiction of Jensen's (1979) claim that the relationship between RT and IQ becomes stronger as the complexity of the RT task increases. But Vernon (1981b) offered an hypothesis regarding why, in this particular study, the correlations increase only across the first three levels of complexity (0, 1, and 2 bits) and then drop to an insignificant value at the highest level of complexity. Vernon's hypothesis is related to the fact that the sample consisted of retarded adults, not university students as Longstreth stated (p. 154). Because Longstreth obviously had access to Vernon (1981a), he could have found that the actual correlations in the university sample between RTs under the four conditions and IQ were $-.13$, $-.25$, -26, and $-.27$, in order of increasing complexity, in accord with Jensen's claim, yet Longstreth did not report this.

In discussing the possible confounding of practice effects on RT with set size in the Hick paradigm, Longstreth stated that subjects in Vernon (1981b) were given practice trials before testing *only with a set size of one,* that is, only in the simple RT condition (0 bits). Besides the fact that this information does not appear in Vernon (1981b), it is incorrect. In each of Vernon's studies, as in most of the studies performed in Jensen's laboratory, the apparatus and procedure have first been demonstrated to subjects with the largest set size (8 bits) exposed. After subjects have practiced and familiarized themselves with this condition, the first frame is placed on the subject's response console and subjects are practiced on the one-light, simple RT condition for several trials, until the subjects indicate they understand the procedure and are ready to begin the RT test proper. Although the details of the familiarization procedure are not described in Vernon's (or Jensen's) articles, neither is the procedure which Longstreth surmises. The actual procedure was described by Vernon (1981b) as follows:

> For the reaction time test, subjects were instructed to press down the "home button" . . . and then watch *for one of the green lights to come on.* . . . When a light appeared, they were told to move their hand and press the pushbutton directly below the light as quickly as possible. (p. 348, emphasis added)

Virtually identical instructions are also reported in Vernon (1983). Obviously, it would be impossible for subjects to watch for *one* of the *lights* to come on if only one light were exposed. Also, contrary to Longstreth's conjecture, it is neither difficult nor "counter-productive" to introduce subjects to the task with all eight lights exposed. Even the mentally retarded subjects in Vernon's (1981b) study

were able to perform adequately in the Hick paradigm with the same rather minimal instructions and practice that were used in studies with university students.

Jensen's Studies.

Longstreth stated, "Increases in IT [i.e., RT] as a positive function of set size are interpreted by Jensen as reflecting the increasing time required to program the movement response" (p. 155). He follows this statement with a one-sentence quotation from Jensen (1982, p. 102) that actually says nothing at all about the programming of the movement response being mainly responsible for the increase in RT as set size increases. Moreover, the whole passage, only about half of which is quoted by Longstreth, indicates that Jensen's conclusion on this point actually *contradicts* the position attributed to him by Longstreth. Also, Jensen (1982, p. 103) presented a graph (reproduced in Longstreth's critique) showing that the increase in RT with increasing set size occurs even when no movement response is required of the subject; the subject merely removes his finger from the home button when the reaction stimulus (RS) appears. As if he were refuting Jensen's conclusion, Longstreth even points out that the RT gradients in Jensen's graph do not differ much in slope whether a movement response is required or not. Any reader who is unfamiliar with what Jensen had actually written about this finding would surely be misled into believing that Jensen had misinterpreted his own data. Yet here is what Jensen (1982) actually stated, only the first sentence of which was quoted by Longstreth:

> When the S is required to make the ballistic response to turn out the light, he apparently cannot remove his finger from the "home" button (i.e., RT) until the ballistic response has been "programmed"; the RT under the double response condition thus reflects in part the programming time for the execution of the specific ballistic response required. [*End of Longstreth quote.*] This outcome is highly suggestive of Fitts' law, which essentially relates the time for beginning the execution of a movement to the required precision of the movement (Fitts, 1954). *The ballistic movement programming time of about 30 ms is only slightly affected by the number of response alternatives [i.e., set size]. The slope of RT over bits [set size] is mainly a function of uncertainty about the RS* [i.e., reaction stimulus]. (pp. 102–103, emphasis added).

The meaning of Jensen's statement is obviously just the *opposite* of Longstreth's representation of it as quoted at the beginning of this section.

Jensen and Munro (1979) reported a correlation of $-.30$ ($p < .06$) between the slope of RT (across 0 to 3 bits) and IQ in a group of 9th-grade girls. Corrected for attenuation, the correlation was increased to $-.36$ ($p < .05$). This is all perfectly legitimate. But Longstreth pokes fun at the idea of testing a disattenuated correlation coefficient for significance, writing, "Lo and behold, it is now

significant! Nevermind that, at least as far as I know, there is no test of significance for corrected correlation coefficients'' (p. 152). It seems a trivial point for such a dramatic reaction in this context, with the significance level changing only from .06 to .05 as a result of the correction for attenuation. Besides, Longstreth is simply wrong here. If he had troubled himself to find this out, we would have been spared his song and dance on this little point. In fact, a disattenuated correlation *can* properly be tested for significance; of course, its standard error and any given confidence interval are slightly larger than those of the uncorrected *r,* but formulas for these have been explicated in the statistical literature, including a well-known textbook (Kelley, 1947), for at least 60 years. (See Forsyth & Feldt [1969] for a discussion of the significance of disattenuated correlations.)

Longstreth (p. 143) quotes a sentence by Jensen, citing ''(Jensen, 1979, p. 105),'' but the quoted sentence is not in the referenced article, and there is no other article by Jensen with this combination of date and page number. The quoted passage is actually from Jensen (1982, p. 105).

LONGSTRETH'S ''MINI-EXPERIMENTS''

Longstreth's own ''mini-experiments'' (so-called no doubt because of their very small *N*s) are purported to discredit some of the findings of Jensen's studies. However, these mini-experiments do not attempt to address the main issue, that is, the relationship between IQ and the RT and MT variables of the Hick paradigm. With such small *N*s, they could not be expected to do so. But we must discount *any* conclusions based on these mini-experiments as irrelevant and ungeneralizable to the findings obtained with Jensen's RT–MT apparatus and procedures. Longstreth conducted his mini-experiments with what he terms a ''modified Jensen'' apparatus. ''Modified'' is a masterpiece of understatement. Even if one intentionally tried to devise a reaction-time apparatus and procedure that is as utterly different from Jensen's apparatus as possible, it would be difficult to invent one more different than Longstreth's ''modified Jensen'' apparatus. It is indeed ironic that, despite all of Longstreth's apparent cautions and warnings about methodological ''dangers'' concerning the sensitivity of procedural and apparatus variables in this kind of research, he should think that the results of his mini-experiments with this extremely different apparatus and procedure could possibly have any cogency whatever regarding the interpretation of Jensen's data. Longstreth's RT apparatus: (1) has no ''home'' button; (2) does not measure RT and MT separately, but a composite of the two; (3) uses digits at the center of a video monitor as the reaction stimuli, rather than spatially distinct, green jeweled lights; (4) presents all the reaction stimuli in the same location (in the center of the video monitor), rather than in different positions all equidistant from the ''home'' button; (5) has four, rather than eight, response buttons; (6) apparently uses no auditory (''beep'') preparatory stimulus; and (7) has a relatively low degree of what is known in choice-RT experiments as ''stimulus-

response compatibility,'' whereas in Jensen's apparatus each of the response buttons is located ½ in. directly below each of their corresponding stimulus lights (see Figure 1 in Longstreth's article). And remember, these are only the few most striking differences between Jensen's apparatus and Longstreth's "modified Jensen" apparatus. It would be otiose to make any further reply to the criticisms of Jensen's RT research that Longstreth has surmised from his scarcely relevant "mini-experiments" with his "modified Jensen" apparatus.

EFFECTS OF PROCEDURAL VARIABLES
IN JENSEN'S STUDIES

Longstreth's critique capitalizes on two truisms in research methodology: (1) In *any* given experiment, one can always list a host of uncontrolled variables, some possibly important, many undoubtedly trivial; (2) the theoretical interpretation of *any* empirical finding in isolation from other findings is necessarily ambiguous.

Longstreth names three types of variables that he supposes are important in our studies with the Hick paradigm, although he is not explicit about just how these variables are important as regards the main purpose of these studies, namely, to determine if there is a relationship between IQ (or psychometric g) and RT measured under varying degrees of stimulus complexity (bits of information) in a task devoid of the kinds of knowledge content, reasoning, and problem-solving demands found in conventional tests. The effects Longstreth emphasizes as problematic are: (1) *order effects,* i.e., the confounding of *order* of presentation of set sizes (always in the order of 0, 1, 2, 3 bits) with the effects of set size per se; (2) *visual attention effects,* or the fact that RT increases as a function of the degree of displacement of the visual reaction stimulus from the fovea of the retina, and in Jensen's apparatus the average angular separation between the stimulus lights (arranged in a semicircle with a 7-in. radius from the home button) increases with set size; and (3) *response bias,* or the possibility that movement toward pushbuttons at different locations (e.g., left or right) may require different amounts of preparation time, which would be confounded with RT per se.

These are all reasonable conjectures, but we will explain why specific experimental investigations of these effects in their own right have not been viewed as top priority in the first stage of this research.

In the case of visual attention effects and response bias, as conceived by Longstreth, it seems most likely that interactions between individual differences in these factors and in speed of information processing per se would only constitute "noise" in the RT measurements of the latter, which would weaken its correlation with g. Longstreth has not claimed that these potentially confounding visual and response effects are themselves responsible for the correlations obtained between RT or MT variables and g. On the contrary, if anything, these

effects would attenuate the correlation. The really important point is that the correlation appears significantly *in spite of* these "noise" factors.

Order effects, on the other hand, could more likely be causally implicated in the correlation between RT *slope* and *g,* presumably because of learning, or practice, effects. That is, the subject's performance on each of the four successively administered tasks, going from 0 to 3 bits, would have the benefit of practice on the prior task, and fast learners would evince more transfer on the later tasks than slow learners. This would decrease the slope of the regression of RT on bits more for fast learners than for slow learners. This could result in a negative correlation between slope and IQ, assuming both that the faster learners have higher IQs than the slower learners in this task and that practice has a substantial effect on slope within the short time frame of the entire test (typically, 15 trials on each set size).

Practice and Order Effects

As Longstreth states, the order of set-size presentation has been 1, 2, 4, and 8 bits in every study. Varying the order of set size across studies would make them not comparable or cumulative, which seemed a more important consideration for establishing a relationship between RT and IQ. Randomizing or balancing order could also add "noise" to the measurement of individual differences, an inefficient procedure when looking for correlations that were not expected to be large. Of course, if we had found any indication of substantial practice effects in this task, investigation of their effect on slope would have been given higher priority. Obviously, systematic experimental variation of order effects would provide the only definitive answer to this question, but research priorities are based on an assessment of the probable importance of variables. The minor cleanup operations concerning the less important variables can always come later.

Our RT data have shown such negligible practice effects in this simple task as to have warranted a low priority for their further investigation. The presently available evidence all pertains only to practice effects over trials *within* set sizes, rather than from one set size to another, which could be determined only by systematically manipulating the order of the four set sizes. However, if practice effects *within* set sizes are found to be negligible, it seems most unlikely that there would be appreciable transfer of practice effects *between* set sizes, as conjectured by Longstreth. There is no apparent rationale for the ad hoc hypothesis that practice effects would exist *between* set sizes but not *within* set sizes.

Let us look at some RT data from several different studies based on independent samples, all college-age students tested on the Jensen apparatus, totaling 569 subjects. Table 1 shows the means of subjects' median RTs to the first 3 and the last 3 of 15 trials, which is the typical number of trials on each set size used in Jensen's studies. There is no discernable evidence here of a practice effect. The RT parameters of interest—intercept and slope—scarcely (and inconsistently)

　　　　　　　　　　JENSEN AND VERNON

TABLE 1

Means of Median RTs to First 3 and Last 3 of 15 Trials in 2 Samples

	Bits						
	0	1	2	3	Intercept	Slope	r
Sample 1							
(N = 100)							
First 3 trials	312.8 ± 8.6[a]		360.4 ± 11.0	389.3 ± 10.5	314.1	24.7	.996
Last 3 trials	315.2 ± 9.4		368.3 ± 11.1	388.0 ± 10.6	319.6	23.7	.989
Sample 2							
(N = 106)							
First 3 trials	309.0 ± 8.2		368.3 ± 11.2	393.8 ± 9.7	313.7	27.5	.990
Last 3 trials	313.1 ± 10.4		375.4 ± 11.3	406.0 ± 12.9	316.0	30.3	.996

[a]Indicates the 95% confidence interval, i.e., $M \pm 1.984 SE_M$.

differ from the first to the last 3 trials, and the correlations (r), which indicate the goodness of fit of the RT means to Hick's law (i.e., the linear regression of RT on bits), are all extremely high, averaging .993 for the first 3 trials and .993 for the last 3 trials. Apparently, practice has neither positive nor negative effects on the data's conformance to Hick's law. Incidentally, it should be noted that the mean RTs yielded by Longstreth's "modified Jensen" setup (see his Table 1 and Figure 3) show a comparatively poor fit to Hick's law, with the slope correlations averaging only about .85. In 27 studies (totaling 1,850 subjects) using Jensen's apparatus, the r index of fit to Hick's law *averages* .995, $SD = .006$, which puts Longstreth's results on this index about 20 standard deviations out of line from the studies conducted with Jensen's apparatus and procedure! (A review and meta-analyses of the results of all of the studies of the Hick paradigm using Jensen's apparatus are presented elsewhere [Jensen, in press], with detailed descriptions of all the subject samples, statistics, etc. mentioned in the present paper.)

Does additional practice beyond the usual 15 test trials at each set size have an appreciable effect? One study (N = 103) which used 30 consecutive trials at each of three set sizes (1, 4, and 8 light/buttons) was analyzed as shown in Figure 1. Practice effects should be quite evenly distributed over the odd- and even-numbered trials, and indeed the odd–even differences in RT (and MT) are small, as shown in the left panel of Figure 1. If practice had an appreciable effect, it would be expected to show up as a difference between the mean RTs (and MTs) obtained in the first set of 15 trials and the second set of 15 trials. But this difference, as we see in the panel on the right, is hardly distinguishable from the odd–even comparison. The regression of RT on bits under these four conditions is shown in Table 2. Again, there is no evidence of practice effects.

What about practice effects from one test session to another, separated by 1 or
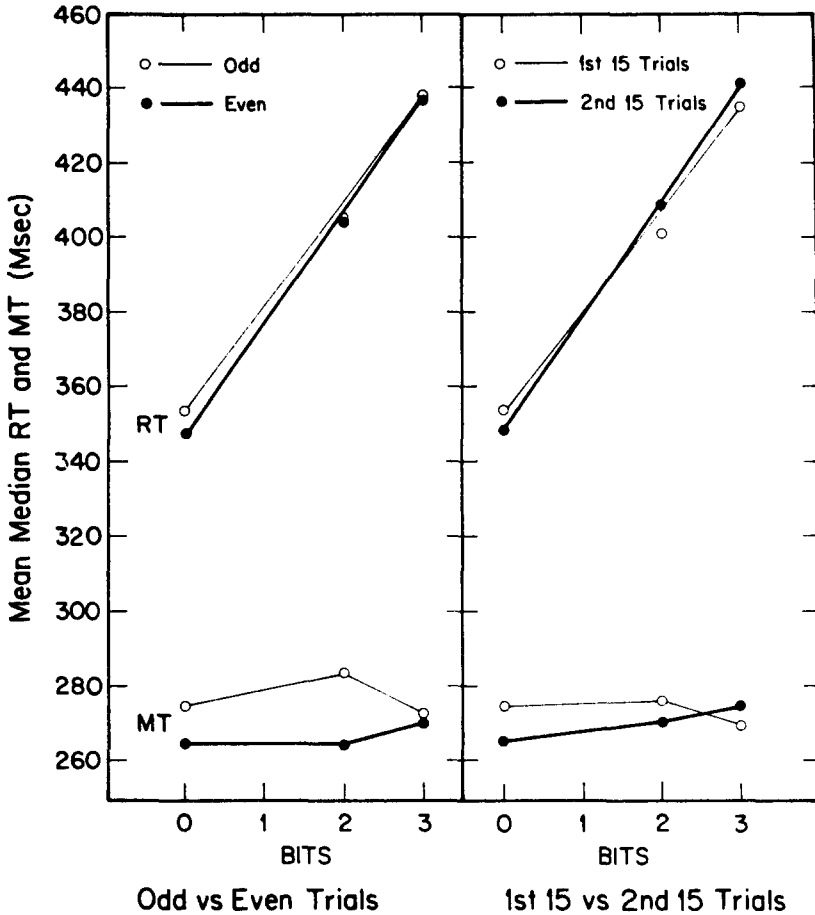
FIG. 1.   RT and MT on odd and even trials of 30 trials (left panel) and on the first set of 15 and second set of 15 trials (right panel), based on $N = 103$.

2 days? The results are shown in Table 3, along with the correlations between the Day 1 and Day 2 measurements of intercept and slope. The intercepts differ significantly ($p < .05$), but the 1 ms difference between the two slopes is nonsignificant. Another study ($N = 50$) showed the following regressions for two sessions of 15 trials on each set size on different days:

Day 1: 306.47 + 26.96(bits), $r = .989$
Day 2: 303.23 + 23.89(bits), $r = .996$

Neither the intercepts nor slopes differ significantly ($t < 1$) between the 2 days. The only statistically significant practice effects we have found are for test

TABLE 2
Regressions of RT on Bits and Index of Fit (r)
to Hick's Law for Conditions in Figure 1 (N = 103)

Condition	Intercept	Slope	Fit (r)
Odd trials	353.5	27.9	.997
Even trials	347.5	30.1	.999
First 15 trials	352.5	26.8	.994
Second 15 trials	348.4	31.1	1.000
Mean	350.5	29.0	.998

sessions extending beyond 2 days, which far exceeds the time frame of the test as typically used in Jensen's studies. Ten subjects were given 15 trials at each set size (0, 1, 2, 3 bits) every other weekday for 9 days. Figure 2 shows the mean intercept of the regression of each subject's median RT on bits across 9 days of practice. The difference between the first 2 days is not significant (correlated $t = 1.55$), but the overall differences between days is significant, $F(8,72) = 3.53, p < .01$, attributable to the difference between the means of the first 2 days and the mean of the last 7 days; there are no significant differences within each of these two sets. The average linear decrement in RT intercept over all 9 days of practice is only 2 ms per day. Figure 3 shows the mean slope of the regression of RT on bits across 9 days of practice. Again, the difference between the first 2 days is nonsignificant (correlated $t = 1.45$), but the overall differences across all 9 days are significant, $F(8,72) = 2.21, p < .05$. The average linear decrement in slope across all 9 days of practice amounts to only 0.45 ms per day. (These data on practice effects are analyzed in further detail by Jensen, in press.) With such small practice effects over trials and over days found *within* set sizes, it would seem highly improbable that there would be any larger practice effects operating *between* set sizes, and so looking for such an unlikely effect has had a low priority in our research.

If such an effect, due to the interaction of practice and order of set size

TABLE 3
Mean Intercept and Slope of Regression of RT on Bits (0, 1, 2, 3)
in Sessions of 15 Trials Each on 2 Days (N = 200)

Occasion	Intercept	Slope	Fit (r)
Day 1	300.2 ± 4.7[a]	25.6 ± 1.5	.998
Day 2	294.3 ± 4.3	26.6 ± 1.4	.980
Correlation, r_{12}	+0.721	+0.341	

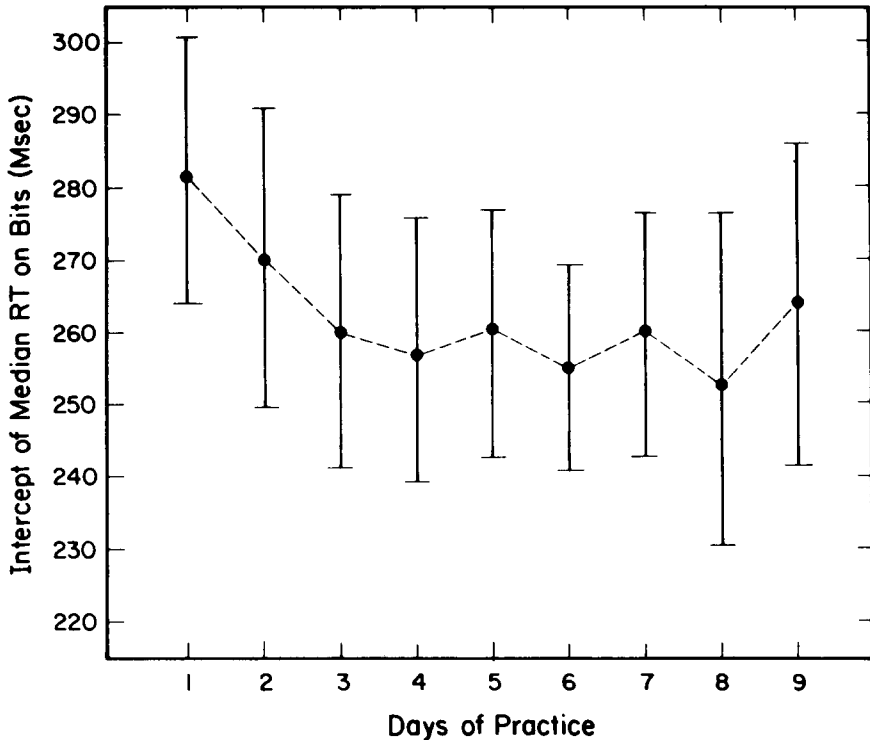[a]Indicates 95% confidence interval, i.e., $M \pm 1.96SE_M$.

FIG. 2.   Intercept of the regression of median RT on bits over 9 days of practice by 10 subjects. The vertical lines indicate the 95% confidence invervals, with 9 df (i.e., $M \pm 2.262SE_M$).

presentation, as hypothesized by Longstreth, actually amounted to anything, it should cause a departure from the linear regression of RT on bits, referred to as Hick's law. Such a departure from Hick's law would also be accentuated, according to Longstreth, by including the *simple* RT condition, that is, 0 bits of information. Longstreth apparently thinks that only RT for different degrees of *choice* will conform to Hick's law. This question is examined in considerable detail by Jensen (in press), and the conclusion, based on 24 independent samples comprising 1,556 subjects, is that both the intercept and the slope of RT on bits are only negligibly affected by the inclusion or exclusion of the RT measurement at 0 bits. Table 4 shows the unweighted and N-weighted means of the intercept, slope, and index of fit (r) to Hick's law under two conditions: (1) when RT at 0 bits is included and (2) when it is excluded.

Transfer of practice effects between set sizes administered in a constant ascending order for all subjects would be expected to distort the *linearity* of the increase in RT as a function of set size, resulting in a negatively accelerated
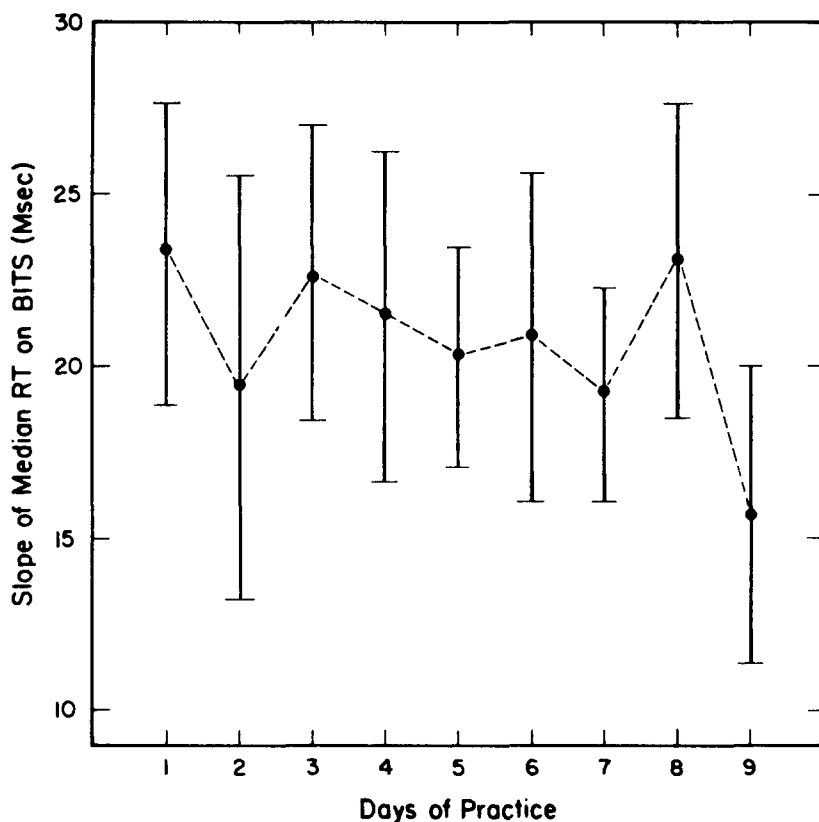
FIG. 3. Slope of the regression of median RT on bits over 9 days of practice by 10 subjects. The vertical lines indicate the 95% confidence intervals.

TABLE 4
Unweighted and N-Weighted Means and Standard Deviations
(Over 24 Samples, with Total N = 1,556) of Intercept, Slope, and Index
of Fit (r) to Hick's Law when RT (in ms) at 0 Bits Is Included or Excluded

Parameter	Condition RT at 0 Bits	Unweighted		N-Weighted	
		M	SD	M	SD
Intercept	Included	333.40	65.90	336.55	62.38
	Excluded	336.75	60.82	341.10	57.94
Slope	Included	33.82	21.47	34.11	19.58
	Excluded	32.27	26.64	32.22	24.31
Fit (r)	Included	.994	.007	.994	.006
	Excluded	.998	.012	.998	.010

166

curve instead of a straight line. Do the RT data obtained with Jensen's apparatus and procedure show this curvature, or any other departure from the linearity known as Hick's law? To examine this, we have averaged the RT data of 27 independent samples totalling 1,850 subjects. Figure 4 shows the unweighted and N-weighted mean of the subjects' median RTs as a function of set size scaled in bits. The regression lines are also shown, with the index of fit, $r = .998$, for both the unweighted and weighted means. The data points are extremely linear. A trend analysis was performed on the 24 samples (total $N = 1,556$) for which there were RT data at every set size (a set size of 2 was not used in two studies and set size of 8 was not used in one study). The linear trend is highly significant, $F(1,69) = 163.36$, $p < .001$, whereas all of the nonlinear variation in the data points is nonsignificant, $F(2,69) < 1$. Thus, there is no evidence of significant departure from linearity, or Hick's law, in the overall mean RT data obtained with Jensen's apparatus and procedure. (The high degree of *individuals'* conformance to Hick's law is discussed in detail elsewhere [Jensen, in press].)

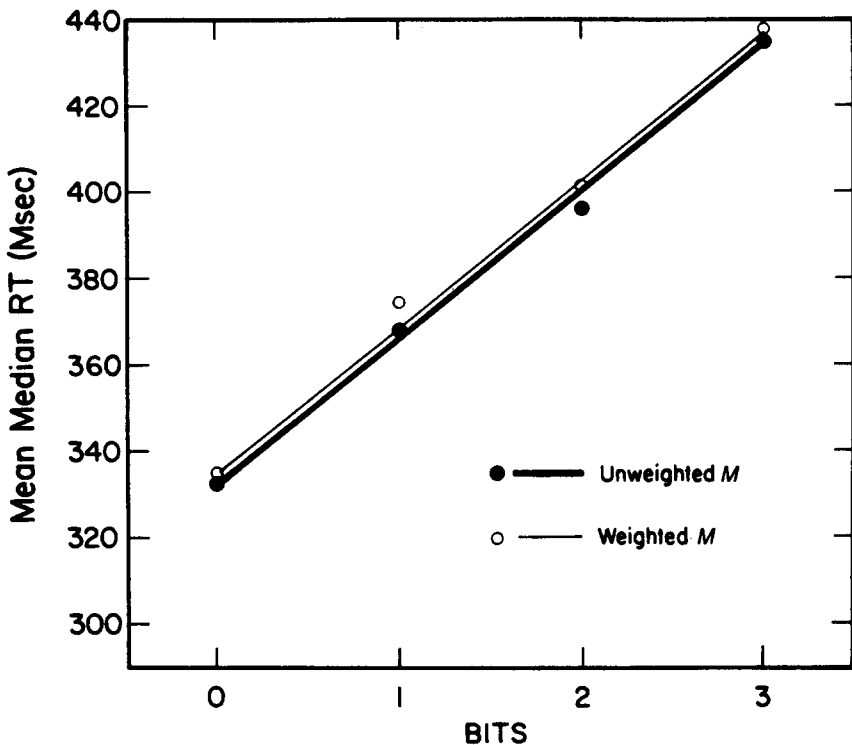There is no support in our data for Longstreth's conjecture that *simple* RT (set

FIG. 4. Unweighted and N-weighted means of the median RT as a function of bits, based on 27 independent samples comprising a total of 1,850 subjects.

size 1 or 0 bits) is somehow different from, or out-of-line with, the *choice* RTs. This is true not only in terms of mean RT, but also in terms of the pattern of intercorrelations among the RTs of every set size. The pattern of intercorrelations is a near-perfect simplex, and the correlations are highly predictable from a simple overlap, or common elements, model. The correlations involving RT at set size 1 fit the simplex model just as closely as do the correlations involving only the RTs for set sizes 2, 4, and 8. (This model and the relevant evidence are explicated in detail by Jensen [in press].) Moreover, as shown later on in Figure 6, the correlation of simple RT with IQ falls exactly on the linear regression line describing the relationship of the RT × IQ correlation as a function of set size scaled in bits.

RESPONSE BIAS

Longstreth's conjectures about response biases are so niggling as to be hardly of interest in the context of our research aims. He argues that, because the response buttons (that turn out the reaction stimulus lights) are located on a semicircle of 6-in. radius from the home button, "some of the responses may require more time for programming or preparation than others" (p. 148). If so, the resulting conjectured variation in increments in RT (i.e., the interval between the onset of a light and the subject's releasing the home button) due to preparation time for the 6-in.-movement response to the button that turns out the light would simply constitute a variable error component in the RT measurements. This "noise" would only attenuate correlations between RT and IQ. This component of error variance due to the supposed response bias, however, would tend to be averaged out over a number of trials, because each response button is targeted an approximately equal number of times.

We have obtained data on the mean RT to each of the 8 light/buttons when subjects are each given 15 trials on set size 8. The results are shown in Figure 5. Longstreth suggests that a left-moving response may take longer to program than a right-moving response, or vice versa. But no very regular pattern is apparent in Figure 5. Overall, the differences are significant, with our large sample size ($N = 309$), although the largest difference between any pair of means is only 20 ms, and the proportion of the total variance accounted for by the main effect of Positions is only .01. The proportion of variance in RT attributable to the interaction of Subjects × Positions is .19. The proportion of variance accounted for by the main effect of Subjects (i.e., individual differences in mean RT) is .80. It is not clear how these small position effects would affect the main object of our investigations, namely, the relationship between RT parameters and IQ, except possibly to *attenuate* the obtained correlations. Position effects certainly have not distorted the overall conformity to Hick's law, although, of course, we cannot rule out the possibility that some individual subjects' deviations from Hick's law could be due to the idiosyncratic position effects that
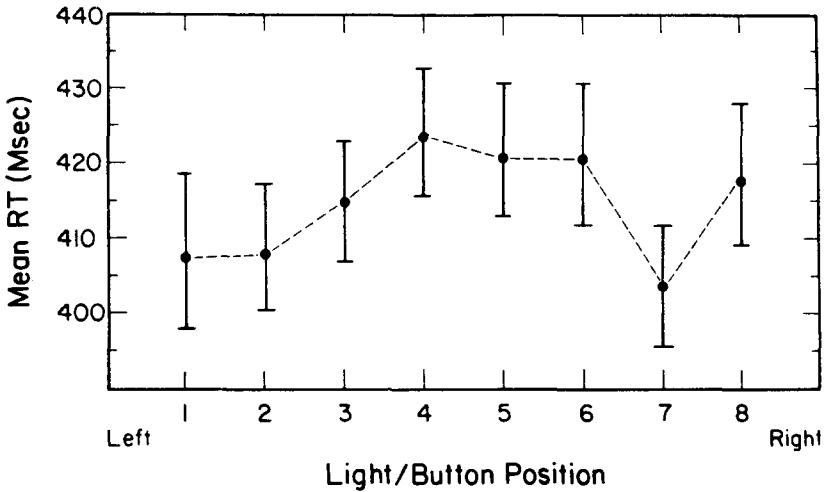
FIG. 5.   Mean RT (of 309 subjects) to each of the light/button pairs in set-size 8; the light/button pairs were arranged in a semicircle of 6-in. radius, and are numbered here 1 to 8, from left to right. The vertical lines are the 95% confidence intervals (i.e., $M \pm 1.96SE_M$).

constitute the Position $\times$ Subjects interaction. Barrett, Eysenck, and Lucking (1986) have made the significant discovery that those subjects who deviate the most from Hick's law are the ones who most attenuate the correlation between RT and IQ, and excluding these nonconformists from the data *increases* the RT $\times$ IQ correlation.

THE SLOPE–IQ CORRELATION

Now we come to the two main issues in Longstreth's critique: (1) the negative correlation between the RT slope parameter and IQ, and (2) the RT–IQ complexity relationship, that is, the negative correlation between RT and IQ as an increasing function of stimulus complexity, or set size. Longstreth considers the claims for both of these correlations unwarranted.

First, the slope–IQ correlation. Longstreth presents a graph from Jensen (1982) as his Figure 4. Jensen (1982, p. 111) had stated that the slopes of the five groups, *A* through *E*, all differed significantly except groups *A* and *B*. Longstreth (p. 151) expresses doubt that *any* of the groups *A* through *E* differ significantly in slope. Both Jensen and Longstreth were in error on this point. Although the *regressions* (i.e., intercepts *and* slopes) of all the groups (except *A* and *B*) differ significantly, they do not all differ significantly in just the *slope*. Three of the nine differences in slope that had been claimed significant by Jensen are, in fact, nonsignificant; of the remaining six differences, three are significant beyond

the .01 level and three are significant beyond the .001 level. Table 5 shows the results of two-tailed t tests for each of the paired comparisons. (The difference is always in the direction of the vertically listed group *minus* the horizontally listed group.) Some of these differences are uninterpretable theoretically, because some of the compared groups differ in age as well as in IQ. One difference that would be predicted theoretically to be significant but is not, however, is the A–D (University–Vocational College) contrast, although the difference is in the predicted direction.

However, we should look more closely at the slope differences between pairs of samples of the same age, all tested on the Jensen apparatus, and in which one group in each pair scores lower on a test of psychometric g than the other. The results of such comparisons based on all the appropriate data available are shown in Table 6. None of these samples contains retarded or borderline subjects. (Mentally retarded and borderline groups, with IQs below 80, have much greater RT slopes than any of the groups compared in Table 6.) It can be seen that the mean slopes of the regression of RT on bits differ significantly between groups that differ in mean IQ, in accord with expectation, that is, the hypothesis that RT slope reflects speed of information processing, which, in Jensen's theory, is a basic component of psychometric g.

Comparison of the means of various criterion groups, such as those in Table 6, has an advantage over the correlation coefficient in testing the relationship of slope to IQ. The reason is that slope has been found to be the least reliable of any of the RT or MT parameters derived from the Hick paradigm. Whereas the low reliability of the slope measures drastically attenuates the correlation between slope and IQ, reliability does not affect the group mean, in which random errors of measurement average out. (Measurement error, or unreliability, does increase the standard deviation, however, and consequently attenuates the difference between means when the difference is expressed in standard deviation or standardized score units [see column headed σ *diff.* in Table 6]. This standardized

TABLE 5

Significance Tests (t) of Differences[a] Between the Mean Slopes of Groups A through E in Jensen's Studies (Longstreth's Figure 4)

Group	B	C	D	E
A. University students ($N = 155$)	−1.58	−5.11**	−1.24	−4.92**
B. Ninth-Grade girls ($N = 39$)		−3.09*	+0.37	−2.92*
C. Sixth graders ($N = 50$)			+3.56**	+0.16
D. Vocational college, white males ($N = 119$)				−3.39*
E. Vocational college, black males ($N = 99$)				

[a]Differences are in the direction of the vertically listed group *minus* the horizontally listed group.

*$p < .01$ (two-tailed)

**$p < .001$ (two-tailed)

TABLE 6
RT Slope Comparisons between Samples of Comparable Age that Differ in Mean IQ

Groups	N	Mean Slope	SD	σ Diff.[a]	t	r_{pbs}^{b}
Above-average 7th graders	72	46.32	21.99			
Gifted 7th graders	60	32.73	15.84			
Difference		13.59		0.70	4.12**	−.305
Above-average 9th graders	39	30.98	10.79			
Gifted 9th graders	76	23.59	15.33			
Difference		7.39		0.53	3.00*	−.243
Vocational college students	324	32.66	16.55			
University students	530	26.48	10.37			
Difference		6.18		0.47	6.03**	−.224
Vocational college blacks	149	35.73	17.76			
Vocational college whites	175	30.04	15.52			
Difference		5.69		0.34	3.04*	−.167

[a]The difference expressed in average standard deviation units:

$$\sigma \text{ Diff} = (M_1 - M_2)/\bar{\sigma}, \text{ where}$$
$$\bar{\sigma} = \sqrt{(N_1 s_1^2 + N_2 s_2^2)/(N_1 + N_2)}$$

[b]Point biserial correlation between group dichotomy and slope, which is negatively correlated with the groups' mean IQs.

*$p < .01$ (two-tailed)
**$p < .001$ (two-tailed)

mean difference [σ *diff.*], like a correlation coefficient, can be corrected for attenuation by dividing it by the square root of the reliability coefficient.) Also, nearly all the groups that have been tested so far are quite restricted on range of ability, a condition that further attenuates the correlations.

The appropriate form of reliability coefficient with respect to the attenuation of the correlation between slope and IQ is the test-retest reliability of the slope measure. The test-retest (1 or 2 days apart) reliability of the slope measure, obtained on 260 university students, is only .37, as compared with a reliability of .72 for the intercept.

With such low reliability and a restriction of IQ range in nearly all of the subject samples that have been tested, we should expect to find very small (negative) correlations between RT slope and IQ. Small correlations due to

unreliability would, of course, be a serious negative indication for the practical usefulness of the slope measure, but such correlations, if statistically significant and if correctable for attenuation and restriction of range, may still be of considerable interest from a theoretical standpoint.

A total of 35 coefficients of correlation between RT slope in the Hick paradigm and one or more IQ or other highly g-loaded psychometric tests, based on 20 independent, or nonoverlapping, samples are presented elsewhere (Jensen, in press) with details of each of the samples, tests, and sources, along with similar information on five other Hick parameters. Of the total of 20 samples, 12 are from Jensen's laboratory (accounting for 22 of the 35 correlations, and comprising a total of 1,055 subjects) and 8 are from studies by other investigators of the Hick paradigm (accounting for 13 of the correlations and comprising a total of 503 subjects). The N-weighted mean of the 35 correlations (based on a total N of 1,558) between RT slope and "IQ" is $-.117$, with a SD of 0.132 over the 35 correlations. (The N-weighted mean r in the studies from Jensen's laboratory is $-.091$, $SD = 0.109$; the N-weighted mean r from studies by other investigators is $-.181$, $SD = 0.147$.) To be sure, these are small correlations, although they are significant beyond the .001 level. The lower mean r obtained from the studies in Jensen's laboratory is most likely attributable to the greater restriction of IQ in his samples, which have consisted mainly of university students or specially selected criterion groups in which the IQ range is fairly restricted. The largest correlation among the whole set of 35 is $-.41$, based on 91 subjects specially selected by Nettelbeck and Kirby (1983) to conform approximately to the normal distribution of IQ in the general population. But Nettelbeck and Kirby's correlation may even be slightly on the high side, due to possible sampling error in their study. The true correlation in the general population is probably closer to $-.30$. If the IQ variance within the restricted samples that were averaged is conservatively estimated at ½ of the IQ variance in the general population, and if we use the best estimate we have of the test-retest reliability of RT slope as .37, and also make the reasonable assumption of a reliability of .90 for the IQ measurements, the obtained N-weighted mean correlation between RT slope and IQ can be corrected for restriction of range and attenuation. Thus corrected by means of the appropriate formulas (Gulliksen, 1950, pp. 101, 137), the mean correlation of $-.117$ becomes $-.285$. Hence a value close to $-.3$ is probably the best available estimate of the true correlation between RT slope and conventional measures of intelligence.

It should be noted that the slope parameter indicates only the average increment in RT per bit (i.e., $\log_2$ of set size). It is a theoretically important discovery that both the intercept of RT and the mean MT, which both reflect a general speed factor, as well as the trial-to-trial intraindividual variability ($RT\sigma_i$) of RT, all show correlations with IQ that are at least as high as the correlation between RT slope and IQ, and even considerably higher in the case of $RT\sigma_i$ and MT. A general speed factor reflected in the mean RT and MT seems to be the main

factor in the correlation between IQ with the various Hick parameters. With correlations for attenuation and restricted range, these Hick parameters (intercept, slope, $RT\sigma_i$, and mean MT) show an overall multiple correlation with IQ of about .50. (See Jensen [in press] for the details of the meta-analyses of the Hick parameters in all the available studies.) This is probably the best estimate of the true overall correlation between IQ and the composite variables derived from the Hick paradigm. It is really a quite remarkable correlation, considering the relatively minimal cognitive demands of the Hick task and its lack of any resemblance to the knowledge content and intellectual skills involved in conventional IQ tests.

THE RT–IQ-COMPLEXITY RELATIONSHIP

Jensen has hypothesized that the relationship of RT to IQ increases as the complexity of the reaction stimulus increases. In the case of the Hick paradigm, this hypothesis predicts an increasing RT–IQ correlation as a function of set size, or bits. (Because RT is negatively correlated with IQ, the *size* of the correlation should increase in the *negative* direction as a function of set size.) Longstreth's selective and biased method of reviewing evidence relevant to this hypothesis would leave uninformed readers with the impression that the hypothesis is refuted, or at least that there is nothing at all in it. For example, he complains about using correlations derived from three combined groups of differing IQs, obtained in a study by Lally and Nettelbeck (1977) and cited by Jensen (1982, p. 109). The correlations between the WAIS Performance IQ and RT for set sizes 2, 4, 6, and 8 were $-.55$, $-.63$, $-.67$, and $-.76$, respectively, in accord with the hypothesis. The correlations are, of course, considerably inflated by the great IQ heterogeneity resulting from combining three groups with IQs of 57 to 81, 90 to 115, and 116 to 130. (In the same article, Jensen [1982, pp. 107, 109] had clearly pointed out the inflation of correlations in highly heterogeneous groups in which IQs are not normally distributed, so Longstreth is not saying anything that was not already reported by Jensen.) But the essential feature of these correlations that is relevant to the hypothesis under consideration (i.e., the increase in the RT–IQ correlation with increasing set size) is not the overall *magnitude* of the correlations, but whether there is a *systematic increase* in the size of the correlations across set sizes 2, 4, 6, and 8. Longstreth apparently would prefer to diminish the size of the correlations by considering only the correlations obtained *within* each of the highly restricted groups. The within-group correlations should be drastically reduced, of course, because the IQ variance *within* each of the groups is only a fraction of the total IQ variance of the combined groups, in fact, only 9.7%, 7.1%, and 7.1%, respectively, of the combined-groups variance. Yet, even when just the *within*-group correlations between RT and IQ are properly averaged (via Fisher's Z transformation) over the three groups, they are $-.15$ $-.23$, $-.27$, and $-.40$, respectively, for set sizes 2, 4, 6, and 8. This

result accords with Jensen's hypothesis, and, in fact, these mean correlations are correlated $-.94$ with bits, indicating a strong *linear* trend. A study by Nettelbeck and Kirby (1983) is also cited by Longstreth, who mentioned only those sets of correlations from Nettelbeck and Kirby's table (p. 52) that are *not* in accord with Jensen's hypothesis. Longstreth does not mention the most representative set of correlations, which is based on 91 subjects selected from each of the three IQ-restricted samples so as to approximate the full range and normal distribution of IQs in the general population. This normally distributed sample shows correlations of ascending size: $-.63$, $-.65$, $-.67$, having a linear correlation with bits of $-.87$, again in accord with the IQ–RT-complexity hypothesis.

But let us move on to some statistics based on much larger studies, and look at the mean within-group correlations obtained in all the studies in the literature (not including the studies by Lally and Nettelbeck [1977] and Nettelbeck and Kirby [1983] already mentioned). We have found a total of 31 RT–IQ correlations (19 from Jensen's laboratory) at each of 0, 1, 2, and 3 bits on the Jensen apparatus, based on 15 independent samples comprising a total of 1,129 subjects. The unweighted and N-weighted mean correlations are shown in Figure 6. The correlation between RT and IQ can be seen clearly to increase linearly as a function of bits. (Also, each of the correlations in Figure 6 is significant beyond the .001 level.) An analysis of variance for trend shows that the linear trend of these mean correlations is highly significant ($p < .001$) and that the nonlinear
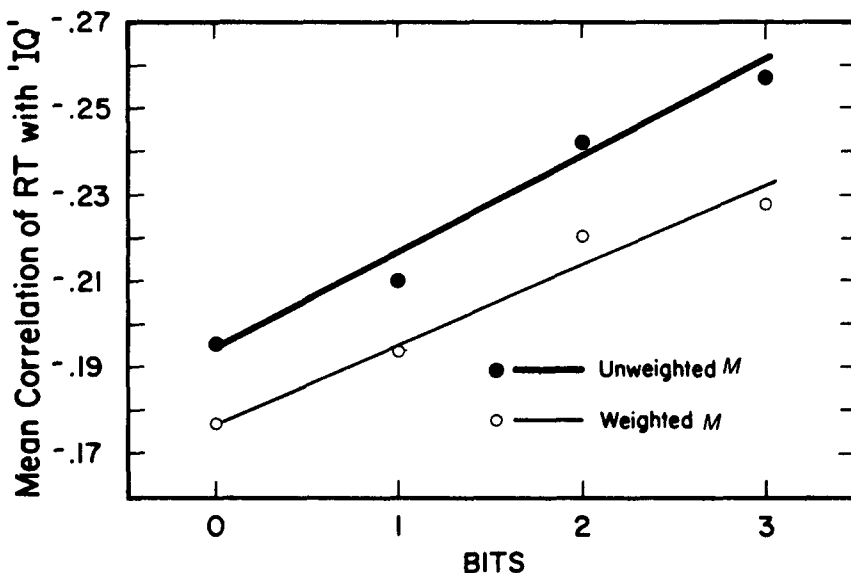

FIG. 6.   Mean unweighted and N-weighted correlations between RT and IQ as a function of set size scaled in bits. The means are based on 31 correlations at each set size obtained in 15 independent samples totalling 1,129 subjects.

components of variance in the data points are nonsignificant ($F < 1$). The linear correlation between the mean RT–IQ correlations and bits is $-.99$ for the unweighted means and $-.98$ for the weighted means. Incidentally, the correlations between MT and IQ, when analyzed in the same way, show no increase at all as a function of bits. (Details of all the studies entering into this meta-analysis are presented elsewhere [Jensen, in press].)

It is also noteworthy, in reference to the hypothesized RT–IQ-complexity relationship, that we have devised other RT tasks that are somewhat more complex than the Hick paradigm and have found considerably higher correlations with these more complex RT tasks than with the Hick paradigm (e.g., Anada, 1985; Cohn, Carlson, & Jensen, 1985; Vernon, 1983; Vernon & Jensen, 1984).

STRATEGIES AND SPEED–ACCURACY TRADE-OFF

Longstreth (p. 157) suggests that the RT–IQ correlation could be due, at least in part, to individual differences in speed–accuracy trade-off. That is, the brighter subjects presumably figure that they can "beat the game" by sacrificing response *accuracy* for an increased *speed* of response. But this hypothesis is unconvincing for two main reasons. First, with the use of a home button, which the subject merely has to release, response accuracy with respect to RT is virtually assured; inaccuracy of response would show up as a failure to touch the button that turns off the reaction stimulus light, thereby affecting movement time, not RT. So simple is the RT task that such response errors are extremely rare in this procedure and are completely absent in the vast majority of subjects. Second, and more important, is the fact that if a speed–accuracy trade-off accounted for any part of the individual differences in either RT or MT, there should be a *negative* correlation between response error rate and RT (or MT). Yet all RT studies in which both RTs and error rates have been measured show a *positive* correlation between RT and errors. The faster responders are also the more accurate responders. We have found no evidence in our RT studies or in anyone else's for a *between*-subjects speed–accuracy trade-off. Speed–accuracy trade-off is a *within*-subjects phenomenon, accounting for *negative* correlations (*within* subjects) between RTs and error rates under different levels of task difficulty. It is not a problem in interpreting the correlation between individual differences in RT and IQ because the *between*-subjects correlation of RT and error rate is a *positive* correlation, and both RT and error rate are negatively correlated with IQ. These relationships can be explained more easily with reference to Figure 7. On the *simple task*, hypothetical persons A, B, and C are shown to have the same short RT and low error rate. On the *complex task*, the latent ability differences between A, B, and C are manifested as variation in their RTs and error rates. Their performances, as reflected jointly by RT and errors, will tend to fall somewhere on each of the arcs that describe the speed–accuracy trade-off and are different for each person. If the same low error rate of the
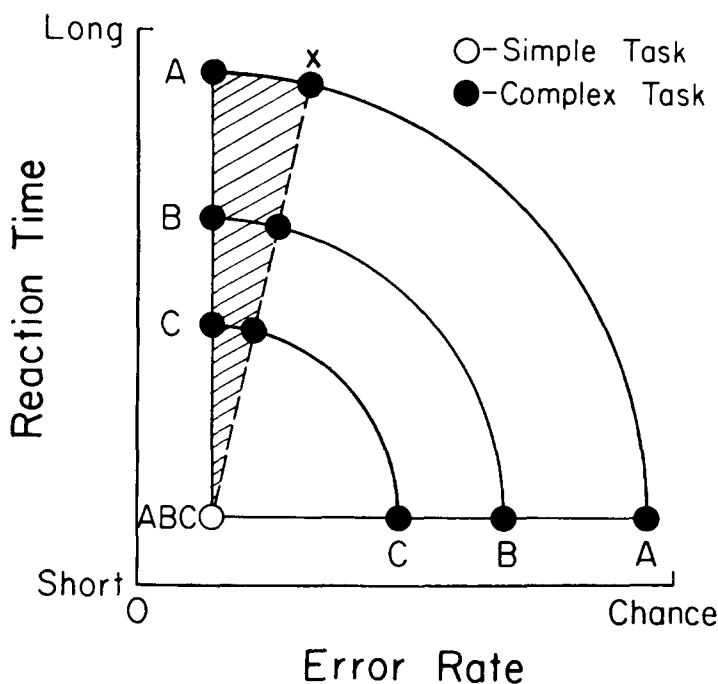
FIG. 7. The idealized relationship between RT and error rate for simple and complex tasks. The arcs describe the speed–accuracy trade-off for hypothetical persons A, B, and C, who are shown here as performing equally on the simple task. Shaded area represents the most desirable region of speed–accuracy trade-off for RT studies.

simple task is to be maintained for the complex task, the RT is greatly increased for all persons (vertical line = zero speed–accuracy trade-off). If the RT in the simple task is to be maintained in the complex task, the error rate is greatly increased for all persons (horizontal line = 100% speed/accuracy trade-off). So the arc for each person describes an *inverse* relationship (or *negative* correlation) between RT and error rate. But *between* persons, the line marked *X* in Figure 7 indicates a fairly high speed–accuracy trade-off for a typical RT study, if the error rate (on the abscissa) is assumed to range between *zero* and *chance*. Thus the shaded area represents the most desirable region for performance when studying the individual differences in RT in that it spreads out individual differences in RT much more than in error rate, a feature observed in all of our RT studies. Hence the observed correlation between RT variables and IQ can in no way be accounted for in terms of speed–accuracy trade-off.

Longstreth (p. 157) refers to Nettelbeck and Kirby (1983) to bolster his contention that some kind of performance strategy effect, rather than speed of information processing per se, may be responsible for the correlations obtained

between the RT parameters derived from Jensen's use of the Hick paradigm and IQ. The discussion by Nettelbeck and Kirby on such a possible strategy effect is based entirely on their observation of the *negative* correlations in their data between individual differences in the *slope* and the *intercept* of the regression of RT on bits (see Table 5 of their article). Nettelbeck and Kirby state:

> There are aspects to these data which are concordant with the possibility that certain strategies for responding have influenced outcome. In the first place, the strong negative correlations between the slope and intercept of regression functions for DT [i.e., RT] and MT within all three groups raise doubts about whether the regression of RT on bits can reliably distinguish rate of processing from fundamental delays in the subject's response system, as Hick's Law proposes. *If it were the case that both variables were interacting with a third, like intelligence, then one would expect a* positive *correlation between slope and intercept.* The negative relationship suggests instead that some subjects have applied different criteria for responding at different levels of choice. One plausible possibility is that some responses have been disproportionately more carefully made when eight stimulus alternatives were involved, although other explanations are equally viable. The important consideration here is that strategies of this kind could increase the slope while decreasing the intercept of the regression function. (pp. 49–50, emphasis added)

The strategy explanation that Nettelbeck and Kirby propose for their observed negative correlations between intercept and slope, however, is quite unnecessary, because the negative correlation that they are trying to explain is largely, perhaps entirely, a statistical artifact, due to the negatively correlated errors of measurement in the slope and intercept. Because measurement error tends to be averaged out in a group's mean when the N is reasonably large, we can observe the correlation between intercept and slope with minimal error by correlating these parameters as the means of a number of different groups, rather than as individual measurements. When we do this, we find that the mean intercepts and mean slopes obtained on 27 independent samples (totalling 1,850 subjects tested on the Jensen apparatus) are *positively* correlated (Pearson $r = +.71$, Spearman rank-order correlation $= +.55$). (The corresponding intercept $\times$ slope correlation based on the means of each of the three groups in the Nettelbeck and Kirby study is $+.99$.) It should be noted that a *positive* correlation between intercept and slope is what Nettelbeck and Kirby expected on the hypothesis that both intercept and slope of RT are correlated with IQ because both of these RT parameters reflect mental processing speed.

But why, then, are intercept and slope *negatively* correlated in *individual* data? We know from the rather low reliabilities of intercept and especially slope that there is considerable error in the measurement of individual differences in these parameters. And when intercept and slope are calculated on the same data,

they share the same errors of measurement, which are *negatively* correlated. That is, measurement error, or random variability, has *opposite* effects on the intercept and the slope; the same errors that increase one variable necessarily decrease the other. The correlation between intercept and slope due solely to their shared errors of measurement can be precisely calculated from the following formula (from Marascuilo & Levin, 1983, p. 161):

$$r_{IS} = \frac{-\bar{X}}{\sqrt{\dfrac{\Sigma X^2}{N}}}$$

where X is the values on the abscissa (in this case, bits) and N is the number of bivariate data points on which center the computation of the regression of the ordinate values on the abscissa values (in this case, the regression of median RT on bits). With 0, 1, 2, 3 bits, as typically used in Jensen's studies of the Hick paradigm, $r_{IS} = -.80$. The true (i.e., error-free) positive correlation between intercept and slope is obscured by this large negative correlation between their errors of measurement, and hence the observed correlation between intercept and slope, measured on individuals' RTs, is often diminished even to the point of being a negative coefficient. If one wants to talk properly about the true correlation between intercept and slope, it is essential that the correlation be based on experimentally independent measurements of each variable, so that their random errors of measurement will be uncorrelated.

Because intercept and slope are negatively correlated through their sharing of the same errors of measurement, one can reduce the effect of measurement error on the correlation between slope and IQ by statistically controlling the variance in intercept, by partial correlation. The resulting partial correlation to some extent, but not completely, rids the slope–IQ correlation of attenuation due to unreliability, and, because the slope-intercept relationship is negative, therefore enlarges the theoretically predicted negative correlation between RT slope and IQ.

REFERENCES

Ananda, S. M. (1985). *Speed of information processing and psychometric abilities in later adulthood.* Unpublished doctoral dissertation, University of California Berkeley.

Barrett, P., Eysenck, H. J., & Lucking, S. (1986). Reaction time and intelligence: A replicated study. *Intelligence, 10,* 9–40.

Cohn, S. J., Carlson, J. S., & Jensen, A. R. (1985). Speed of information processing in academically gifted youths. *Personality and Individual Differences, 6,* 621–629.

Eysenck, H. J. (1986). Intelligence and reaction time: The contribution of Arthur Jensen. In S. Modgil & C. Modgil (Eds.), *Arthur Jensen: Consensus and controversy.* Barcombe, England: Falmer Press.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology, 47*, 381–391.

Forsyth, R. A., & Feldt, L. S. (1969). An investigation of empirical sampling distributions of correlation coefficients corrected for attenuation. *Educational and Psychological Measurement, 29*, 61–71.

Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology, 4*, 11–26.

Jensen, A. R. (1979). g: Outmoded theory or unconquered frontier? *Creative Science and Technology, 2*, 16–29.

Jensen, A. R. (1982). Reaction time and psychometric g. In H. J. Eysenck (Ed.), *A model for intelligence*. Heidelberg: Springer-Verlag.

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

Jensen, A. R., & Munro, E. (1979). Reaction time, movement time, and intelligence. *Intelligence, 3*, 121–126.

Kelley, T. L. (1947). *Fundamentals of statistics*. Cambridge, MA: Harvard University Press.

Lally, M., & Nettelbeck, T. (1977). Intelligence, reaction time, and inspection time. *American Journal of Mental Deficiency, 82*, 273–281.

Longstreth, L. E. (1984). Jensen's reaction time investigations of intelligence: A critique. *Intelligence, 8*, 139–160.

Marascuilo, L., & Levin, J. R. (1983). *Multivariate statistics in the social sciences: A researcher's guide*. Monterey, CA: Brooks/Cole.

Nettelbeck, T., & Kirby, N. H. (1983). Measures of timed performance and intelligence. *Intelligence, 7*, 39–52.

Posner, M. I. (1978). *Chronometric explorations of mind*. Hillsdale, NJ: Erlbaum.

Roth, E. (1964). Die Geschwindigkeit der Verarbeitung von Information und ihr Zusammenhang mit Intelligenz. *Zeitschrift für Experimentelle und Angewandte Psychologie, 11*, 616–622.

Spearman, C. (1927). *The abilities of man*. New York: Macmillan.

Sternberg, S. (1966). High speed scanning in human memory. *Science, 153*, 652–654.

Vernon, P. A. (1981a). *Speed of information-processing and general intelligence*. Unpublished doctoral dissertation, University of California, Berkeley.

Vernon, P. A. (1981b). Reaction time and intelligence in the mentally retarded. *Intelligence, 5*, 345–355.

Vernon, P. A. (1983). Speed of information processing and general intelligence. *Intelligence, 7*, 53–70.

Vernon, P. A., & Jensen, A. R. (1984). Individual and group differences in intelligence and speed of information processing. *Personality and Individual Differences, 5*, 411–423.

Welford, A. T. (Ed.). (1980). *Reaction times*. New York: Academic Press.