THEORETICAL REVIEW

Mental chronometry and individual differences: Modeling reliabilities and correlations of reaction time means and effect sizes

Jeff Miller · Rolf Ulrich

Abstract We used a general stage-based model of reaction time (RT) to investigate the psychometric properties of mean RTs and experimental effect sizes (i.e., differences in mean RTs). Using the model, formulas were derived for the reliabilities of mean RTs and RT difference scores, and these formulas provide guidance about the number of trials per participant needed to obtain reliable estimates of these measures. In addition, formulas were derived for various different types of correlations computed in RT research (e.g., correlations between a mean RT and an external non-RT measure, between two mean RTs, between a mean RT and an RT effect size). The analysis revealed that observed RT-based correlations depend on many parameters of the underlying processes contributing to RT. We conclude that these correlations often fail to support the inferences drawn from them and that their proper interpretation is far more complex than is generally acknowledged.

Keywords Individual differences · Reaction time analysis · Correlation analysis

J. Miller (✉)
Department of Psychology, University of Otago, PO Box 56,
Dunedin 9054, New Zealand
e-mail: miller@psy.otago.ac.nz

R. Ulrich (✉)
Department of Psychology, University of Tübingen, Schleichstr. 4,
Tübingen 72076, Germany
e-mail: ulrich@uni-tuebingen.de

Classically, reaction time (RT) research has compared performance across different conditions to assess the effects of various experimental manipulations and group differences on RT. Most often, the results have been summarized in terms of the mean RTs for each condition or group, although occasionally comparisons are made with respect to variances or full RT distributions (e.g., Luce, 1986; Posner, 1978; Sanders, 1998; Smith, Ratcliff, & Wolfgang, 2004).

Increasingly, RT researchers also have examined the correlations of different RT-based measures with each other and with other measures. Such correlations are of interest in at least two kinds of research. First, RTs offer a promising tool for assessing individual differences (Cattell, 1890). For example, many researchers have looked for correlations of intelligence with overall mean RT and with the sizes of particular effects (i.e., differences in mean RT) used to assess the time needed for specific mental operations (e.g., Beauducel & Brocke, 1993; Helmbold, Troche, & Rammsayer, 2007; Hunt, 1978; Keating & Bobbitt, 1978; Kirby & Nettelbeck, 1989; Neubauer, Riemann, Mayer, & Angleitner, 1997; Smith & Stanley, 1983; Vernon & Mori, 1992; for a review, see Vernon, 1990). RT means and effect sizes have also been used to assess individual differences within many different areas, including social psychology (e.g., Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005; Nosek & Smyth, 2007; Richeson & Shelton, 2003; Wiers, Van Woerden, Smulders, & De Jong, 2002), personality psychology (e.g., Indermühle, Troche, & Rammsayer, 2011; Karwoski & Schachter, 1948; Smulders & Meijer, 2008), aging research (e.g., Eckert, Keren, Roberts, Calhoun, & Harris, 2010; Myerson, Robertson, & Hale, 2007; Wood, Willmes, Nuerk, & Fischer, 2008), and assessment of brain damage and psychopathology (e.g., Godefroy, Lhullier, & Rousseaux, 1994; Nettlebeck, 1980; Stuss, Pogue, Buckle, & Bondar, 1994).

Second, even where individual differences are not the focus of investigation, correlations between different RT effects may intuitively be thought to provide important information about the underlying mechanisms responsible for those effects. For example, Corballis (2002) examined the correlation between two RT-based effects thought to depend on the time needed for interhemispheric transmission (i.e., redundancy gain and the crossed–uncrossed difference) in order to find out whether the two effects were mediated by a common neural pathway. Likewise, Maloney, Risko, Preston, Ansari and Fugelsang (2010) examined correlations of numerical distance effects obtained with different number formats in order to see whether these different effects assessed the same underlying numerical representations and comparison mechanisms. Similarly, Stolz, Besner and Carr (2005; see also Waechter, Stolz, & Besner, 2010) examined the intercorrelations of se-mantic priming effects across different RT sessions in order to assess the extent to which individual variation in these effects reflect systematic differences in semantic associations between individuals, as opposed to merely statistical noise.

The main purpose of this article was to examine formally the factors influencing the reliabilities and correlations of RT-based measures. Although much has been written about important statistical considerations involved in correlational research using RTs (e.g., Brown, 2011; Jensen, 2006; Sriram, Greenwald, & Nosek, 2010; Stolz et al., 2005), the fundamental questions of exactly what determines RT re-liabilities and correlations and of how these quantities are related to the durations of underlying mental processes have not been addressed. In short, we ask in this article, "What aspects of mental processing times affect the reliabilities and correlations of RT-based measures?" The answers to these questions are important not only for the proper interpretation of correlations involving RT, but also for the evaluation of contemplated correlational research protocols (e.g., how power will depend on the number of trials per individual).

The main conclusion of this article is that interpretations of RT-based correlations are far more complicated than has typically been acknowledged. We analyzed several different types of RT-based correlations (e.g., mean RTs, RT effect sizes), each of which has been studied across a wide range of substantive areas. The common finding running through our analyses is that the meanings of these correlations are far more complex than intuition would suggest. Specifically, our analyses reveal that RT-based correlations are difficult to interpret because they are influenced by many factors that are not intuitively obvious. To motivate our analyses, we start with brief descriptions of four prototypical examples arising in quite different research domains. The implications of our analyses extend far beyond these prototypical exam-ples, of course, extending to all situations involving corre-lations of RT-based measures.

As a first example of the complications associated with RT-based correlations, consider the question of how strongly gen-eral intelligence correlates with a given RT effect size—the latter measured either as the difference between experimental and control conditions (e.g., Hunt, 1978) or, equivalently, as the slope relating RT to some quantitative independent variable (e.g., Jensen & Munro, 1979; but see Beauducel & Brocke, 1993). Intuitively, it seems that this correlation should be rela-tively strong to the extent that extra processing required in the experimental condition is specifically associated with general intelligence. Our analysis shows, however, that this intuition is vastly oversimplified. In fact, the correlation of the RT effect size with intelligence is also influenced by numerous other factors, including the correlation of intelligence with perfor-mance in the control condition and the correlation of perfor-mance in the experimental and control conditions.

A second and somewhat similar example involves the correlations of explicit measures of socially sensitive atti-tudes with RT-based measures of those same attitudes obtained with the implicit association test (e.g., Greenwald, McGhee, & Schwartz, 1998; Nosek & Smyth, 2007). Small correlations of explicit and implicit attitude measures have been taken as evidence that these measures tap into different underlying attitude representations or sys-tems (e.g., Greenwald et al., 1998), which seems intuitively to be quite a plausible interpretation of the small correlation. Our analysis suggests, however, that this interpretation is unwarranted and the influences of other parameters could cause the observed correlations to be quite low even if the two measures are driven by a single attitude system.

A third example involves the correlation between costs and benefits—measured relative to a common neutral con-dition—in the Stroop color-naming task (e.g., Brown, 2011). Intuitively, it seems clear that this correlation should be large if costs and benefits are determined by the same mechanism, and researchers have therefore measured this correlation to assess single-mechanism accounts of Stroop effects (e.g., Brown, 2011; Lindsay & Jacoby, 1994). Our analysis shows, however, that this intuition is wrong, because the correlation of costs and benefits can be very small or negative even when a single-mechanism account is correct. In particular, the observable correlation of costs and benefits is influenced markedly by the correlations of performance in the two experimental conditions (i.e., facilitation and interference) with that in the neutral condition. Identical complications arise, of course, when interpreting correlations of costs and benefits obtained with a variety of other paradigms (e.g., precuing; Jonides & Mack, 1984).

Finally, as a fourth example, consider the correlation between two RT effects measured under different condi-tions, such as negative-priming effects assessed with normal versus degraded visual inputs (e.g., Kane, May, Hasher, Rahhal, & Stoltzfus, 1997). Again, intuition leads one to

expect that these two effects should be strongly positively associated if they are signs of the same underlying mechanism, as would be expected. The present analysis, however, indicates that the correlation of such effects can be small even when they are driven by a common mechanism, because other factors can easily conceal the expected association between such effects. Analogous problems can even arise when two measures of the same RT effect are correlated across different testing sessions (e.g., Stolz et al., 2005). Low correlations between these two measures may simply indicate low test–retest reliability even though both measures seem necessarily to be driven by a single mechanism.

The individual differences in reaction time (IDRT) model

To explore the effects on RT-based reliabilities and correlations, we developed a general framework called the *individual differences in RT* (IDRT) model. IDRT is a specific classical test theory model, thereby allowing standard measures of classical test theory (e.g., reliabilities and correlations) to be investigated within an RT modeling framework. (Appendix 1 provides selected material on aspects of classical test theory especially relevant to this article.) Broadly, the IDRT model is intended to relate measurable RTs to underlying mental processes across a wide variety of tasks and to subsume standard RT models as special cases (e.g., diffusion models, accumulator models). It is also general enough to capture the key features of prominent models in the literature on individual differences in RT (e.g., Cerella, 1985; Fisher & Glaser, 1996; Hartley, 2001). We used the IDRT to analyze various different types of reliabilities and correlations that are computed using RT-based measures. The results of the analyses provide insights about how various parameters characterizing the times needed for mental processing would influence these quantities. Table 1 summarizes the different dependent measures whose reliabilities are considered in this article, and Table 2 summarizes the different types of correlational analyses studied—each in its own section. Readers interested in a particular type of analysis can focus on the section of interest but should first read the description of the IDRT model and the sections about the reliability of their measures (i.e., mean RT or difference score). The derivations of IDRT's predicted reliabilities and correlations are presented in Appendix 2, and the main text provides numerical illustrations of the predictions for various combinations of the model's parameters. Readers interested in other combinations of model parameters can, of course, use the general equations in Appendix 2 to examine the predicted reliabilities and correlations under any scenario of interest.

According to the general IDRT model, the total RT is the sum of latencies of several processing stages intervening between the stimulus and the response (e.g., Sternberg, 1969, 2001). The model is agnostic with respect to the nature of the processing within each stage, so this processing could conform to assumptions of diffusion models (e.g., Ratcliff, 1978), accumulator models (e.g., Usher & McClelland, 2001; Vickers, 1970), parallel models (e.g., Townsend & Nozawa, 1995), and so on. In addition, consistent with prior literature exploring individual differences in RT, IDRT proceeds from the assumption that an individual participant's average latency in an RT task depends on certain latent processing time variables that differ across participants. Thus, IDRT attempts to integrate individual differences into a general framework that can be used to study the psychometric properties of mean RTs and their differences across a variety of tasks.

In general, IDRT represents the observed mean reaction time $\mathbf{RT}_k$ of a single participant, k, with the equation[1]

$$\mathbf{RT}_k = (A + B + C) \cdot G_k + B \cdot \Delta_k + R_k + \mathbf{E}_k. \qquad (1)$$

As is described next, according to this model, the observed mean $\mathbf{RT}_k$ is determined by three conceptually separate sets of components.

The first component consists of the mental processing required in each stage of a task. In keeping with a long tradition within RT research, this component of the model characterizes task requirements rather generically as a set of three sequential stages—A, B, and C—that must be carried out between the onset of the stimulus and the initiation of the response (e.g., Donders, 1868/1969; Pashler, 1994; Smith, 1968; Sternberg, 1969). Stages A and C can be conceived as perceptual input and motor output stages, respectively, whereas stage B is a task-specific central stage such as response selection. The constants A, B, and C represent the amounts of work that need to be done in each of the three stages, and these are assumed to depend on the task but not on the person performing it.

The second component of the model represents individual differences in processing time that would allow RT to differ across people performing the same task. In keeping

---

[1] We adopted two related notational conventions. First, consistent with the notation of Lord and Novick (1968, Chap. 2), we differentiated between measurements taken for a fixed person k and measurements taken for a randomly selected person. In our notation, the subscript k is used to indicate a value for a fixed person k. When the subscript k is omitted, we refer to measurements taken for randomly selected participants. Second, unlike Lord and Novick, we denote random variables with boldface type and constants with normal type. This somewhat elaborate notation is designed to distinguish clearly between constants and random variables, as is essential for our analysis.

Table 1 Dependent measures involved in correlational analyses

Measure	Meaning
$\mathbf{Y}$	an external (non-RT) measure
$\mathbf{RT}$, $\mathbf{RT_x}$, and $\mathbf{RT_y}$	observed mean RTs in arbitrary tasks
$\mathbf{RT'}$	a second observed measure of $\mathbf{RT}$ from a separate set of trials
$\mathbf{RT_c}$ and $\mathbf{RT_e}$	observed mean RTs in control and experimental conditions
$\mathbf{D}$ (also called $\mathbf{D_{ce}}$)	an observed experimental effect $\mathbf{RT_e} - \mathbf{RT_c}$
$\mathbf{RT_u}$ and $\mathbf{RT_v}$	observed mean RTs in a second pair of control and experimental conditions
$\mathbf{D_{uv}}$	a second observed experimental effect $\mathbf{RT_v} - \mathbf{RT_u}$
$\mathbf{RT_n}$	observed mean RT in a neutral condition used to assess cost and benefit
$\mathbf{RT_f}$	observed mean RT in a facilitation condition used to assess benefit relative to a neutral condition
$\mathbf{RT_i}$	observed mean RT in an interference condition used to assess cost relative to a neutral condition
$\mathbf{D_f}$	observed RT benefit, $\mathbf{RT_n} - \mathbf{RT_f}$
$\mathbf{Di}$	observed RT cost, $\mathbf{RT_i} - \mathbf{RT_n}$

with much current thinking about the relationship of RT to intelligence (e.g., Vernon, 1990), the ability of each participant to perform cognitive tasks is modeled in terms of a general processing time parameter, G_k, which could be related to overall neural processing speed (e.g., Eysenck, 1986; Miller, 1994; Reed & Jensen, 1991; Vernon & Mori, 1992). This parameter may be thought of as the amount of time needed to carry out a single, arbitrarily defined unit of cognitive work, which we will sometimes refer to as the *processing time*. For example, participant k's total processing

Table 2 Types of correlation analyses and notation used

Type of correlation analysis	Notation
Reliability of a mean RT	$\mathrm{Corr}[\mathbf{RT}, \mathbf{RT'}]$
Correlation of a mean RT with an external measure, Y	$\mathrm{Corr}[\mathbf{RT}, \mathbf{Y}]$
Correlation of two mean RTs	$\mathrm{Corr}[\mathbf{RT_x}, \mathbf{RT_y}]$
Reliability of a difference between two mean RTs	$\mathrm{Corr}[\mathbf{D}, \mathbf{D'}]$
Correlation of a difference between two mean RTs with an external measure, Y	$\mathrm{Corr}[\mathbf{D}, \mathbf{Y}]$
Correlation of two differences between mean RTs	$\mathrm{Corr}[\mathbf{D_{ce}}, \mathbf{D_{uv}}]$
Correlation of two differences between mean RTs computed with a common term	$\mathrm{Corr}[\mathbf{D_i}, \mathbf{D_f}]$
Correlation of a difference between two mean RTs and the smaller mean	$\mathrm{Corr}[\mathbf{RT_c}, \mathbf{D}]$
Correlation of the average and difference of two mean RTs	$\mathrm{Corr}[\overline{\mathbf{RT}}, \mathbf{D}]$

Note. The meanings of the dependent measures are given in Table 1.

time in the perceptual stage (A) is $A \cdot G_k$. Analogous approaches in modeling stage processing times or speed parameters can be found in virtually all RT models (e.g., S. Brown & Heathcote, 2005; Kieras & Meyer, 1997; Miller & Ulrich, 2003; Navon & Miller, 2002; Ratcliff & Smith, 2004; Smith, 2000; Tombu & Jolicœur, 2003; Townsend & Nozawa, 1995; Usher & McClelland, 2001; for a review of older examples, see Luce, 1986).

In addition to the general processing time, G_k, the second component also includes a processing time parameter reflecting participant k's particular facility at the central processing required by a particular task, Δ_k. For example, this participant's total time for central stage (B) processing is $B \cdot (G_k + \Delta_k)$. Because of the parameter Δ_k, the size of an experimental effect on central processing time can vary across participants who have the same general processing time, G_k. Similar parameters could be added to capture individual differences in the durations of stages A and C, but in order to keep the analysis tractable, we considered only tasks varying in the requirements at the central stage (B). In developing this model, we allowed the general and task-specific processing times, G_k and Δ_k, to be correlated across participants, because a positive correlation is predicted by the idea that all cognitive operations depend to some extent on a common underlying neural processing speed (e.g., Eysenck, 1986; Vernon & Mori, 1992).

The second component also includes a residual term, R_k, which reflects individual differences in RT that are uncorrelated, or at least negligibly correlated, with the processing times represented by G_k and Δ_k. These seem most likely to involve fairly peripheral latency components that we assumed are constant across conditions within the same task. For example, R_k may include very early sensory processes starting with light transduction processes within the retina and very late motor processes ending with the activation of single muscle fibers (e.g., Ulrich & Wing, 1991). Although there is evidence that residual processes have at most a weak correlation with overall processing time or intelligence (e.g., Reed & Jensen, 1991; Vernon & Mori, 1992), we nonetheless included the correlation parameters $\rho_{\Delta R}$ and ρ_{GR} in the model development for completeness and generality.

The third component of the model is a purely statistical error term, $\mathbf{E}_k$. This term arises from the random trial-to-trial variability of RT for a given participant in a given condition. To model this within-condition variability, we assumed that the standard deviation of RT was proportional to the mean, on the basis of prior evidence that the ratio of the standard deviation of RT to the mean RT—also known as the coefficient of variation (CV)—is approximately constant in a number of RT models and data sets (e.g., Luce, 1986; Wagenmakers & Brown, 2007). For a given participant k,

then, the variance associated with the error term **E** for a mean across N trials is

$$\text{Var}[\mathbf{E}_k] = (CV \cdot \text{E}[\mathbf{RT}_k])^2/N. \tag{2}$$

This assumption implies that the mean and within-subjects standard deviation of RT would be strongly correlated across participants, as is commonly observed (e.g., Jensen, 1992). Although the predictions derived below do not depend on any distributional assumptions about $\mathbf{E}_k$, the central limit theorem implies that this error term would have a nearly normal distribution with a reasonably large number of trials, as is typical in RT research.

Following classical test theory, the IDRT model for a particular participant (i.e., Eq. 1) is generalized to a randomly selected participant by regarding the person-specific general processing time (G_k), task-specific processing time (Δ_k), and residual time (R_k) as random variables **G**, $\boldsymbol{\Delta}$, and **R**. Thus, the RT for a randomly selected participant is

$$\mathbf{RT} = (A + B + C) \cdot \mathbf{G} + B \cdot \boldsymbol{\Delta} + \mathbf{R} + \mathbf{E}. \tag{3}$$

To produce more compact expressions, we define $S = A + B + C$, so that Eq. 3 can be written as[2]

$$\mathbf{RT} = S \cdot \mathbf{G} + B \cdot \boldsymbol{\Delta} + \mathbf{R} + \mathbf{E}. \tag{4}$$

The expected value of **RT** is given by

$$\text{E}[\mathbf{RT}] = S \cdot \text{E}[\mathbf{G}] + B \cdot \text{E}[\boldsymbol{\Delta}] + \text{E}[\mathbf{R}], \tag{5}$$

because $\text{E}[\mathbf{E}] = 0$. In order to simplify the notation, we generally denote expected values of parameters with terms like μ_G and μ_Δ so that the above expectation may be written

$$\text{E}[\mathbf{RT}] = S \cdot \mu_G + B \cdot \mu_\Delta + \mu_R, \tag{6}$$

and we similarly denote variances as σ_G^2, σ_Δ^2, and so on.

The variance of **RT** is derived in Appendix 2 for the most general case (Eq. 38). As would be expected, this variance increases with task difficulty (i.e., the times needed for the perceptual, central, and motor stages [A, B, C]) and with the variability across participants of the individual processing time and sensory–motor residual parameters (i.e., **G**, $\boldsymbol{\Delta}$, and **R**), as well as the variance associated with measurement error (**E**). It also increases with the correlation between the general and task-specific processing times, **G** and $\boldsymbol{\Delta}$.

For further details concerning standard simplifying assumptions and plausible estimates of parameter values that we used in exploring the predictions of this model, the interested reader can consult Appendix 3. In addition, a

preliminary check on the overall plausibility of the model involves its ability to produce realistic Brinley plots, which is examined in Appendix 4.

As was mentioned earlier, IDRT rests on rather general assumptions and subsumes many specific RT models that provide detailed descriptions of the latency mechanisms contributing to individual stages. It should be emphasized that these specific models describe total RT as the sum of a particular modeled stage and a residual component that is beyond the scope of the model (e.g., Jepma, Wagenmakers, & Nieuwenhuis, 2012; Ratcliff, 1978). For example, the information accumulation stage of a diffusion model would correspond to IDRT's central stage B, and that model's extra parameter for encoding and motor time would correspond to the sum of IDRT's times for stages A and C. Similarly, almost all models of detection (e.g., Ollman, 1973; Smith, 1995), discrimination (e.g., Usher & McClelland, 2001), visual and memory search (e.g., Shiffrin & Schneider, 1977), and other cognitive processes postulate a sum of times for the key modeled process and the unmodeled residual processes (Luce, 1986), and this sum corresponds directly to IDRT's additive stage times. Furthermore, many models that do not describe RT as a sum (e.g., McClelland, 1979) can be approximated well by an additive model (e.g., Miller, Van der Ham, & Sanders, 1995; Molenaar & Van der Molen, 1986). In further work it would also be possible to consider predicted reliabilities and correlations within more elaborate RT models that are not approximately additive, but it seems self-evident that such models would yield relationships even more complicated than the ones emerging from our simple model. In short, the complexity of interpreting correlations that emerges from the present analysis of our simple RT model, IDRT, is likely to provide a lower bound on such complexities within the space of all existing RT models. As was noted by Hillis (1993), "when a system is too complex to understand, it often helps to understand a simpler system with analogous behavior" (p. 80).

Mean reaction times

The most basic RT measure that might be used in correlational research is the mean RT, and this measure has been used by researchers in numerous fields, including intelligence (e.g., Jensen, 1985) and neurological assessment (e.g., Godefroy et al., 1994; Stuss et al., 1994). To interpret the correlations obtained in such studies, it is essential to have a picture of the psychometric properties of mean RTs. In this section, we investigated the reliabilities and correlations of mean RTs using the IDRT model; RT-based difference scores are considered in the next section.

---

[2] Equation 4 is an elaborated version of the basic model of classical test theory shown in Appendix 1 (Eq. 20). Specifically, **RT** is the observable measure comparable to **X**, and it is the sum of a true score, **T** = $S \cdot \mathbf{G} + B \cdot \boldsymbol{\Delta} + \mathbf{R}$, plus a measurement error, **E**.

Reliability of mean reaction times

We began by studying the reliability of mean RTs, because the reliability of any measure limits the correlations it might have (see Eq. 25). As in classical test theory, the reliability of a mean RT is the correlation across individuals of two parallel measures, $\text{Corr}[\mathbf{RT}, \mathbf{RT}']$ (Lord & Novick, 1968, p. 61).

Figure 1 shows the reliability of mean RTs under various parameter combinations chosen to reveal the effects of task properties on reliability. Of course, the exact choices of parameter values were necessarily somewhat arbitrary, but we attempted to vary each parameter over its widest possible range of plausible values in order to see how large its effects might be.

Not surprisingly, the reliability of mean RTs is quite strongly affected by the number of trials, N. With these parameter combinations, reliability is often less than .5 for small numbers of trials but generally exceeds .85 with even 10–20 trials. By 100 trials—a range that is sometimes attainable in practice—reliability virtually always exceeds .95. Thus, these results

provide some reassurance to researchers planning correlational studies using mean RT: High reliability can be obtained with only moderate numbers of trials.

As is also illustrated by Fig. 1, the reliability of mean RTs is strongly influenced by the RT distribution's coefficient of variation, CV. The trial-to-trial error variance σ_E^2 increases with this CV, so reliability decreases as the CV increases. Naturally, researchers should take steps to minimize trial-to-trial fluctuations in arousal, attention, and other factors that might increase RT variability. The effect of the coefficient of variation, CV, diminishes rapidly as the number of trials increases, however, so such steps would not be very important with more than approximately 20–30 trials per participant.

Perhaps surprisingly, Fig. 1 also shows that task difficulty has only a rather small effect on RT reliability. This may at first be surprising, because it is well known that RTs are more variable in slower tasks, which would tend to increase error variance and thereby reduce reliability. On the other hand, increasing task difficulty also tends to increase the true score variance of the individual participants' mean RTs and, hence, to increase the covariance across two independent measures

Fig. 1 Reliability of mean reaction time, $\text{Corr}[\mathbf{RT}, \mathbf{RT}']$, as a function of the number of trials (N) and the task-related variables A, B, and the coefficient of variation (CV). Default values of the other parameters are shown in Table 15. $\text{Corr}[\mathbf{RT}, \mathbf{RT}']$ was computed using Eq. 24 from the covariance given in Eq. 40 and from the variances given in Eqs. 38 and 39
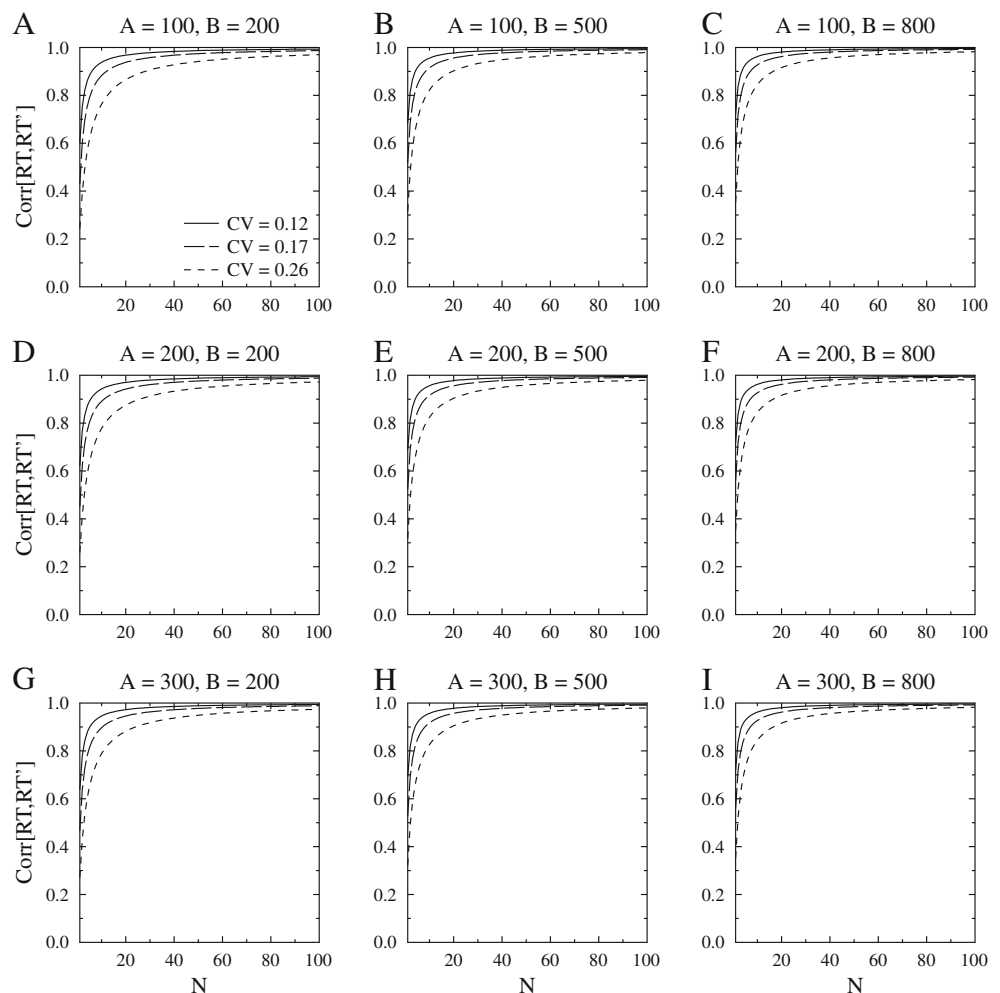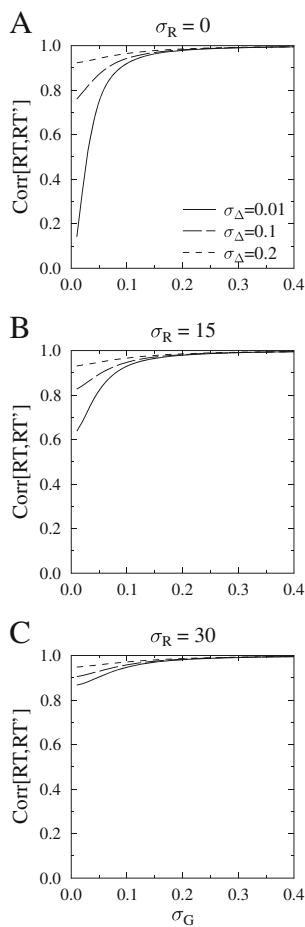
(Eq. 40). The results shown in Fig. 1 suggest that these two counteracting forces balance each other out to a good approximation, leaving reliability fairly independent of task difficulty.

Figure 2 illustrates how the reliability of mean RTs depends on the population variability in the individual cognitive processing and sensory–motor residual time parameters, **G**, **Δ**, and **R**. Somewhat arbitrarily, these levels of population variability were chosen to range from very small values—simulating homogeneous populations—to values large enough to yield visible effects on reliability. Reliability increases with variability in all three of these parameters, consistent with the well-known phenomenon that the reliability of any measure depends not only on the measuring instrument itself, but also on the population to which it is applied (e.g., Graham, 2006). In particular, the general rule within classical test theory is that reliability increases with the amount of true score variance (Eq. 24; Lord & Novick, 1968). The most important new message of this figure is that the reliability of
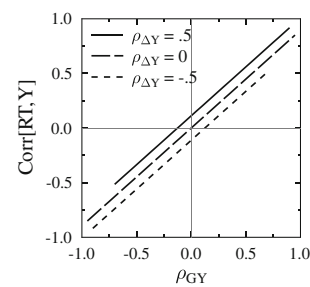
a mean RT tends to be high as long as there is variability in *at least one* of these three population parameters. This message is both good news and bad news for researchers studying correlations of mean RTs with other variables. It is good news because mean RTs can be expected to be highly reliable except in the worst case where there is severe range restriction on all three variables simultaneously (i.e., **G**, **Δ**, and **R**). For example, even in a sample that is restricted with respect to the cognitive processing times **G** and **Δ** (e.g., university students), the reliability of mean RTs will be high as long as there is adequate variability in the residual peripheral sensory and motor processing time, **R**. At the same time, this message is bad news because good reliability per se does not imply adequate sample variation in any particular type of processing time. If mean RTs are highly reliable only because of variation in **R**, for example, they might fail to correlate with cognitive measures (e.g., IQ) because of range restriction on the critical cognitive processing time parameters reflecting general and task-specific abilities, **G** and **Δ**.

Correlation of mean reaction times with another measure

Researchers often correlate the mean RT in one task with some external (i.e., non-RT) measure **Y** (e.g., IQ, total score on a symptom checklist). Figure 3 illustrates how Corr[**RT**, **Y**] is determined by the correlations of **Y** with the underlying cognitive processing time parameters determining mean RTs (i.e., **G** and **Δ**), because researchers computing such correlations seem to be interested mainly in the correlations of **Y** with these underlying processing times. Clearly, the main determinant of Corr[**RT**,**Y**] is ρ_{GY}. The relationship between these two correlations is remarkably linear, and they are nearly equal in most cases. There is also an effect of $\rho_{\Delta Y}$ on Corr[**RT**,**Y**], however, and Corr[**RT**,**Y**] best matches ρ_{GY} and $\rho_{\Delta Y}$ when the latter two correlations are equal to each other.

Fig. 3 Correlation of mean reaction time, **RT**, with an external measure, **Y**, as a function of the underlying correlations ρ_{GY} and $\rho_{\Delta Y}$. Default values of the other parameters are shown in Table 15. The lines do not cover the full left–right range from $\rho_{GY} = -1$ to $\rho_{GY} = 1$ values because some combinations of ρ_{GY} and $\rho_{\Delta Y}$ are incompatible with the assumption of $\rho_{\Delta G} = .2$. Corr[**RT**, **Y**] was computed using Eq. 23 from the covariance given in Eq. 41 and from the variances given in Eqs. 38 and 39

Fig. 2 Reliability of mean reaction time, Corr[**RT**, **RT**′], as a function of the population-related variables σ_G, σ_Δ, and σ_R. Default values of the other parameters are shown in Table 15. Correlations were computed using the same equations as those indicated in Fig. 1

Figure 3 has two important implications for researchers trying to correlate external measures with mean RTs. First, researchers interested in correlating some such measure (e.g., IQ) with general processing time (**G**) must be alert to the fact that their observed correlations will inevitably also be influenced to some extent by the task-specific processing time (**Δ**) of whatever task they choose to use. The general solution to this problem is to use a variety of tasks and extract **G** as a common factor across all of them, correlating factor scores with the external measure (Jensen, 1993). Second and conversely, researchers trying to study the relationship between an external measure and some task-specific processing time (i.e., to correlate **Y** with some particular **Δ**) must be alert to the fact that their observed correlations will be strongly influenced by the correlation of **Y** with the general processing time **G**. For example, if **G** and **Y** are uncorrelated, the correlation of **RT** and **Y** will be less than the correlation between **Δ** and **Y**, due to the diluting influence of extraneous variability contributed by **G**. In fact, the results in Fig. 3 suggest that the correlation of the external measure with general processing time, ρ_{GY}, almost completely dominates the observed correlation, Corr[**RT**, **Y**], making it almost impossible to assess the external measure's correlation with the task-specific processing time, $\rho_{\Delta Y}$. Similar influences could also arise from the residual component **R** if it were correlated with **Y** and varied substantially across individuals.

Correlation of two mean reaction times

In some situations, researchers might want to examine the correlation across individuals between the mean RTs from two different tasks, **RT$_x$** and **RT$_y$**, possibly to assess the extent to which these tasks tap into the same versus different mental processes or to validate the equivalence of different RT-based measures of individual differences (e.g., Chen, Myerson, Hale, & Simon, 2000; Kauranen & Vanharanta, 2001; Seashore & Seashore, 1941; Simonen, Videman, Battie, & Gibbons, 1995). Since any general influence on processing time would be the same in both tasks by definition, the between-task correlation would presumably vary mainly with the correlation of the task-specific processing times, $\rho_{\Delta_x \Delta_y}$. Under IDRT, the observed mean RTs for randomly selected individuals in the two tasks (i.e., $i = x, y$) are

$$\mathbf{RT}_i = S_i \cdot \mathbf{G} + B_i \cdot \mathbf{\Delta}_i + \mathbf{R}_i + \mathbf{E}_i. \tag{7}$$

Example results shown in Fig. 4 clearly indicate that the correlation between the two RTs is not a good index of the correlation between the two task-specific processing times, $\rho_{\Delta_x \Delta_y}$. In particular, Corr[**RT$_x$**, **RT$_y$**] is virtually always much higher than $\rho_{\Delta_x \Delta_y}$. This overestimation increases with σ_G, so it is evidently driven by the common contribution of the general processing time, **G**, to both tasks. The overestimation does

lessen somewhat with increases in the variability of the task-specific processing times, σ_{Δ_x} and σ_{Δ_y}, because increases in their variability allow the task-specific times to determine more of the variance in the corresponding RTs. The inadequacy of Corr[**RT$_x$**, **RT$_y$**] as a measure of the correlation of task-specific times, $\rho_{\Delta_x \Delta_y}$, is also obvious because of the shallow slopes of the lines in Fig. 4, indicating that large changes in $\rho_{\Delta_x \Delta_y}$ produce much smaller changes in Corr[**RT$_x$**, **RT$_y$**]. Thus, without detailed information about other parameters, the correlation between two RTs provides virtually no information about the correlation of the underlying task-specific processing times, $\rho_{\Delta_x \Delta_y}$.

Reaction time difference scores

Often, researchers measure an individual's performance using a difference between two mean RTs rather than a single overall mean. The difference score is generally used in order to focus more specifically on a particular kind of mental processing. For example, differences in mean memory scanning RTs to small versus large memory set sizes have been used to assess an individual's speed of retrieval from short-term memory (e.g., Keating & Bobbitt, 1978; Neubauer et al., 1997). Similarly, differences in visual search RTs with displays of different sizes have been used to index perceptual inspection and comparison time (e.g., Schweizer, 1989). Other RT differences have been used to assess interhemispheric communication (e.g., Corballis, 2002; Iacoboni & Zaidel, 2000; Schulte, Pfefferbaum, & Sullivan, 2004), semantic memory access (Hunt, 1978), executive function (e.g., Larson & Clayson, 2011), and multisensory integration (e.g., Barutchu et al., 2011). Within social psychology, the implicit association test (IAT) uses differences in the mean RTs of different stimulus–response mapping conditions to assess an individual's implicit attitudes about socially sensitive matters such as racial bias (e.g., Greenwald et al., 1998), although there is considerable debate about the meaning of the differences obtained in this way (e.g., Blanton, Jaccard, Gonzales, & Christie, 2006).[3]

Using IDRT, it is also possible to study correlations involving differences in mean RT. For simplicity, we denote the two conditions generically as "experimental" and "control" and, hence, use the subscripts "e" and "c" to distinguish them. Also, we assume that the residual sensory–

---

[3] Some researchers have transformed RTs nonlinearly to remove skew before computing differences (e.g., Nosek, Banaji, & Greenwald, 2002), and others have suggested IAT measures that are more elaborate than pure difference scores (e.g., Greenwald, Nosek, & Banaji, 2003). The analysis of correlations obtained with such more complex methods is beyond the scope of the present investigation.

Fig. 4 Illustrations of how the correlation of two mean reaction times, $\mathbf{RT_x}$ and $\mathbf{RT_y}$, varies as a function of the correlation of the underlying task-specific processing times ($\rho_{\Delta_x \Delta_y}$), the correlation of each of these processing times with $\mathbf{G}$ ($\rho_{\Delta G}$), and the variability of the processing times σ_G and $\sigma_{\Delta_x} = \sigma_{\Delta_y} = \sigma_\Delta$, with $N = 30$ trials per participant in each task. A value of $\rho_{R_x R_y} = .40$ was assumed. Default values of the other parameters are shown in Table 15. Corr[$\mathbf{RT_x}$, $\mathbf{RT_y}$] was computed using Eq. 23 from the covariance given in Eq. 42 and from the variance given in Eq. 38 for each term
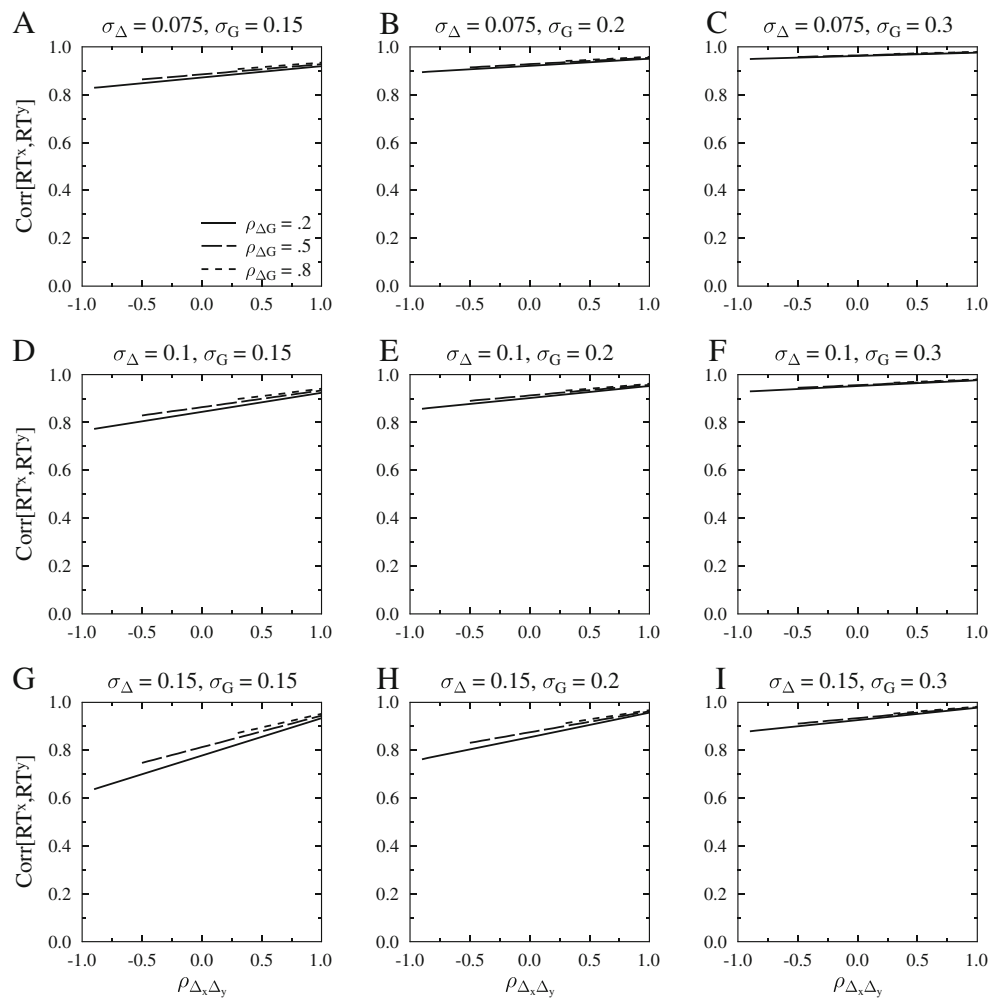
motor component is the same in all conditions of a given task, so we omit the subscript on $\mathbf{R}$. Thus, the observed mean difference score is

$$\mathbf{D} = \mathbf{RT_e} - \mathbf{RT_c}$$
$$= [S_e \cdot \mathbf{G} + B_e \cdot \mathbf{\Delta_e} + \mathbf{R} + \mathbf{E_e}] - [S_c \cdot \mathbf{G} + B_c \cdot \mathbf{\Delta_c} + \mathbf{R} + \mathbf{E_c}].$$
(8)

We assume that the experimental manipulation only influences the amount of processing needed in the central stage (B), so that $A_e = A_c$ and $C_e = C_c$, in which case this simplifies to

$$\mathbf{D} = (B_e - B_c) \cdot \mathbf{G} + B_e \cdot \mathbf{\Delta_e} - B_c \cdot \mathbf{\Delta_c} + \mathbf{E_e} - \mathbf{E_c}. \quad (9)$$

One immediate implication of this model is that the measured difference score does not completely isolate the individual ability of interest—that is, the time needed for task-specific processing in the experimental condition, $\mathbf{\Delta_e}$. Equation 9 shows that the difference also depends on both overall processing time, $\mathbf{G}$, and the task-specific processing time in the control condition, $\mathbf{\Delta_c}$.

Common versus opposing task-specific processes

It is important to distinguish between two extreme types of RT difference scores that are typically measured in RT experiments. We will refer to these as differences in which the control and experimental conditions have *common* versus *opposing* task-specific processes. The distinction is important because the analyses reported below indicate that these two types of RT differences have very different psychometric properties. As later numerical examples illustrate, for example, reliabilities tend to be higher for differences based on opposing processes than for those based on common processes. In contrast, correlations with an external measure tend to be stronger for differences based on common processes.

Comparisons involving *common* task-specific processes are those in which the control and experimental conditions differ with respect to the *amount* of some hypothesized mental processing. For example, researchers interested in mental rotation processes might compute the difference between mean RTs to stimuli rotated (say) 90° versus 180°, reasoning that the larger mean RT for the 180°

condition reflects the extra time needed for the larger rotation (e.g., Cooper & Podgorny, 1976; Just & Carpenter, 1985). It is usually assumed that each individual participant rotates at approximately the same rate in both conditions, which implies within IDRT that the values of Δ_c and Δ_e would be approximately equal for each individual. Thus, across individuals, the difference shown in Eq. 9 would have strongly positively correlated values of Δ_c and Δ_e—that is, $\rho_{\Delta_c \Delta_e} \gg 0$ —so it seems natural to refer to these experimental and control conditions as having *common* task-specific processes. As other examples of differences based on common task-specific processes, researchers might assess the speed of visual search by comparing target-detection RTs in displays with smaller versus larger numbers of items (e.g., Schweizer, 1989), and researchers might assess the speed of short-term memory search by comparing target-detection RTs in conditions with smaller versus larger numbers of items held in memory (e.g., Chiang & Atkinson, 1976; Wilson & O'Donnell, 1986). In such comparisons, the control and experimental conditions differ in the amount of the common process needed (e.g., more mental rotation in the 180° experimental condition, more memory search with more items held in memory).

In contrast, comparisons involving *opposing* task-specific processes are those in which the control and experimental conditions differ with respect to the *consequences* of some hypothesized mental processing. As one example, consider color name interference as it is often measured in the Stroop (1935) paradigm. Participants are presented with words displayed in colored letters, and they must name the color of the letters. In a congruent *control* condition, the word matches the letter color (e.g., the word "red" displayed in red letters). In an incongruent *experimental* condition, the word is a conflicting color name (e.g., the word "blue" displayed in red letters). The difference between these two conditions reflects the effects of an automatic word reading process. Critically, this effect has opposite consequences for the two conditions. Specifically, stronger automatic processing of irrelevant word name tends to speed responses in the congruent condition (i.e., to reduce Δ_c) but tends to slow them in the incongruent one (i.e., to increase Δ_e). Thus, within IDRT, the difference would involve a strong negative correlation across individuals of Δ_c and Δ_e—that is, $\rho_{\Delta_c \Delta_e} \ll 0$ —so it seems natural to refer to these experimental and control conditions as having *opposing* task-specific processes.

Exactly analogous arguments suggest that opposing task-specific processes are involved in many other tasks assessing different kinds of congruence effects, including the flanker effect, the Simon effect, the crossed–uncrossed difference in simple RT tasks (e.g., Hasbroucq, Kornblum, & Osman, 1988), the SNARC effect (e.g., Dehaene, Bossini, & Giraux, 1993), and so on (e.g., Keye, Wilhelm, Oberauer, & Van Ravenzwaaij, 2009; Larson & Clayson, 2011; McConnell &

Shore, 2011). In every case, the same processing that speeds responses in the congruent condition tends to slow responses in the incongruent one, so this processing has opposing consequences in the two conditions. Other examples include cue validity effects with spatial, semantic, and response cues (e.g., Huang, Mo, & Li, 2012; McConnell & Shore, 2011; Versace, Mazzetti, & Codispoti, 2008). The cues evoke selective preparation for a particular stimulus location, stimulus identity, or response; that preparation leads to especially fast responses in the valid cue condition, where it is appropriate, but to especially slow responses in the invalid cue condition, where it is inappropriate (i.e., a different stimulus location or identity was presented or a different response was required).

The above distinction focuses on the extremes of strongly correlated task-specific processing times (i.e., $\rho_{\Delta_c \Delta_e} \gg 0$ and $\rho_{\Delta_c \Delta_e} \ll 0$), but intermediate cases are also possible. With the mental rotation task, for example, comparing rotations of 0° versus 180° would lead to less positively correlated task-specific processing times, because rotation is involved in only one of the two conditions. In the Stroop task example, comparing the incongruent condition against a neutral condition of colored Xs would lead to less negatively correlated task-specific processing times, because automatic word reading processes would have little or no effect with the Xs. Similarly, comparisons between primed and unprimed conditions (e.g., Tipper, 1985) and between cued and uncued conditions (e.g., Fan, McCandliss, Sommer, Raz, & Posner, 2002) would tend to be intermediate, because the processes responsible for the priming and cuing effects would simply be absent from the unprimed and uncued control conditions. In the following, to investigate the psychometric properties of various types of difference scores, we used task-specific speed correlations of $\rho_{\Delta_c \Delta_e} = -.8$, $\rho_{\Delta_c \Delta_e} = 0$, and $\rho_{\Delta_c \Delta_e} = .8$ to represent difference scores based on opposing, unrelated, and common processes, respectively.

Reliability of reaction time difference scores

The reliability of a difference score is defined as the correlation of two separate estimates of that difference, $\mathrm{Corr}\left[\mathbf{D}, \mathbf{D}'\right]$. Figure 5 shows how the reliability of RT difference scores depends on a number of task-related variables that might be expected to influence these differences. In general, the number of trials, N, has a large effect, as expected. With some combinations of parameters, though, many more trials are needed to obtain reliable RT difference scores than were required to obtain reliable mean RTs. Hundreds of trials per condition are sometimes needed to produce reliabilities exceeding .8; although not shown in the figure, thousands are sometimes needed for reliabilities exceeding .9. Unless it is practical to obtain thousands of trials per participant and condition, these results raise a caution for researchers studying correlations of RT differences: It may be difficult to obtain high reliability.

Fig. 5 Reliability of a mean reaction time difference score, **D**, as a function of the number of trials (N), the duration of stage A, the duration of stage B in the experimental condition (B_e), as compared with a duration of $B_c = 200$ ms in the control condition, and the correlation of task-specific processing times in the two conditions involved in the difference score ($\rho_{\Delta_c \Delta_e}$). Default values of the other parameters are shown in Table 15. Corr$[\mathbf{D}, \mathbf{D}']$ was computed using Eq. 24 from the covariance given in Eq. 45 and from the variance given in Eq. 43
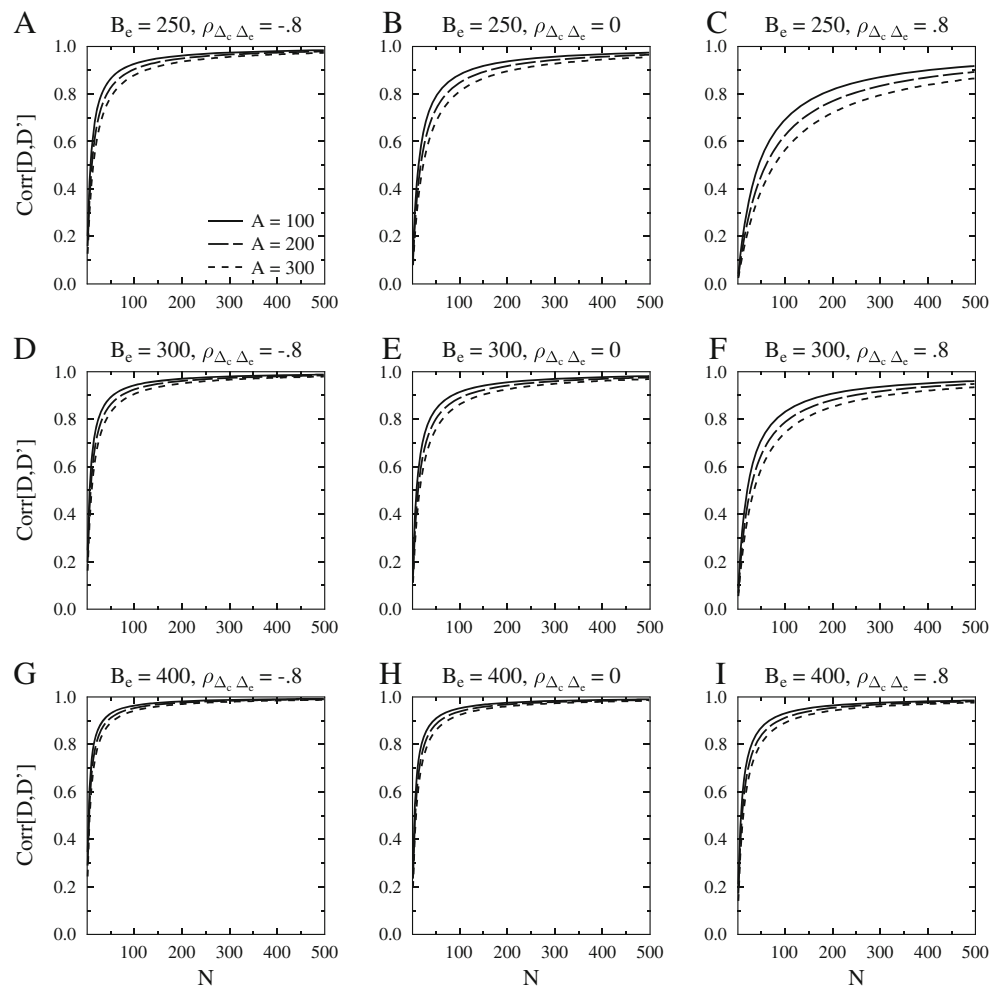

It is clear from Fig. 5 that RT difference scores computed from opposing tasks (5a, 5d, and 5g: $\rho_{\Delta_c \Delta_e} = -.8$) are more reliable than those computed from common tasks (5c, 5f, and 5i: $\rho_{\Delta_c \Delta_e} = .8$). RT differences computed from unrelated tasks (5b, 5e, and 5h: $\rho_{\Delta_c \Delta_e} = 0$) are intermediate. Furthermore, the Corr$[\mathbf{D}, \mathbf{D}']$ difference between common and opposing tasks can be quite large. For example, keeping other parameters constant, reliability could be .8 for opposing tasks but only .1 for common tasks.

As would be expected, reliability tends to increase with a larger effect size (i.e., larger values of central stage processing time B_e relative to the fixed B_c). Interestingly, increases in the duration of the perceptual stage A decrease reliability, despite the fact that these stage times are removed from the difference score by the subtraction, and the same pattern is also found with the durations of the motor stage (C). This pattern results from increased trial-to-trial error variance, which increases with the overall RT when perceptual or motor time (i.e., A or C) increases. For the same reason, reliability would decrease with increases in the coefficient of variation CV, although this is not illustrated in the figure.

Figure 6 illustrates the increases in difference score reliability resulting from increases in the variabilities of the general and task-specific processing times (σ_G and σ_Δ), with the range of parameter values again chosen to produce clear effects. In general, these effects are not too large over the range of parameter values examined here. Again, the difference between common and opposing task-specific processes plays a crucial role, with reliability decreasing dramatically as $\rho_{\Delta_c \Delta_e}$ increases from −.8 to .8.

In summary, researchers wanting to study correlations involving an RT difference score must be aware that many more trials—often an order of magnitude more—are needed to obtain adequate reliability than are needed with mean RTs. Moreover, the relationship between the two conditions entering into the difference score must also be considered, because this relationship has a big effect on the number of trials needed for adequate reliability. Indeed, thousands of trials per condition may be needed when the two conditions involve common task-specific processing, especially if the effect is not too large. Fortunately, the equations presented in Appendix 2 can be used to estimate reliability and, thereby, help determine the number of trials needed to

Fig. 6 Reliability of a mean reaction time difference score, **D**, as a function of σ_G, σ_Δ, $\rho_{\Delta_c \Delta_e}$, and the number of trials (*N*). Default values of the other parameters are shown in Table 15. Correlations were computed using the same equations as those indicated in Fig. 5
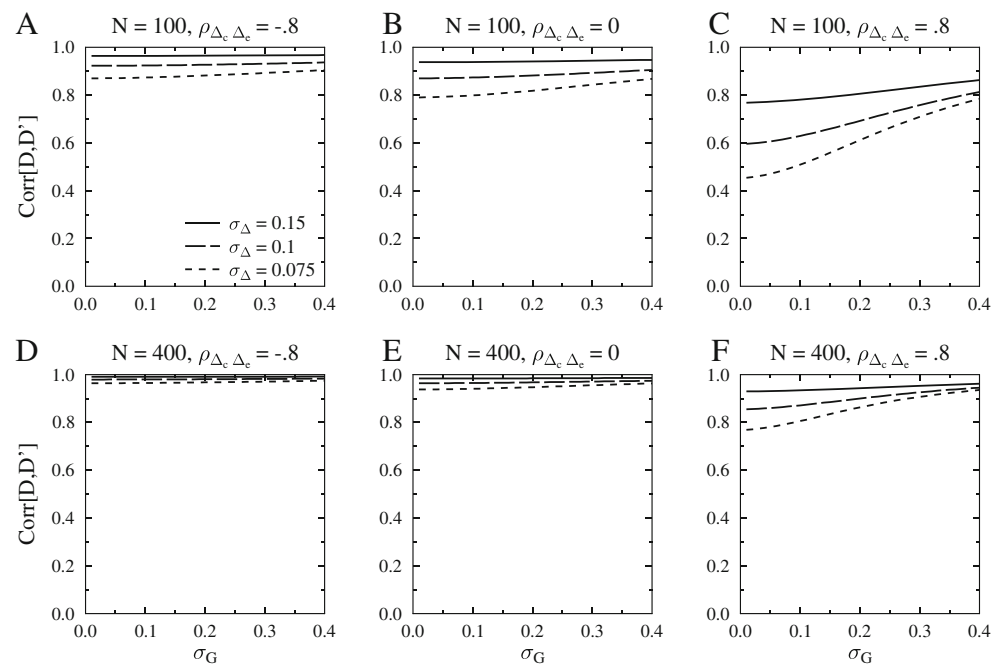
obtain a desired reliability level for a specific set of assumed parameter values that seem appropriate for the difference score under study.

In practice, unfortunately, it might be difficult to estimate the extent to which a given RT difference score involves common versus opposing processes, because this involves estimating the value of $\rho_{\Delta_c \Delta_e}$. At present, we know of no way to do that empirically and, thus, could rely only on a theoretical analysis of the tasks entering into the difference score. Although such analysis seems convincing in most of the cases that have been considered here (e.g., set size effect in memory scanning, Stroop congruency effect), it certainly need not do so in all cases.

Correlation of reaction time difference scores
with another measure

One of the most common uses of RT in correlational research is to study the relationship of an RT difference score, $\mathbf{D} = \mathbf{RT_e} - \mathbf{RT_c}$, with some external measure, **Y** (e.g., Greenwald et al., 1998; Hunt, 1978; Williams, Light, Braff, & Ramachandran, 2010). Intuitively, the RT difference score is used in order to remove unwanted influences of general processing time, **G**, and of residual sensory–motor time, **R**. The usual goal of the correlation is to assess the relationship between the external measure and the task-specific processing time in the experimental condition, $\mathbf{\Delta_e}$.

For example, the difference between RTs in primed and unprimed conditions in the negative-priming paradigm is thought to isolate the effects of inhibitory processes (e.g., Tipper, 1985), and researchers have correlated this measure of inhibitory processes with the severity of schizophrenic

symptoms in order to examine the hypothesis that inhibition is disrupted by schizophrenia (e.g., Moritz & Andresen, 2004; Moritz & Mass, 1997). Others have correlated an RT-based measure of hemispheric disconnection known as the crossed–uncrossed difference with other behavioral (e.g., Cherbuin & Brinkman, 2006) and neurophysiological (e.g., Iacoboni & Zaidel, 2004) measures of such disconnection, essentially attempting to determine the validity of the RT-based disconnection measure. Likewise, RT difference scores are sometimes used to isolate particular cognitive processes, such as memory retrieval, that are thought to be especially strongly related to standard psychometric measures of IQ (e.g., Hunt, 1978; Keating & Bobbitt, 1978). As a final example, RT difference scores are now used extensively in social psychology within the context of the IAT (Greenwald et al., 1998), as was mentioned earlier. In the IAT people must classify examples based on two different categorical distinctions (e.g., flowers vs. insects and words having pleasant vs. unpleasant meanings). There are only two possible responses (e.g., left and right hands), and across two experimental conditions, the response assignments for the two distinctions are paired in opposite ways (e.g., flowers + pleasant words vs. insects + unpleasant words in one condition, flowers + unpleasant words vs. insects + pleasant words in the other). If the two distinctions are semantically related, responses should presumably be faster in the condition with two associated categories assigned to the same response (e.g., flowers + pleasant words vs. insects + unpleasant words) than in the condition with two unassociated categories assigned to the same response (e.g., flowers + unpleasant words vs. insects + pleasant words). Thus, the RT difference between these two

conditions may be a measure of the strength of the semantic associations between categories (but see Blanton et al., 2006). This RT difference measure is thought to be implicit because it is extracted from performance measures rather than explicit questions about attitudes, and such RT difference scores are often correlated with corresponding explicit attitude measures (for a recent review and meta-analysis, see Hofmann et al., 2005). Moreover, the difference seems to be based on opposing processes, because stronger semantic associations would speed responses with associated categories and slow responses with unassociated categories.

In this section we analyze the correlation of an RT difference score, $\mathbf{D} = \mathbf{RT_e} - \mathbf{RT_c}$, with an external measure, $\mathbf{Y}$. As in the section on difference score reliability, we assume that a randomly selected participant's difference score is described by Eq. 9. Figure 7 illustrates the correlations predicted by IDRT under a wide range of combinations of true correlations among model terms. When the researcher's goal is to assess the correlation between $\mathbf{\Delta_e}$ and $\mathbf{Y}$, the perfect outcome would be $\text{Corr}[\mathbf{D}, \mathbf{Y}] = \rho_{\Delta_e Y}$, which would imply that all points in each panel of the figure would lie exactly on the positive diagonal. As can be seen in the figure, the observable correlation $\text{Corr}[\mathbf{D}, \mathbf{Y}]$ does tend to increase linearly with the correlation of $\mathbf{Y}$ with the underlying task-specific processing time, $\rho_{\Delta_e Y}$, which is good. Nonetheless, the values of $\text{Corr}[\mathbf{D}, \mathbf{Y}]$ and $\rho_{\Delta_e Y}$ are often quite different (i.e., many points are far from the diagonal), so the former is not necessarily a good estimate of the latter. Moreover, depending on the other parameters, the observable correlation $\text{Corr}[\mathbf{D}, \mathbf{Y}]$ can be either larger or smaller than the underlying correlation of interest, $\rho_{\Delta_e Y}$, so researchers cannot even be certain whether $\text{Corr}[\mathbf{D}, \mathbf{Y}]$ will tend to underestimate or overestimate the true target value of $\rho_{\Delta_e Y}$. Comparisons across panels indicate that the observable $\text{Corr}[\mathbf{D}, \mathbf{Y}]$ tends to increase with increases in the correlation of $\mathbf{Y}$ with general processing time, ρ_{GY}, and to decrease with increases in the correlation of $\mathbf{Y}$ with task-specific processing time in the control condition, $\rho_{\Delta_c Y}$.

The results shown in Fig. 7 also indicate that there is a substantial effect of whether the RT difference score is based on common, unrelated, or opposing processes (i.e., $\rho_{\Delta_c \Delta_e} = .8$, 0, or −.8). Thus, the relation between task-specific processes, $\rho_{\Delta_c \Delta_e}$, has an important effect on the observable correlation $\text{Corr}[\mathbf{D}, \mathbf{Y}]$ even though this parameter does not directly involve $\mathbf{Y}$. As can be seen within each panel, the lines relating the observable $\text{Corr}[\mathbf{D}, \mathbf{Y}]$ to the underlying $\rho_{\Delta_e Y}$ are steepest with differences based on common processes and shallowest with differences based on opposing processes. In this sense, $\text{Corr}[\mathbf{D}, \mathbf{Y}]$ may be regarded as a better indicator of $\rho_{\Delta_e Y}$ with common rather than opposing processes, although the actual numerical difference between the observable $\text{Corr}[\mathbf{D},\mathbf{Y}]$ and the target $\rho_{\Delta_e Y}$ depends on many parameters and is in many cases smaller with opposing

processes than with common ones. In the final analysis, then, an observed value of $\text{Corr}[\mathbf{D},\mathbf{Y}]$ by itself conveys little information about the correlation between $\mathbf{Y}$ and the time needed for task-specific processing in the experimental condition. For example, it seems quite risky to assess the relation between schizophrenia and inhibitory processes by correlating the extent of schizophrenic symptoms with the difference in RTs between a condition with inhibitory negative priming and a neutral control condition, because the observable correlation is influenced by too many factors to provide a good estimate of the association between schizophrenia and inhibitory processes.

Correlation of two distinct reaction time difference scores

Researchers might also want to examine the correlation across participants between two different experimental effects, perhaps in order to estimate the degree to which the effects are determined by the same versus distinct mental processes. Typically, the size of each effect is estimated for each participant by the difference in mean RTs between two conditions of a particular task, and these effect sizes are then correlated.

For example, Fan et al. (2002) developed the Attention Network Test (ANT) in order to obtain separate assessments of three previously suggested attentional networks involved in alerting, spatial orienting, and conflict resolution. Specifically, their goal was "to assess whether or not subjects' efficiency within each of the [three attentional] networks was correlated" (p. 343) as a test of whether these attentional networks "engage separate brain mechanisms" (p. 344). On each trial, participants were presented with a row of five left- or right-pointing arrows and were required to respond with the left or right hand in accordance with the direction of the target arrow in the center of the row, with the other arrows being distractors (cf. Eriksen & Eriksen, 1974). The time needed for conflict resolution was measured as the difference in mean RT between trials with congruent distractor arrows (i.e., pointing in the same direction as the center target) and trials with incongruent distractors (i.e., pointing in the opposite direction). In addition, the spatial location of the row of arrows was either cued or unpredictable, and the difference between the mean RTs in these conditions was used to assess the efficiency of spatial orienting. Finally, the onset time of the row of arrows was either cued or unpredictable, with the difference between these conditions used to assess alerting. There were no statistically reliable correlations among these three effects, leading the authors to conclude that these effects are mediated by separate processes, although subsequent analyses of the psychometric properties of these measures have called this conclusion into question (e.g., MacLeod et al., 2010). The following analysis using IDRT also suggests that correlations may be small even if there are common

Fig. 7 Correlation of a mean reaction time difference score, **D**, and an external measure, **Y**, as a function of the true correlations ρ_{GY}, $\rho_{\Delta_c Y}$, $\rho_{\Delta_c \Delta_e}$, and $\rho_{\Delta_e Y}$. Default values of the other parameters are shown in Table 15. Corr[**D**, **Y**] was computed using Eq. 47
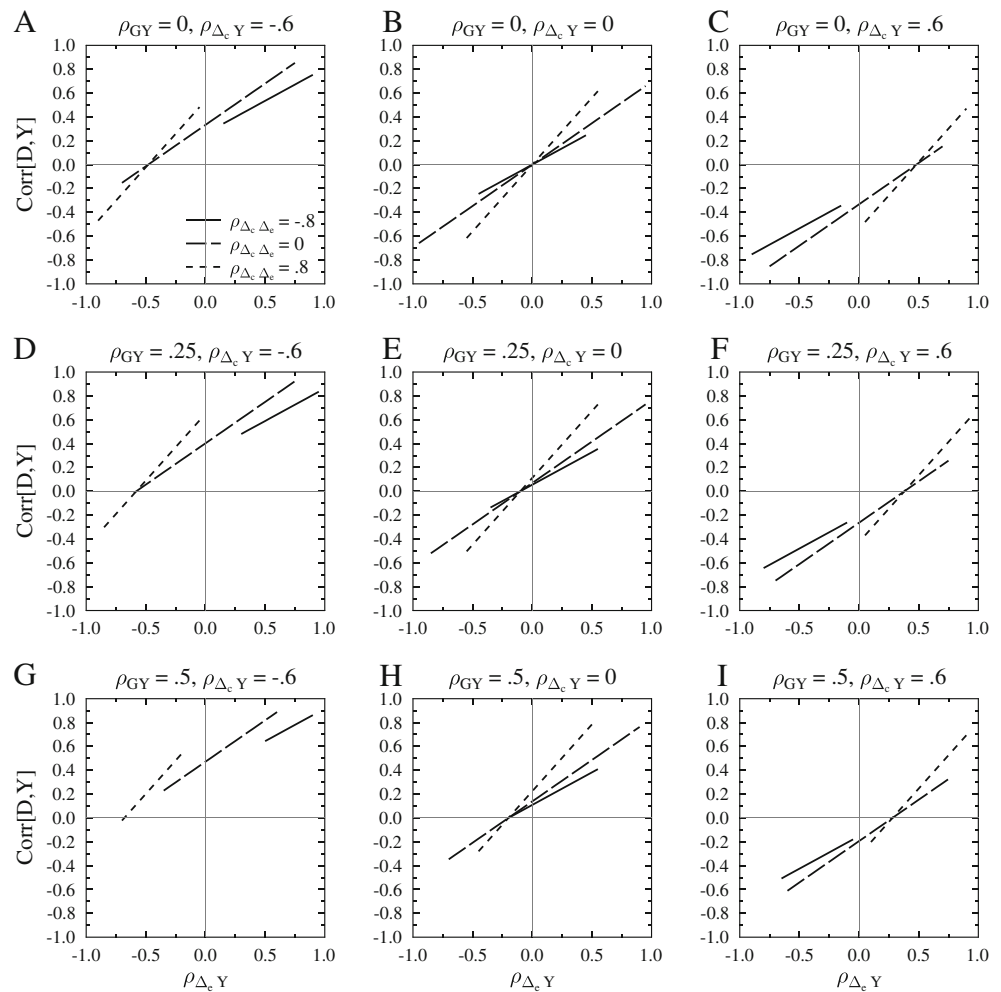
components and that Fan et al.'s conclusion is, therefore, only weakly supported by the findings. This is especially true because only approximately 25 trials per condition and participant were included in some of the difference scores, weakening reliability.

Again, we consider one effect to be a comparison between experimental and control conditions, $D_{ce} = RT_e - RT_c$, so that the difference for each individual in a particular task is given by Eq. 9. Let the second effect be denoted as $D_{uv} = RT_v - RT_u$, so that the analogous equations apply substituting u and v for c and e. As before, we assume that both experimental effects involve changes in the time needed for the central stage, B.

In most situations, researchers seem mainly interested in the correlation of the task-specific processing times of the two experimental conditions, $\rho_{\Delta_e \Delta_v}$, so one question is clearly how well that parameter is estimated by Corr[D_{ce}, D_{uv}] (e.g., Kane et al., 1997). Figure 8 illustrates how the true correlation Corr[D_{ce}, D_{uv}] varies as a function of the correlation between the task-specific processing times of the two experimental conditions, $\rho_{\Delta_e \Delta_v}$, as well as the correlations between other

pairs of task-specific processing times (e.g., $\rho_{\Delta_c \Delta_e}$, $\rho_{\Delta_u \Delta_v}$). Specifically, the different panels represent correlations in which the difference scores represent different combinations of common, unrelated, and opposing task-specific processes. For example, Fig. 8b represents a correlation between one difference score involving unrelated task-specific processes (i.e., $\rho_{\Delta_c \Delta_e} = 0$) and one difference score involving opposing processes (i.e., $\rho_{\Delta_u \Delta_v} = -.8$), which corresponds to the correlation of the flanker effect (opposing) and the spatial cuing effect (unrelated) in the study of Fan et al. (2002).

One important fact influencing the patterns shown in most panels of Fig. 8 is that the interrelationships of the different processing time parameters are tightly constrained. Because of these constraints, some values of $\rho_{\Delta_e \Delta_v}$ are impossible given specified values of the other correlations, which causes left and right truncation of the lines in most panels. For example, consider the correlation of two differences both involving opposing task-specific processes (Fig. 8a). If the task-specific processing times of the two control conditions are uncorrelated (i.e., $\rho_{\Delta_c \Delta_u} = 0$) and all of the task-specific processing times have the same small correlation with **G**

Fig. 8 Correlation of two mean reaction time difference scores, $\mathbf{D_{ce}}$ and $\mathbf{D_{uv}}$, as a function of the correlations between task-specific processing times $\rho_{\Delta_c \Delta_e}$, $\rho_{\Delta_u \Delta_v}$, $\rho_{\Delta_c \Delta_u}$, and $\rho_{\Delta_e \Delta_v}$. Correlations between two differences involving two opposing, two unrelated, and two common task-specific processes, are depicted in **a**, **c**, and **f**, respectively, whereas **b**, **d**, and **e** depict correlations of two differences involving other combinations of task-specific differences. Values of $\rho_{\Delta_c \Delta_v} = \rho_{\Delta_e \Delta_u} = 0$ were assumed. Default values of the other parameters are shown in Table 15. Corr[$\mathbf{D_{ce}}$, $\mathbf{D_{uv}}$] was computed using Eq. 23 from the covariance given in Eq. 48 and from the variance given in Eq. 43 for each term
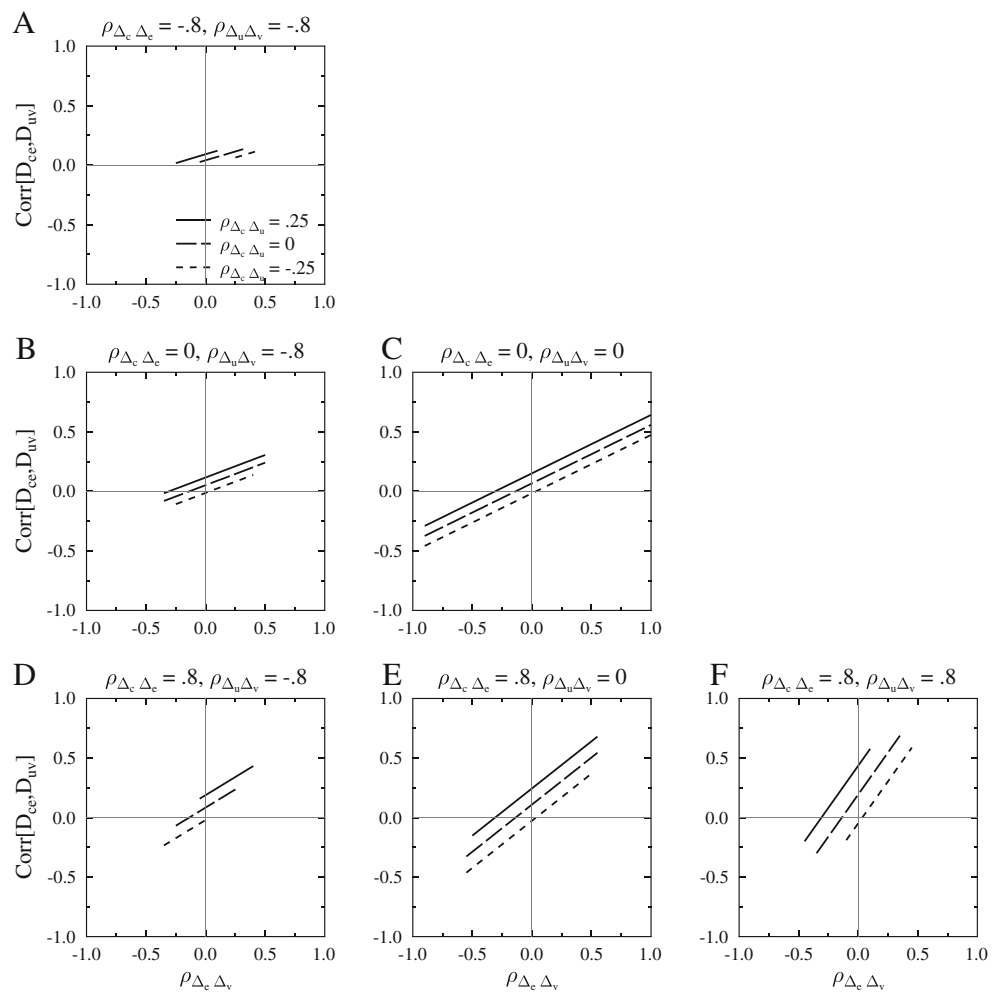
(i.e., $\rho_{\Delta_c G} = \rho_{\Delta_e G} = \rho_{\Delta_u G} = \rho_{\Delta_v G} = .2$), then the task-specific processing times of the two experimental conditions cannot possibly have a strong positive or negative correlation. Instead, their correlation must lie within the range of approximately −.1 to +.3. Thus, researchers interested in estimating such correlations (e.g., the correlation of the Stroop effect with the flanker effect) should keep in mind that the correlation under investigation is part of a complex network of interrelated quantities—not an entirely free parameter like mean effect size.

Each panel of Fig. 8 reveals a highly linear relation between the observable correlation of difference scores, Corr[$\mathbf{D_{ce}}$, $\mathbf{D_{uv}}$], and the underlying correlation of the task-specific processing times in the two experimental conditions, $\rho_{\Delta_e \Delta_v}$, suggesting that the former could provide a good estimate of the latter within the narrow range of possible values. Unfortunately, the observable and target values (i.e., Corr[$\mathbf{D_{ce}}$, $\mathbf{D_{uv}}$] and $\rho_{\Delta_e \Delta_v}$, respectively) are not generally equal, raising complications for such estimates. For example, consider first the correlation of two differences both involving opposing task-specific processes

(Fig. 8a). The observable values of Corr[$\mathbf{D_{ce}}$, $\mathbf{D_{uv}}$] are systematically closer to zero than the underlying target values of $\rho_{\Delta_e \Delta_v}$, making it less likely that researchers will find a statistically significant correlation even when the two task-specific processes are truly correlated (i.e., $\rho_{\Delta_e \Delta_v} \neq 0$). For example, Fan et al. (2002) might not have obtained correlations among attentional effects even if the same mechanisms influenced task-specific processes in the experimental conditions, depending upon the relationships among the task-specific processes in the control conditions.

In addition, the separate lines within Fig. 8a illustrate that the observable Corr[$\mathbf{D_{ce}}$, $\mathbf{D_{uv}}$] increases with the correlation between task-specific processing times in the two control conditions, $\rho_{\Delta_c \Delta_u}$. This effect indicates that the observable Corr[$\mathbf{D_{ce}}$, $\mathbf{D_{uv}}$] is not a pure measure of the target correlation $\rho_{\Delta_e \Delta_v}$. Moreover, it shows that the observable Corr[$\mathbf{D_{ce}}$, $\mathbf{D_{uv}}$] can differ from zero even when the target correlation $\rho_{\Delta_e \Delta_v}$ equals zero, leading to the possibility that researchers will incorrectly reject the null hypothesis of interest (i.e., $H_0 : \rho_{\Delta_e \Delta_v} = 0$). For example, researchers might incorrectly

conclude that two experimental conditions are correlated when, in fact, only the two control conditions are related.

Similar problems for the estimation of the underlying $\rho_{\Delta_e \Delta_v}$ by the observable Corr[$\mathbf{D_{ce}}$, $\mathbf{D_{uv}}$] are also evident in Fig. 8b–f. Specifically, the lines do not generally lie on the positive diagonal, making it difficult for researchers to estimate the underlying correlation of interest, $\rho_{\Delta_e \Delta_v}$, from the observable value of Corr[$\mathbf{D_{ce}}$, $\mathbf{D_{uv}}$]. For example, even across the relatively restricted range of ± .25 examined here, the nuisance parameter $\rho_{\Delta_c \Delta_u}$ always has a clear effect on Corr[$\mathbf{D_{ce}}$, $\mathbf{D_{uv}}$], demonstrating that the observable value depends noticeably on factors other than the underlying correlation of interest.

Correlation of two reaction time difference scores involving the same baseline

A prominent variant of the RT difference score analysis considered in the previous section arises when researchers assess two effects on RT by computing difference scores that involve the same baseline condition. One case in which researchers examine such RT differences, on which we focus here, involves what is sometimes called *cost–benefit analysis* (e.g., Jonides & Mack, 1984). In this type of analysis, the researcher obtains the mean RTs for three conditions. Condition n is a neutral baseline condition, whereas conditions f and i represent conditions with some type of facilitation and interference, respectively. In the Stroop (1935) color naming task, for example, facilitation is expected when the irrelevant word matches the to-be-named color and inhibition is expected when the irrelevant word names an alternative color, relative to a neutral condition in which the word has no color association (e.g., Brown, 2011).[4]

In cost–benefit analysis, the measured cost of interference is $\mathbf{D_i} = \mathbf{RT_i} - \mathbf{RT_n}$, the measured benefit of facilitation is $\mathbf{D_f} = \mathbf{RT_n} - \mathbf{RT_f}$, and the correlation between facilitation

---

[4] A slightly different case in which the same baseline enters into two differences is when two different experimental effects are measured relative to the same control condition. That is, the researcher would obtain mean RTs in a control condition c and in two experimental conditions e and v. The two experimental effects would then be measured as $\mathbf{D_{ce}} = \mathbf{RT_e} - \mathbf{RT_c}$ and $\mathbf{D_{cv}} = \mathbf{RT_v} - RT_c$. For example, a human-factors researcher might measure separate effects of alcohol and sleep loss on RTs in a simulated driving task by comparing RTs obtained with each of these manipulations against RTs in a normal control condition. This situation differs from the cost–benefit analysis in that the baseline condition enters into both differences with the same sign. This implies that a positive correlation between $\mathbf{RT_e}$ and $\mathbf{RT_v}$, which would result from a positive correlation between $\mathbf{\Delta_e}$ and $\mathbf{\Delta_v}$, would tend to produce a positive correlation between the RT differences, Corr[$\mathbf{D_{ce}}$, $\mathbf{D_{cv}}$]. Thus, the correlation in this case is simply the negative of the correlation for the cost–benefit case in which the baseline condition enters into the two differences with opposite signs. Because of the simple relationship between the two cases, we considered only one of them, and we chose the cost–benefit case because it is more frequent in the literature.

and interference is often of theoretical interest. Note that the neutral condition mean $\mathbf{RT_n}$ enters into the two differences with opposite signs, which tends to produce a negative correlation between these two differences (e.g., Brown, 2011).

Considering as usual the case in which only the central stage B varies across conditions, the RTs for the interference and neutral conditions involved in the cost–benefit analysis can be represented under IDRT as

$$\mathbf{RT_i} = (A + B_i + C) \cdot \mathbf{G} + B_i \cdot \mathbf{\Delta_i} + \mathbf{R} + \mathbf{E_i}, \text{ and} \quad (10)$$

$$\mathbf{RT_n} = (A + B_n + C) \cdot \mathbf{G} + B_n \cdot \mathbf{\Delta_n} + \mathbf{R} + \mathbf{E_n}. \quad (11)$$

It is most intuitive, however, to represent the mean RT in the facilitation condition as

$$\mathbf{RT_f} = (A + B_f + C) \cdot \mathbf{G} - B_f \cdot \mathbf{\Delta_f} + \mathbf{R} + \mathbf{E_f}, \quad (12)$$

with $B_f \cdot \mathbf{\Delta_f}$ subtracted from rather than added to the overall total $\mathbf{RT_f}$. With this definition of $\mathbf{RT_f}$, larger values of $\mathbf{\Delta_f}$ produce smaller values of $\mathbf{RT_f}$, so $\mathbf{\Delta_f}$ reflects the amount of facilitation, as one intuitively expects. Furthermore, with this definition a positive correlation of $\mathbf{\Delta_f}$ and $\mathbf{\Delta_i}$ means that larger facilitation is associated with larger inhibition, as one also intuitively expects. Using these definitions, the measured cost of interference, $\mathbf{D_i} = \mathbf{RT_i} - \mathbf{RT_n}$, and benefit of facilitation, $\mathbf{D_f} = \mathbf{RT_n} - \mathbf{RT_f}$, are

$$\mathbf{D_i} = (B_i - B_n) \cdot \mathbf{G} + B_i \cdot \mathbf{\Delta_i} - B_n \cdot \mathbf{\Delta_n} + \mathbf{E_i} - \mathbf{E_n}$$
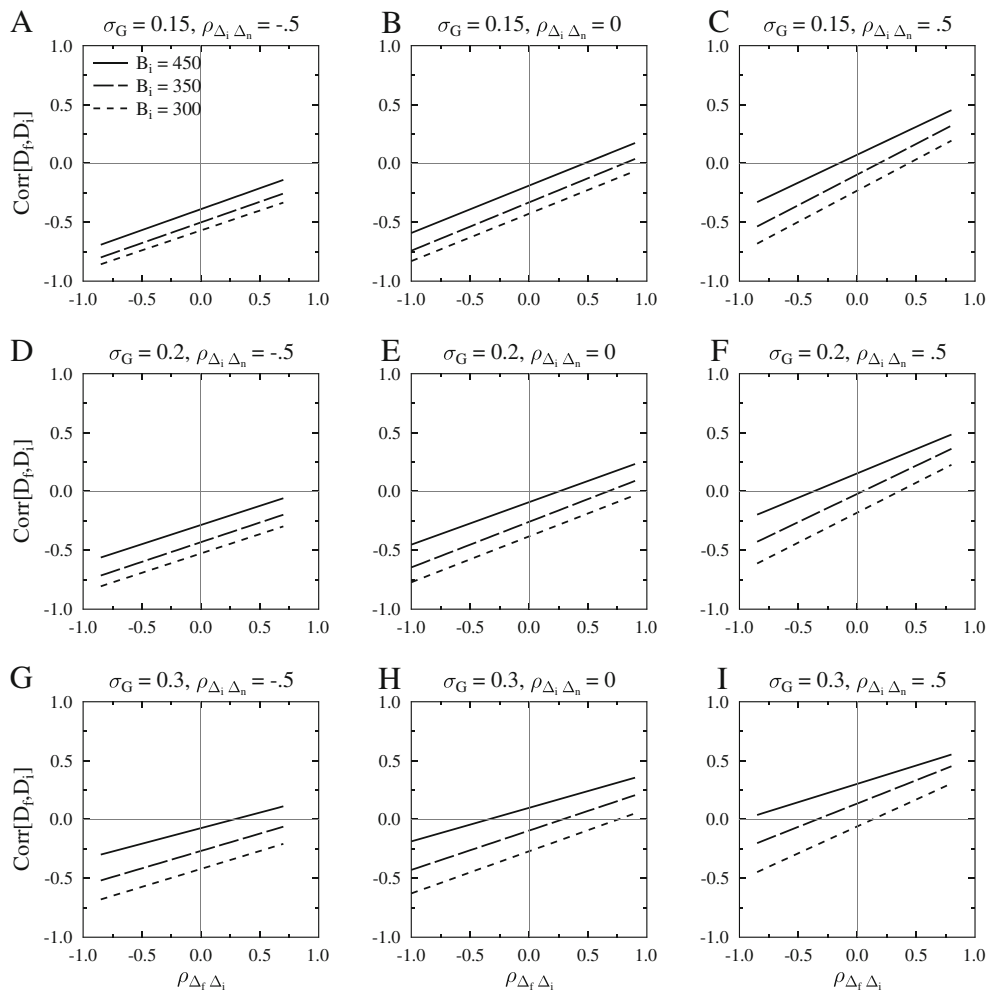$$(13)$$

and

$$\mathbf{D_f} = (B_n - B_f) \cdot \mathbf{G} + B_n \cdot \mathbf{\Delta_n} + B_f \cdot \mathbf{\Delta_f} + \mathbf{E_n} - \mathbf{E_f}. \quad (14)$$

Figure 9 shows the correlation between observable costs and benefits, Corr[$\mathbf{D_f}$, $\mathbf{D_i}$], as a function of the correlation of the underlying task-specific processing times in the facilitation and inhibition conditions, $\rho_{\Delta_f \Delta_i}$, as well the overall effect size indexed by central processing time in the interference condition, B_i, the variability associated with the general processing time $\mathbf{G}$, and the correlation of task-specific processing times in the inhibition and neutral conditions, $\rho_{\Delta_i \Delta_n}$. The figure reveals that the observable Corr[$\mathbf{D_f}$, $\mathbf{D_i}$] is linearly related to the underlying correlation of task-specific costs and benefits that is of interest, $\rho_{\Delta_f \Delta_i}$, but is not generally equal to it, even approximately. Indeed, Corr[$\mathbf{D_f}$, $\mathbf{D_i}$] and $\rho_{\Delta_f \Delta_i}$ often differ in sign, so the observed value need not even capture the true direction of the correlation of interest. Furthermore, when the value of the underlying target correlation $\rho_{\Delta_f \Delta_i}$ approaches the extremes of ±1, the observable Corr[$\mathbf{D_f}$, $\mathbf{D_i}$] is far too small in absolute value. Both within and across panels, it is

Fig. 9 Correlation of mean reaction time difference scores measuring the benefit of facilitation, $\mathbf{D_f} = \mathbf{RT_n} - \mathbf{RT_f}$, and the cost of interference, $\mathbf{D_i} = \mathbf{RT_i} - \mathbf{RT_n}$, computed using a common neutral condition, $\mathbf{RT_n}$. Correlations are displayed as a function of σ_G; the amount of stage B processing needed in the inhibition condition, B_i, which determines the mean cost relative to the fixed stage B processing amounts in the facilitation and neutral conditions (i.e., $B_f = 200$ ms and $B_n = 250$ ms); the correlation of the task-specific processing times in the neutral and inhibition conditions, $\rho_{\Delta_i \Delta_n}$; and the correlation of the task-specific processing times in the facilitation and inhibition conditions, $\rho_{\Delta_f \Delta_i}$. The value of $\rho_{\Delta_f G} = -.2$ was assumed—in contrast to the values of $\rho_{\Delta_i G} = \rho_{\Delta_n G} = .2$ —because the task-specific processing time was subtracted from the total RT in the benefit condition (Eq. 12), rather than adding to it as in the cost and neutral conditions (Eqs. 10 and 12). Default values of the other parameters are shown in Table 15. Corr[$\mathbf{D_f}$, $\mathbf{D_i}$] was computed using Eq. 23 from the covariance given in Eq. 49. The variance of $\mathbf{D_i}$ is given by Eq. 43, with the neutral and inhibition conditions corresponding to the control and experimental conditions, respectively, and the variance of $\mathbf{D_f}$ is given by Eq. 44

also clear that the other parameters have substantial effects on the observable Corr[$\mathbf{D_f}$, $\mathbf{D_i}$] (e.g., the population variability of general processing time, σ_G), making it impossible to estimate the target $\rho_{\Delta_f \Delta_i}$ from the observable Corr[$\mathbf{D_f}$, $\mathbf{D_i}$] without precise information about the values of these other parameters. It is especially noteworthy that the observable Corr[$\mathbf{D_f}$, $\mathbf{D_i}$] depends on the correlations of task-specific processing time in the neutral condition with the time in the inhibition condition ($\rho_{\Delta_i \Delta_n}$, illustrated across panels) and with the time in the facilitation condition ($\rho_{\Delta_f \Delta_n}$, not illustrated). Thus, researchers studying the relationship between facilitation and interference must allow for effects of these auxiliary neutral-condition correlations on the observable Corr[$\mathbf{D_f}$, $\mathbf{D_i}$]. Debates

about the appropriateness of different neutral conditions have previously focused on mean RT (e.g., Brown, 2011), but this analysis shows that the correlation of this condition with the facilitation and inhibition conditions must also be considered when effect-size correlations are being investigated.

One particularly interesting special case of a correlation involving difference scores with a common neutral term arises in the analysis of Stroop effects (e.g., MacLeod, 1991). As was reviewed by Brown (2011), many models of the Stroop task posit that facilitation in the congruent condition and interference in the incongruent condition are driven by a single underlying mechanism for automatic word recognition. Within these models, one would naturally

expect both a perfect correlation of the underlying facilitation and inhibition (i.e., $\rho_{\Delta_f \Delta_i} = 1$) and, consequently, a strong correlation of the measured costs and benefits (i.e., $\text{Corr}[\mathbf{D_f}, \mathbf{D_i}] \gg 0$).

Figure 10 shows example correlations computed for this single-mechanism special case. As is evident from the figure, the correlation of the observed RT facilitation and inhibition (i.e., $\text{Corr}[\mathbf{D_f}, \mathbf{D_i}]$) can be quite low even when the single-mechanism model is correct (i.e., $\rho_{\Delta_f \Delta_i} = 1$). In fact, the correlation can even be negative, which is possible because the random error component of the neutral condition, $\mathbf{E_n}$, enters into the facilitation and interference measures with opposite signs (Brown, 2011). Thus, the clear implication of Fig. 10 is that researchers cannot confidently reject single-mechanism accounts of facilitation and interference based on small correlations of measured facilitation and interference. This example also illustrates another way in which IDRT can be helpful; namely, by providing exact numerical values—possibly rather unexpected ones—for correlations that might be predicted by qualitative theories such as the single-process mechanism.

Correlations between mean reaction times and reaction time difference scores

In some situations, researchers correlate mean RTs with difference scores. As was discussed by Chapman, Chapman, Curran and Miller (1994), for example, one main motivation for such correlations is that slower individuals or groups generally show larger effects of experimental manipulations in most RT tasks. This is to be expected, they argued, because "slow subjects tend to be slow in most aspects of performance with the result that they show greater differences than fast subjects between long-latency and short-latency tasks. By analogy, slow typists tend to show a larger difference in completion times between a long manuscript and a short one than do fast typists" (p. 162). It is particularly important to understand correlations between mean RT and effect size because these correlations complicate the interpretation of different-sized effects found for groups that differ in overall ability (Chapman et al., 1994). In this section we examine the correlation of a mean RT with an RT difference score.[5]

One simple and intuitive way to quantify the relationship between overall RT and effect size is by correlating the observed mean RT in a control condition, $\mathbf{RT_c}$, with the size

of the experimental effect, $\mathbf{D} = \mathbf{RT_e} - \mathbf{RT_c}$. On the basis of the idea that the effects of the experimental manipulation might depend on a participant's overall processing time, one might expect $\text{Corr}[\mathbf{RT_c}, \mathbf{D}]$ to reflect primarily the correlation of the general processing time with the task-specific processing time in the experimental condition, $\rho_{\Delta_e G}$. On the other hand, given that $\mathbf{RT_c}$ enters with opposite signs into the two measurements being correlated (i.e., $\mathbf{RT_c}$ and $\mathbf{D}$), one might expect the correlation to be negative. Given these two conflicting expectations, it is not surprising that the true situation is more complicated than either one suggests.

The influences of various factors on the observable $\text{Corr}[\mathbf{RT_c}, \mathbf{D}]$ can be studied within IDRT. Again we assume that the RTs in the control and experimental conditions are given by Eq. 7, with $i = c, e$, and that these two conditions differ only with respect to processing in the central stage B, so the observed experimental effect $\mathbf{D}$ is given by Eq. 9. Figure 11 illustrates how the correlation of the control RT and the difference score depends on several key parameters. First, there is a clear tendency for the observable $\text{Corr}[\mathbf{RT_c}, \mathbf{D}]$ to increase with the underlying correlation of the general processing time with the task-specific processing time in the experimental condition, $\rho_{\Delta_e G}$, as is intuitively expected. Nonetheless, these two correlations may be quite different, as is indicated by the deviations from the positive diagonal. In fact, depending on the values of the other parameters, the observable $\text{Corr}[\mathbf{RT_c}, \mathbf{D}]$ can be substantially larger or smaller than the underlying target correlation $\rho_{\Delta_e G}$. For example, as is shown within each panel, the observable $\text{Corr}[\mathbf{RT_c}, \mathbf{D}]$ tends to increase with the population variability of general processing time, σ_G, as would be expected because both the mean and the difference tend to increase with $\mathbf{G}$. Not surprisingly, the observable $\text{Corr}[\mathbf{RT_c}, \mathbf{D}]$ also tends to increase with the size of the experimental effect (i.e., with the duration of the central stage B_e). In addition, the observable correlation increases with the correlation of task-specific processing times, $\rho_{\Delta_c \Delta_e}$, which implies that larger observable correlations would be expected when differences involve common, rather than opposing, task-specific processes. Finally, although it is not shown in this figure, the observable $\text{Corr}[\mathbf{RT_c}, \mathbf{D}]$ also tends to increase with the variability of general processing times, σ_G, which is to be expected because the true score variability of each of the two terms involved in the correlation tends to increase with the variability of $\mathbf{G}$.

A slightly different way to assess the relationship between effect size and processing time is to examine the correlation between the size of an experimental effect and the average of the RTs in the control and experimental conditions, $\overline{\mathbf{RT}} = (\mathbf{RT_c} + \mathbf{RT_e})/2$ (e.g., Chapman et al., 1994; Gignac & Vernon, 2004). The analysis of this correlation shows effects that are quite similar to those shown in

---

[5] The general properties of correlations between linear combinations and their components (e.g., $\text{Corr}[\mathbf{X}, \mathbf{Y} - \mathbf{X}]$) have previously been examined outside of the context of RT models (e.g., Brown, 2011; Fuguitt & Lieberson, 1973; Sriram et al., 2010). In contrast, the present investigation focuses on the specific implications of such correlations within IDRT.
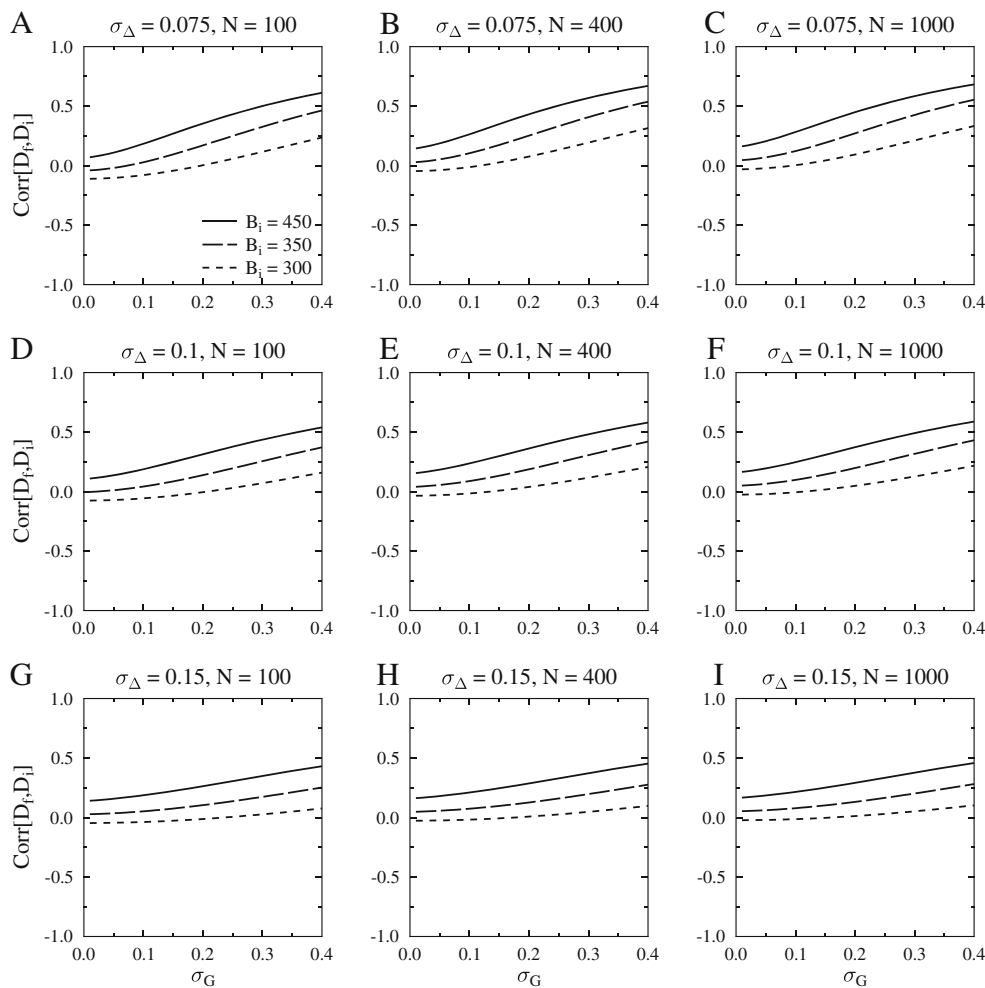
Fig. 10 Correlation of mean reaction time difference scores measuring the benefit of facilitation, $\mathbf{D_f} = \mathbf{RT_n} - \mathbf{RT_f}$, and the cost of interference, $\mathbf{D_i} = \mathbf{RT_i} - \mathbf{RT_n}$, computed using a common neutral condition, $\mathbf{RT_n}$, for the special case in which a single mechanism is responsible for both facilitation and inhibition (i.e., $\rho_{\Delta_f \Delta_i} = 1$). Correlations are displayed as a function of the number of trials, the amount of cost indexed by B_i relative to fixed values of $B_f = 200$ and $B_n = 250$, the standard deviation of $\mathbf{G}$, and the common standard deviation of all task-specific processing times ($\sigma_{\Delta_f} = \sigma_{\Delta_n} = \sigma_{\Delta_i} \equiv \sigma_\Delta$). Values of $\rho_{\Delta_f G} = \rho_{\Delta_i G} = 0$ were assumed, in contrast to $\rho_{\Delta_n G} = .2$, because of two constraints inherent in the single-mechanism model. First, $\mathbf{\Delta_f}$ and $\mathbf{\Delta_i}$ necessarily have the same correlation with $\mathbf{G}$ if they are perfectly correlated with one another. Second, $\mathbf{\Delta_f}$ and $\mathbf{\Delta_i}$ have opposite correlations with $\mathbf{RT_f}$ and $\mathbf{RT_i}$, respectively, because the facilitation term is subtracted from $\mathbf{RT_f}$, whereas the cost term is added to $\mathbf{RT_i}$ (Eqs. 12 and 10). If $\mathbf{G}$ is to be equally correlated with $\mathbf{RT_f}$ and $\mathbf{RT_i}$, then, $\rho_{\Delta_f G} = \rho_{\Delta_i G} = 0$ is the only possibility. Default values of the other parameters are shown in Table 15. Correlations were computed using the same equations indicated in Fig. 9

Fig. 11, so the situation apparently only changes minimally when using $(\mathbf{RT_c} + \mathbf{RT_e})/2$ rather than $\mathbf{RT_c}$ to index faster versus slower participants. The main difference between the two cases is that the correlations are slightly higher using $\overline{\mathbf{RT}}$. This makes sense, because the correlation would tend to be increased by the positive contribution of the experimental condition mean, $\mathbf{RT_e}$, to both of the terms being correlated. The bottom line, then, is that many factors influence the correlation between overall processing time and effect size within IDRT, whether processing time is indexed by $\mathbf{RT_c}$ or $(\mathbf{RT_c} + \mathbf{RT_e})/2$. In particular, the model suggests that it will not be easy to determine whether a slower group shows a larger effect size just because of general slowing (i.e.,

changes in $\mathbf{G}$). Instead, the appropriate adjustment in effect size for general slowing depends on a number of parameters, and this adjustment can be determined only within the context of a specific model.

General discussion

Analyses of correlations involving RT play a prominent role in research investigating both individual-differences (e.g., Jensen, 1985, 1993; Sheppard & Vernon, 2008) and basic cognitive processes (e.g., Corballis, 2002; Stolz et al., 2005). Although researchers conducting such analyses have often

Fig. 11 Correlation of the mean reaction time in a control condition, **RT_c**, with the difference between means, **D = RT_e − RT_c**. Correlations are displayed as a function of the size of the experimental effect, indexed by B_e, the standard deviation of **G**, and the correlations $\rho_{\Delta_c \Delta_e}$ and $\rho_{\Delta_e G}$. Default values of the other parameters are shown in Table 15. Corr[**RT_c**, **D**] was computed using Eq. 23 from the covariance given in Eq. 50 and from the variances given in Eqs. 38 and 43
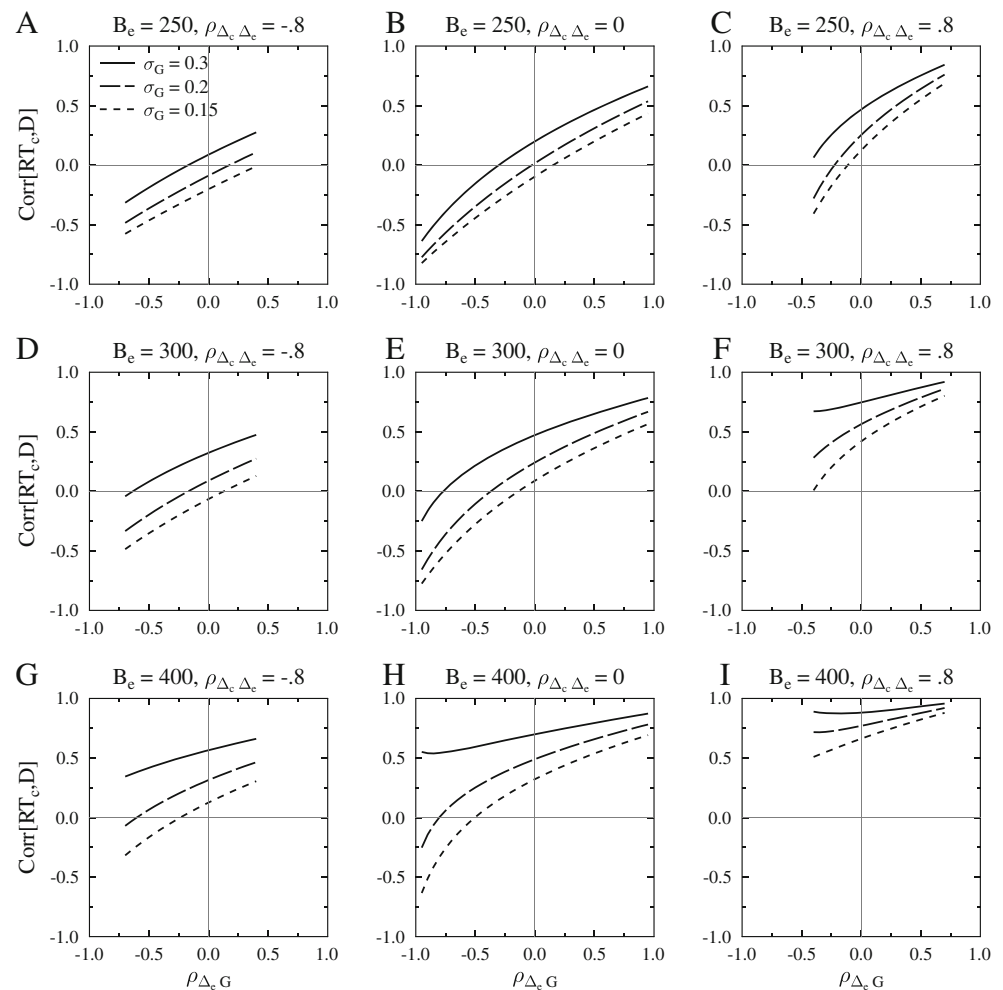
considered classical psychometric testing concepts (e.g., reliability) in assessing their correlations, no attempt has been made to assess the precise meanings of these correlations within the framework of standard RT models. Therefore, the goal of this article was to investigate how RT-based correlations would be influenced by the various underlying processes within standard RT models. To achieve that, we developed a general model of individual differences in RT, called IDRT, and linked this model to psychometric concepts from classical test theory. This linkage was especially direct because IDRT involves a linear combination of random variables. We explored the consequences of this model for several different types of correlational analyses involving RTs. In particular, the model's predictions can be determined regarding correlations involving both mean RTs and difference scores (see Table 2).

Our model is based on the simple assumption that processing from stimulus input to response output proceeds via a series of computational steps or stages whose durations sum to produce the overall RT. Within IDRT, individuals of course differ in the time needed to carry out each of the stages (e.g., Vernon, 1990). This model makes it possible to distinguish between general and task-specific processing

times, and the influences of these times on various observable correlations can be assessed, thereby helping to elucidate the precise meanings of such correlations. This model is attractive because of its simplicity, generality, and extensive theoretical development (e.g., Donders, 1868/1969; Smith, 1969; Sternberg, 1969, 2001). At the same time, as was discussed in the section "The individual differences in reaction time (IDRT) model," it seems plausible that the conclusions emerging from this simple model system would also be applicable within more detailed models providing a richer description of specific RT tasks (Hillis, 1993).

Implications regarding correlations

The most important general conclusion emerging from our analysis is that the observable correlations involving RT means and difference scores depend on many factors influencing performance within the task, including characteristics of both the task (e.g., times needed for perceptual, central, and motor processing [A, B, C]) and the population (e.g., variability of general and task-specific processing times, and their correlation [σ_G, σ_Δ, and $\rho_{\Delta G}$]). Obviously,

the fact that the observable correlations are influenced by so many parameters greatly complicates the interpretation of any particular observed correlation. This finding critically underscores the need for extreme caution in interpreting observed correlations, especially because there are cases in which correlations can be expected to be far higher or far lower than the correlations of internal parameters that they might be intuitively assumed to measure. Ultimately, this finding raises the question of just what is actually being learned about individual differences and mental processes by studying such correlations. Although the present general model provides a first step toward understanding the implications of RT-based correlations, it is clear that there is a long way to go before it will be possible to draw strong conclusions from the size or in some cases even from the direction of an RT-based correlation.

The equations for the correlations predicted by the model clearly illustrate the above general conclusion for each of the different scenarios we examined (see Table 2). For example, Eq. 42 shows that even the correlation between two mean RTs—one of the simplest cases—depends on at least ten parameters. From an observed correlation of means, then, it is therefore impossible to estimate the value of a single parameter of interest without detailed knowledge about the other parameters. As a second example, Eq. 50 shows how the correlation between an effect size and the mean RT in the control condition depends on many parameters affecting the mean RTs in both the experimental and control conditions. Thus, although it seems intuitively reasonable to ask about such a correlation (i.e., whether an effect size increases for individuals who are slower in the control condition), the observed correlation value simply has no straightforward interpretation in terms of the underlying RT processes involved in the two conditions.

The present findings weaken many previous conclusions based on correlations of RT measures. As one example, consider the finding of low correlations among several different attentional effects reported by Fan et al. (2002), which was discussed in the earlier section "Correlation of two distinct reaction time difference scores". It is certainly possible that these low correlations reflect truly independent neural mechanisms involved in the different attentional systems assessed via the RT difference scores, which is what Fan et al. concluded. Figure 8 shows, however, that—depending on the values of the other parameters—there could actually be a rather high correlation between the times needed for the task-specific mechanisms used in the experimental conditions (i.e., $\rho_{\Delta_e \Delta_v}$), despite the fact that the correlation of the RT difference scores is very low. Thus, it is possible that a low correlation of difference scores actually provides only illusory evidence of functional dissociations between the task-specific mechanisms under study. Before accepting a low correlation as evidence of a functional dissociation, then, it would be necessary to show

that the values of the other parameters were not responsible for the low observed correlation. Unfortunately, it is not yet clear how to do this, because none of these parameters can be estimated directly.

Similarly, the present results raise doubts about the interpretation of weak correlations between RT-based measures of facilitation and interference. Some models of the Stroop (1935) task suggest, for example, that interference and facilitation are driven by the same underlying word recognition mechanism. Because the two effects are opposite sides of the same coin within these models, the models seem intuitively to predict that the effects should be strongly correlated. On that basis, findings of weak correlations have been regarded as evidence against single-mechanism models (Brown, 2011). As is illustrated in Fig. 10, however, such models need not predict a strong correlation between facilitation and interference. Depending on the values of other parameters, they may predict small or even negative correlations. Thus, the absence of a strong correlation between facilitation and interference does not actually imply the existence of separate mechanisms underlying the two effects.[6]

Despite the complications evident in the formulas, the numerical results indicate that some types of observed correlations do sometimes provide very good estimates of underlying relationships of interest to researchers. For example, Fig. 3 shows that the correlation Corr[**RT**, **Y**] between mean RT and the score on an external (i.e., non-RT) measure, **Y**, is often quite similar to the correlation of general processing time, **G**, with that measure. This result provides encouragement that it may be possible to use mean RTs for fairly accurate assessments of correlations of external measures with general processing time, and it is therefore quite consistent with the large literature suggesting that the mean RTs of many tasks correlate well with general intelligence (e.g., Jensen, 2006).

What can be concluded from RT-based correlations?

Even if RT-based correlations are highly replicable empirical phenomena, they may be devilishly complicated to interpret. Given the multiplicity of factors influencing RT-based correlations of each type shown in Table 2, researchers must obviously be cautious in interpreting observed values of these correlations. Statistical reliability of

---

[6] In theory, the present analysis could be used to derive more appropriate tests of single-mechanism models. Using the model developed here, it would be possible to determine the exact correlation predicted by a single-mechanism model, ρ_p. Testing this prediction would then provide a direct assessment of the model. In practice, however, this approach would require numerical values for each of the parameters determining the predicted correlation, and it is not clear how these values could be determined.

the observed values should be assessed as usual, but the interpretations of both statistically significant and nonsignificant correlations must also take into account the many possible influences that could be responsible for the results.

Consider, for example, a significant correlation between the mean RTs of two distinct tasks, RT_x and RT_y. Although it may be tempting to attribute this correlation to a common task-specific central process hypothesized to be involved in both tasks (i.e., $\Delta_x = \Delta_y$), the correlation may actually be produced mainly by something else entirely, such as a general processing time parameter (G) common to all tasks (e.g., Jensen, 2006) or a correlation of the sensory–motor residual times (R) of the two tasks.

It seems clear that relatively sophisticated research strategies will be required to reach strong conclusions from between-task correlations of mean RTs. Specifically, researchers will need to base their conclusions on the patterns of correlations across a range of tasks—not just pairs of tasks. For example, suppose that the correlation of RT_x and RT_y is demonstrably higher than the correlation of either of these with a third task's RT_z (within the same sample of participants). If the third task were constructed so that RT_z involved the same sensory–motor residual times as RT_x and RT_y, and if all three tasks appeared to depend equally strongly on general processing time G (e.g., because all correlated equally well with IQ), then the researcher would clearly be on stronger ground in attributing at least part of the RT_x/RT_y correlation to a hypothesized common central process contributing to RT_x and RT_y but not RT_z.

In view of the multiplicity of influences on correlations, it is perhaps surprising that the present results do suggest that very strong correlations can have quite specific implications. Consider, for example, Iacoboni and Zaidel's (2004) report of a .9 correlation between the crossed–uncrossed difference in simple RT (i.e., stimulus light on the same vs. opposite side of the body midline as the respond hand) and an fMRI-based measure of activity in the right superior parietal cortex. From this strong relationship, they concluded that this area has "a key role . . . in the type of interhemispheric visuo-motor integration required by [the task]" (Iacoboni & Zaidel, 2004, p. 423), but even more specific conclusions can be reached on the basis of IDRT. First, it seems clear that this RT difference must reflect opposing or unrelated task-specific processes in the two RT conditions being compared (i.e., $\rho_{\Delta_c \Delta_e} \leq 0$), because such a strong correlation is not found with common task-specific processes (i.e., $\rho_{\Delta_c \Delta_e} \gg 0$; Fig. 7). Second, activity in the right superior parietal cortex must have been both negatively correlated with RT in the uncrossed condition ($\rho_{\Delta_c Y} \ll 0$) and positively correlated with RT in the crossed condition ($\rho_{\Delta_e Y} \gg 0$), because very strong correlations of RT difference scores with an external measure are not found unless both of these requirements are met (e.g., Fig. 7). In short, given the multiplicity of influences on RT-based

correlations, extreme values near ±1 can be found only when most or all of the relevant parameters have certain required settings.

On the other hand, the interpretations of small correlations are much more poorly constrained. Consider, for example, possible interpretations of the finding that two RT-based effect sizes are only weakly correlated, as in the case of Fan et al.'s (2002) attentional effects. It would be tempting to conclude that different mechanisms are responsible for the effects in the two experimental conditions, but our analysis shows that other interpretations are possible. For instance, the correlation of the two effect sizes is also strongly influenced by the correlation of the RTs in the two control conditions (Fig. 7), and the effect sizes could be weakly correlated even when the same mechanism was responsible for both effects if the RTs in the two control conditions were negatively correlated. Indeed, given the rich set of constraints among the correlations of the four conditions involved in the difference scores (i.e., two control conditions and two experimental conditions), it seems clear that the entire set of correlations needs to be examined when assessing the mechanisms involved in producing the different experimental effects.

Again, more sophisticated research strategies can help to strengthen conclusions from correlations of RT-based effects. As an example, consider the study of Miles and Proctor (2012), who examined correlations of Simon compatibility effects obtained with three different types of stimulus materials (i.e., locations, arrows, and words). They found a significant correlation between the compatibility effects obtained with arrows and words, but no correlation between either of these effects and the compatibility effect obtained with location stimuli. This pattern of changing correlations among fairly similar tasks provides stronger support for the claim that the underlying mechanisms responsible for Simon effects with arrows and words have more in common with each other than either one does with the mechanisms responsible for location-based Simon effects. On the other hand, it is impossible to be certain about this conclusion in the absence of a complete model for the task, because it is might be possible to construct models that produce unequal correlations despite having common mechanisms for all three stimulus types. The present work strongly suggests that the proper interpretation of RT correlations requires explicit models, perhaps even more so than the interpretation of mean RT results.

It should be emphasized that our conclusions about RT-based correlations apply only to situations where correlations are computed *across participants*—not where they are computed *across trials* for a given participant. In studies of the performance of two successive tasks within the psychological refractory period paradigm, for example, many researchers have examined the correlation across trials

between the RTs of the two tasks (e.g., Davis, 1959; Pashler & Johnston, 1989; Sigman & Dehaene, 2006; Way & Gottsdanker, 1968). Both the task parameters and the individual-difference parameters of the present version of IDRT (e.g., A, B, $\mathbf{G}$, $\boldsymbol{\Delta}$, $\mathbf{R}$) would be held constant across trials within such correlations, so an elaborated version of the model would be needed for the analysis of such correlations.[7]

Would partial correlations avoid the problems of difference scores?

As has been discussed already, the main rationale for using an RT difference score like $\mathbf{D} = \mathbf{RT_e} - \mathbf{RT_c}$ is usually to isolate the influence of a specific processing stage lengthened in the experimental condition, removing the effects of stages common to the experimental and control conditions. For example, the correlation of RT benefits and costs measured relative to a common neutral condition, $\mathrm{Corr}[\mathbf{RT_n} - \mathbf{RT_f}, \mathbf{RT_i} - \mathbf{RT_n}] = \mathrm{Corr}[\mathbf{D_f}, \mathbf{D_i}]$, is intended to assess the relationship between the task-specific processes generating those costs and benefits, $\rho_{\Delta_f \Delta_i}$. We have seen, however, that the observable correlation $\mathrm{Corr}[\mathbf{D_f}, \mathbf{D_i}]$ does not accomplish its intended task (e.g., Fig. 9), because this correlation is influenced by many parameters other than the intended one.

Some readers might wonder whether partial correlations would avoid the difficulties associated with difference scores. For example, researchers could compute the partial correlation of the mean RTs in the facilitation and inhibition conditions, partialling out the mean RT in the neutral condition, $\mathrm{Corr}[\mathbf{RT_f}, \mathbf{RT_i}|\mathbf{RT_n}]$. The partial correlation seems to have some intuitive appeal for this purpose, given its usual interpretation as "removing the effect of" the variable being partialled out. Thus, for example, this partial correlation might intuitively be expected to provide another way to assess the relationship between benefits and costs after removing effects that were common with the neutral condition, just as the difference score was meant to do. In fact, however, despite the intuitive similarity of the partial correlation and the difference score in removing effects of neutral condition performance, these measures are not identical (i.e., $\mathrm{Corr}[\mathbf{RT_f}, \mathbf{RT_i}|\mathbf{RT_n}] \neq \mathrm{Corr}[\mathbf{D_f}, \mathbf{D_i}]$). Therefore, it is mathematically possible that the partial correlation would directly assess the desired relationship of the task-specific

processing times in the facilitation and inhibition conditions, $\rho_{\Delta_f \Delta_i}$, even though correlation of the difference scores, $\mathrm{Corr}[\mathbf{D_f}, \mathbf{D_i}]$, does not.

Fortunately, this possibility can also be examined within IDRT. In general, a partial correlation measures the relationship between $\mathbf{X}$ and $\mathbf{Y}$ after each of these two variables have been adjusted by linear regression to remove any association with $\mathbf{Z}$ (i.e., partialling out $\mathbf{Z}$'s contribution to $\mathbf{X}$ and to $\mathbf{Y}$). The partial correlation is therefore defined as $\mathrm{Corr}[\mathbf{X}, \mathbf{Y}|\mathbf{Z}] = \mathrm{Corr}[\mathbf{U}, \mathbf{V}]$, where $\mathbf{U}$ and $\mathbf{V}$ are the residuals when $\mathbf{X}$ and $\mathbf{Y}$ are regressed on $\mathbf{Z}$. Conveniently, the correlation of the residuals can be computed from the three pairwise correlations of the original variables $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$. For example, the partial correlation between $\mathbf{RT_f}$ and $\mathbf{RT_i}$ controlling for $\mathbf{RT_n}$ is

$$
\mathrm{Corr}[\mathbf{RT_f}, \mathbf{RT_i}|\mathbf{RT_n}]
$$
$$
= \frac{\mathrm{Corr}[\mathbf{RT_f}, \mathbf{RT_i}] - \mathrm{Corr}[\mathbf{RT_f}, \mathbf{RT_n}] \cdot \mathrm{Corr}[\mathbf{RT_i}, \mathbf{RT_n}]}{\sqrt{\left(1 - \mathrm{Corr}\left[\mathbf{RT_f}, \mathbf{RT_n}\right]^2\right) \cdot \left(1 - \mathrm{Corr}\left[\mathbf{RT_i}, \mathbf{RT_n}\right]^2\right)}}.
$$
(15)

It is possible to investigate partial correlations within IDRT in basically the same manner that we have used to investigate simple correlations. The pairwise correlation between any two mean RTs can be derived using the methods discussed in the section "Correlation of two mean reaction times," so the full predicted correlation matrix for any set of mean RTs can be obtained by repeated pairwise applications of these methods. Then, the partial correlation of two variables controlling for one or more additional variables can be determined from these pairwise correlations (e.g., Eq. 15). In short, predicted partial correlations can be derived from IDRT because they depend only on predicted pairwise correlations.

It is beyond the scope of this article to present or illustrate the partial correlations predicted by IDRT, but the overall conclusion of such an analysis is clear. Just like simple correlations, partial correlations are also influenced by numerous parameters beyond the ones of interest to researchers, so these are subject to the same sorts of interpretation difficulties that plague difference score correlations.

Ultimately, it appears that neither correlations of difference scores nor partial correlations can measure the simple relationships of intuitive interest, because the underlying RT process is not completely additive.[8] Consider, for example,

---

[7] Within bottleneck models of psychological refractory period tasks, for example, the correlation between the RTs of the two tasks arises because (1) the individual stage times within each task are random variables and (2) task 2 central processing cannot start until task 1 central processing is finished. Within IDRT, the random components of single-trial RTs were aggregated into a single term ($\mathbf{E}$) for simplicity, and it would be necessary to elaborate IDRT to have separate random components for each stage in order to derive its predictions about correlations computed across trials for a single individual.

[8] The same basic problems also arise with a further type of correlation known as *semi-partial* correlation. This type of correlation assesses the relation between two variables when the influence of a third variable is removed from one—but not both—of the first two variables. Although we have not discussed semi-partial correlations explicitly, the same sorts of complications with these correlations arise within IDRT, preventing any straightforward interpretations of their values in terms of relationships between particular cognitive processes.

the general processing time parameter **G**, which is related to a hypothesized neural processing speed influencing all stages. Within any general model of RT, this parameter has a multiplicative effect [i.e., $\mathbf{RT} = (A + B + C) \times \mathbf{G} + \mathbf{R}$] rather than a purely additive one as is assumed by the general model underlying difference scores and partial correlations. This is even true for IDRT despite the fact that it was constructed to have a relatively simple additive structure in the first place, so it is extremely doubtful that correlations would have simpler interpretations within other, less additive RT models.

Implications regarding reliability

As is well known within classical test theory and has been acknowledged by many researchers focusing on RT (e.g., Jensen, 1985), the reliabilities of RT-based measures are crucial. In general, reliability is determined by true score variance and error variance, and it is important because it places an upper limit on the correlations that can be observed.

Regarding mean RTs, IDRT provides some grounds for optimism. Although the number of observations needed for satisfactorily high reliability of a mean RT depends on the exact situation under study, as few as 15–30 trials are often enough under realistic parameter settings, and it is usually feasible to obtain at least that many trials per participant in all conditions of an RT study. On the other hand, IDRT indicates that high reliability per se is not a sufficient indication that a population has adequate variance in the population parameters of interest. As is shown in Fig. 2, reliability tends to be high if there is a reasonable amount of variability in at least one of the general processing time, **G**, the task-specific processing time, **Δ**, or the sensory–motor residual component, **R**. Thus, mean RT reliability could be high without any variability in the cognitive processing times that are generally of interest, **G** and **Δ**—and thus, without any opportunity for observing correlations of RT with other cognitive measures—as long there was large variability in the residual component.

The situation is somewhat less promising with respect to the reliability of difference scores, because more observations—possibly two orders of magnitude more —are needed for a reliable difference score. The reliability of difference scores is reduced partly because the sensory–motor residual term **R** does not contribute to the difference (Eq. 9), reducing the true score variation. Happily, this means that when the reliability of a difference score is high—unlike with mean RTs—the researcher can be sure that there was substantial variability in at least one of the cognitive processing time parameters (i.e., **G**, **Δ_c**, and **Δ_e**).

Critically, the reliability of a difference score depends greatly on the relation between the two conditions involved in the difference. Specifically, the reliability of the difference between the mean RTs of two tasks depends on the similarity of the underlying task-specific processes involved in those tasks. When the two tasks involve common task-specific processes, so that these processes are positively correlated across participants (i.e., $\rho_{\Delta_c \Delta_e} \gg 0$), reliability tends to be relatively low. Difference scores computed from two conditions requiring different degrees of mental rotation, different memory loads, different display sizes in visual search, different temporal offsets between two overlapping tasks, and so on would be examples. In such cases, thousands of trials per condition might be required to obtain satisfactory levels of reliability. Not surprisingly, then, the literature contains numerous reports of low reliability for difference scores computed from such tasks (e.g., Neubauer et al., 1997).

In contrast, when the two tasks involve opposing task-specific processes, so that these processes are negatively correlated across participants (i.e., $\rho_{\Delta_c \Delta_e} \ll 0$), reliability tends to be relatively high. A difference score computed from two conditions involving congruent versus incongruent trials in the Stroop task would be one example, and others might involve congruent versus incongruent conditions in the SNARC effect, Simon effect, flanker effect, stimulus–response compatibility tasks, and so on. In these cases, as few as 100–200 trials per condition might be sufficient to achieve adequate levels of reliability. Since correlations are inherently limited by reliability, the similarity of the underlying task-specific processes also has strong implications for correlations of difference scores with mean RTs and with external measures, as well as for reliability.

The practical implications of IDRT for reliability are nicely illustrated by considering a recent pair of studies examining the reliability of priming effects on word recognition. Stolz et al. (2005) found that semantic priming effects were rather unreliable across two blocks of trials, suggesting that these effects are driven by "uncoordinated processes specific to semantic memory" (Waechter et al., 2010, p. 553). In contrast, using a closely-matched experimental protocol, Waechter et al. (2010) found that repetition priming effects were noticeably more reliable than semantic priming. The higher reliability of repetition priming was taken as further support for the idea of uncoordinated semantic memory processes because it ruled out explanations of low reliability based on uncoordinated processes at presemantic (i.e., featural, lexical) levels. Within the context of IDRT, however, it is easy to imagine another possible interpretation of the higher reliability for repetition priming than for semantic priming. Specifically, the repetition priming effect was approximately twice as large

as the semantic priming effect (M = 88 vs. 37 ms). Given that the reliability of a difference score increases with the effect size (e.g., effect of B_e in Fig. 5), the larger effect size could have been responsible for the greater reliability of repetition priming, negating its support for the uncoordinated nature of semantic processes. Similarly, Maloney et al. (2010) found much lower reliability for numerical distance effects obtained with numbers presented in symbolic formats (e.g., "4") than with those presented nonsymbolically (e.g., four squares), and these reliability differences may also have been due at least partly to the fact that the distance effect was much smaller with symbolic than with nonsymbolic stimuli (i.e., approximately 50 vs. 500 ms). Thus, as was the case with RT correlations, IDRT is useful in elucidating the many factors that need to be considered when interpreting changes in RT reliabilities.

Further uses of IDRT

The model developed here could also be useful in addressing various methodological questions affecting the exact choice of data analyses. Consider, for example, the issue of whether it is better to use common versus separate estimates of RT means contributing to both of two terms being correlated. As was discussed by Brown (2011), this question arises in correlating the sizes of Stroop facilitation and interference, because each of these effect sizes is estimated relative to the mean RT in a common neutral condition. Including all of the available neutral trials in a single estimate of the neutral mean has the advantage of yielding an estimate based on a larger number of trials but has the corresponding disadvantage of creating an artificial dependence between the two effect sizes being correlated. In contrast, dividing the neutral trials into two sets and computing separate estimates has the advantage of yielding independent estimates but the disadvantage of yielding estimates based on smaller numbers of trials. Exactly analogous questions arise when correlating an effect size with the control condition mean, $\text{Corr}[\mathbf{RT_c},\mathbf{D}]$, or with the average of the control and experimental condition means, $\text{Corr}[\overline{\mathbf{RT}}, \mathbf{D}]$, because in both of these cases, at least one of the condition means, $\mathbf{RT_c}$ and $\mathbf{RT_e}$, contributes to both of the terms being correlated.

Within the present model, the implications of using common versus separate RT estimates can be examined precisely under any desired set of assumptions about the task and individual-difference parameters. As was discussed in the section "Correlation of two reaction time difference scores involving the same baseline," trial-to-trial error variance contributes to the covariance when a common RT mean is used, but not when separate RT means are used. Furthermore, the error variance

of a single mean varies with the number of trials used to compute it (Eq. 2). Thus, the true underlying correlation of interest (e.g., $\text{Corr}[\mathbf{D_f}, \mathbf{D_i}]$) can be computed exactly for both analysis procedures (i.e., common vs. separate estimates), allowing a fully informed choice about which is the better method under a particular configuration of parameter values.

Future directions

As an initial investigation of the psychometric properties implied by RT models, the present work has a number of limitations that should be addressed in future extensions. One is that we have only examined measures based on mean RTs and their differences. Many other summaries of RT have also been used in computing correlations, including median RT, within-condition standard deviation of RT, and parameters obtained by fitting particular models to RT distributions (e.g., Fjell, Ostby, & Walhovd, 2007; Forstmann et al., 2008; Jensen, 1992; Schmiedek, Lovden, & Lindenberger, 2009; Schmiedek, Oberauer, Wilhelm, Süß, & Wittmann, 2007). Future research could extend the present analysis to such other RT summary measures. Such extensions would appear to be quite straightforward in some cases. For example, IDRT's predictions about the within-condition standard deviation of RT are dictated by Eq. 2, so predicted standard deviations can easily be computed from the same parameters used to compute predicted means.

A second limitation of the present work is that we have only considered the true values of the reliabilities and correlations implied by the model. In any empirical study, of course, the reliabilities and correlations would be estimated from observed data and would therefore fluctuate randomly around the true values provided by our formulas. We have not explored the implications of the IDRT model for these purely statistical fluctuations due to sampling error, and these implications could be important. For example, the reliability of a mean RT or a difference score could be estimated using a test–retest procedure, a split-half analysis, or some other technique, and it is not clear which of these would provide the reliability estimate with the best statistical properties (i.e., least bias and lowest standard error).

A third limitation of this work is that we have examined in detail only the relationships of observable correlations to the correlations between certain pairs of underlying parameters. Figure 7, for example, shows how the observable correlation of an external measure $\mathbf{Y}$ with an RT difference score is related to the underlying correlation of $\mathbf{Y}$ with the time needed for task-specific processing in the experimental condition, $\mathbf{\Delta_e}$. One might ask, instead, how the observable correlation is related to $\mathbf{Y}$'s correlation with the time needed for task-specific processing in

the control condition ($\mathbf{\Delta_c}$), or even how it is related to $\mathbf{Y}$'s correlation with the difference in task-specific processing times (i.e., $\mathbf{\Delta}^* = B_e\mathbf{\Delta_e} - B_c\mathbf{\Delta_c}$). It is beyond the scope of this initial investigation to consider all such possible relationships, but two points can be made. First, the equations developed in Appendix 2 could be used to study IDRT's predictions concerning any such relationship of particular interest. As one illustration, Fig. 12 shows how Corr[$\mathbf{D}$, $\mathbf{Y}$] is related to the underlying correlation of $\mathbf{Y}$ with the difference in experimental versus control central processing times just defined, ρ_{Δ^*Y}, under various assumptions about the other parameters. Second, and in keeping with the more general conclusions from this research, the complexity of these equations strongly suggests that all relationships between observable correlations and underlying parameters will be complicated, making it difficult to reach straightforward conclusions about the correlations of underlying processing durations from observable correlations.

A fourth limitation of this work is that it ignores the possibility that the observed RTs are contaminated by speed–accuracy trade-offs. The possibility of such contamination plagues all RT research (e.g., Pachella, 1974), of course—not just research focusing on correlations. Within correlational RT research, it appears that the possibility of such contamination can be addressed only by using a formal model to combine RT and accuracy into a single measure of processing efficiency (e.g., Brown & Heathcote, 2005; Ratcliff, 1978; Yellott, 1971).

Finally, it might also be worthwhile to extend the present approach to other analytical techniques beyond the computation of reliabilities and correlations. As one example, an alternative approach for investigating individual differences is to divide participants into groups based on one variable (e.g., age or IQ) and then to compare mean RTs and RT differences across the groups (e.g., Der & Deary, 2006; Dickman & Meyer, 1988; Eaton & Ritchot, 1995; Ellermeier, Eigenstetter, & Zimmer, 2001; Exposito & Andres-Pueyo, 1997; Myerson, Hale, Chen, & Lawrence, 1997; Smulders & Meijer, 2008). A persistent problem within this approach is to make a fair comparison of the sizes of RT differences across groups differing in overall mean RT. For example, Chapman et al. (1994) suggested that "slower or less accurate individuals tend generally to show larger differences between pairs of scores [RTs], and this may explain the finding in many kinds of tasks that slower and less accurate groups . . . show heightened [RT] priming difference scores" (p. 160). Researchers have developed a number of ad hoc procedures to adjust RT effect sizes in order to more fairly compare effect sizes across groups (e.g., ANCOVA), but none of these procedures has been developed from an explicit model of the underlying RTs. If the IDRT model could be extended to this situation, it might be useful for comparing the effectiveness of different suggested adjustment procedures or even to find a new model-based adjustment procedure for making the desired comparisons.

Appendix 1 Classical test theory background

This appendix reviews briefly some major concepts and results from classical test theory that underlie the analysis of correlational research using any dependent measure, including RT. Naturally, we emphasize aspects of classical test theory that are especially relevant for the analysis of RT-based correlations, including some equations not presented in its standard references (e.g., Eqs. 32–35). As is developed in the main text of this article, it is possible to address these issues more specifically within the RT domain by considering a general model of RT within the classical framework. In the present appendix, we adapt somewhat the standard notation used within classical test theory in order to match more naturally the standard RT notation used in the main text.

Basic model for an individual

The basic model of classical test theory is that the observed score $\mathbf{X}_k$ for a single individual k in a particular test is the sum of that individual's true score, T_k, on that test, which is regarded as a fixed value for each individual, and a random measurement error, $\mathbf{E}_k$:

$$\mathbf{X}_k = T_k + \mathbf{E}_k. \tag{16}$$

A basic assumption in classical test theory is that the expected value of the observed score equals the true score,

$$\mathrm{E}[\mathbf{X}_k] = T_k, \tag{17}$$

and this implies that the expected value of the error score equals zero,

$$\mathrm{E}[\mathbf{E}_k] = 0. \tag{18}$$

This basic model also implies that the variance of a single individual k's observed scores equals the variance of the random error term,

$$\mathrm{Var}[\mathbf{X}_k] = \mathrm{Var}[\mathbf{E}_k]. \tag{19}$$

Distribution across individuals

Naturally, true scores differ across individuals. Therefore, for a randomly selected individual, the above additive model becomes

Fig. 12 Correlation of a mean reaction time difference score, **D**, and an external measure, **Y**, as a function of the true correlations ρ_{GY}, $\rho_{\Delta_c\Delta_e}$, and ρ_{Δ^*Y} and of the assumed relationship between ρ_{Δ_cY} and ρ_{Δ_eY}. Default values of the other parameters are shown in Table 15. As in Fig. 7, Corr[**D**, **Y**] was computed using Eq. 47. Values of $\rho_{\Delta^*\mathbf{Y}}$ were computed using Eq. 55
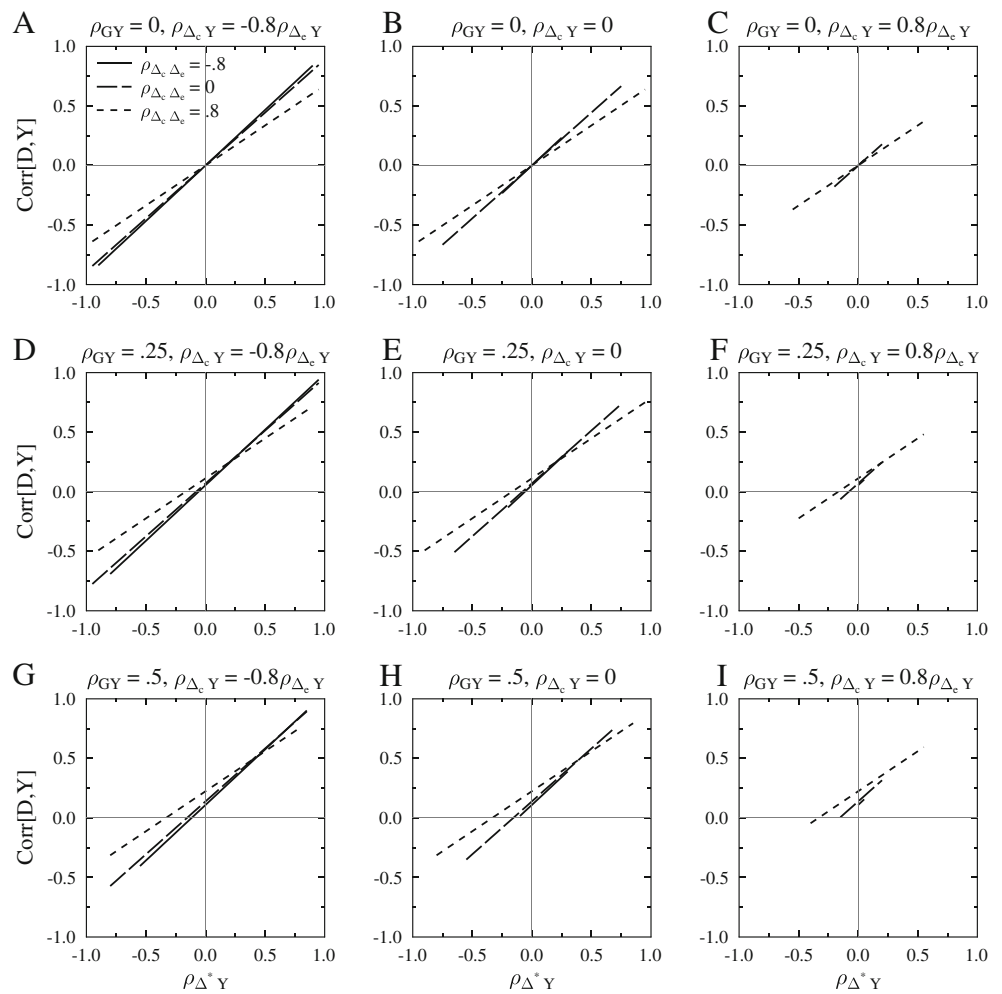
$$\mathbf{X} = \mathbf{T} + \mathbf{E}. \tag{20}$$

Note that a true score is a fixed value when considered for one particular individual, T_k, but it is a random variable, **T**, when considered for a randomly selected individual.

As is depicted by the solid line in Fig. 13 and is also illustrated in Table 3, the full distribution of **X** modeled by Eq. 20 is a *mixture distribution*. The components of this mixture are the distributions associated with the particular individuals in the population, depicted by the dotted lines in Fig. 13 and modeled by Eq. 16. Thus, one can imagine obtaining a value of **X** by a two-step process: (1) Select a random individual k from the population, and (2) select a particular observed score $X_k = T_k + E_k$ from that person's individual distribution. Within the domain of RT research, for example, the X_k values might represent individual participants' mean RTs in a certain condition. Each true value T_k would be participant k's true mean RT in that condition, which could be measured only across an infinite number of trials. The error term, E_k, would reflect the standard error of the mean for that participant, which would naturally depend on the variability of that participant's RT distribution and on the number of trials in which that participant was tested.

As one might expect, the properties of the mixture distribution are fully determined by the properties of the individual component distributions that contribute to it (see Everitt & Hand, 1981, for formal details). Because the components of the mixture are equally likely in this case, and because $E[E_k] = 0$ must hold, it follows from Eq. 20, that the mean of **X** is

$$E[\mathbf{X}] = E[\mathbf{T}]. \tag{21}$$

Furthermore, the variance of **X** is

$$\text{Var}[\mathbf{X}] = \text{Var}[\mathbf{T}] + E[\text{Var}[E_k]]. \tag{22}$$

Equation 22 expresses the fact that the total variance of the mixture distribution of observed scores is the sum of (1) the true score variance Var [**T**] over individuals' true scores and (2) the average error variance $\text{Var}[\mathbf{E}] = E[\text{Var}[E_k]]$ over individuals' error variances. Note that $\text{Var}[E_k]$ is the error variance associated with a particular individual k, and the
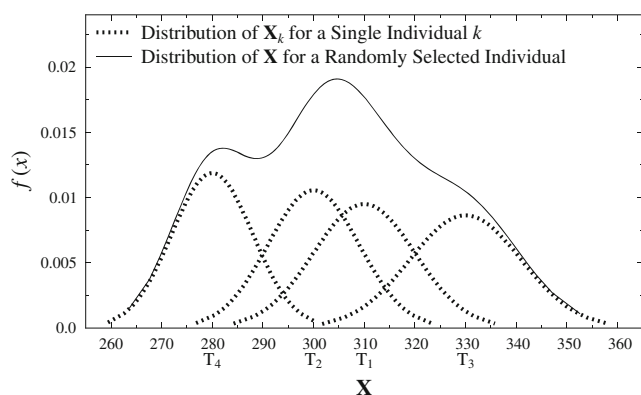
Fig. 13 Illustration of the model represented by Eq. 20 using an idealized population with only four individuals. As is listed in Table 3, each individual k has a true score T_k and a measurement error variance Var [E_k]. The dotted lines show the probability density functions (PDFs) of the observed scores for each individual (i.e., $X_k = T_k + E_k$); these are normal distributions centered at the true scores. The solid line shows the PDF of the observed score from a randomly selected individual, X; this is a mixture distribution, with scores equally likely to come from each of the individual distributions. As is summarized in Table 3, the mean and variance of the population distribution X can be computed from the properties of the individual distributions, T_k and Var [E_k]. The plotted PDFs of the individuals have been rescaled to an area one fourth that of the mixture distribution to improve visual clarity and to reflect their contributions to the mixture distribution

expectation of this quantity is the average error variance across all individuals.

Reliability of single scores

A fundamental concept of classical test theory is the *reliability* of a measure (e.g., a psychometric test or—as is mainly considered in this article—an individual measure derived from one or more mean RTs). Although many definitions of reliability are possible, a common and intuitively appealing one involves the notion of *parallel measures*, say X and X',

Table 3 Parameter values for the idealized population of four individuals depicted in Fig. 13.

Person k and Parameter		
k	T_k	Var [E_k]
1	310	100
2	300	81
3	330	121
4	280	64

$\mathrm{E}[\mathbf{T}] = 305$

$\mathrm{Var}[\mathbf{T}] = 325$

$\mathrm{Var}[\mathbf{E}] = \mathrm{E}[\mathrm{Var}[\mathbf{E_k}]] = 91.5$

$\mathrm{Var}[\mathbf{X}] = \mathrm{Var}[\mathbf{T}] + \mathrm{E}[\mathrm{Var}[\mathbf{E_k}]] = 416.5$

which share the fundamental property that the two tests assess the same true score and have the same observed score variance for every individual k. Specifically, if $\mathbf{X}_k = T_k + \mathbf{E}_k$ and $\mathbf{X}'_k = T_k + \mathbf{E}'_k$ are the two parallel measures, then the reliability of each one is the correlation between these two measures across individuals. In general, a convenient definition of the correlation between any two random variables $\mathbf{X}$ and $\mathbf{Y}$ is

$$\rho_{XY} \equiv \mathrm{Corr}[\mathbf{X}, \mathbf{Y}] \equiv \frac{\mathrm{Cov}[\mathbf{X}, \mathbf{Y}]}{\sqrt{\mathrm{Var}[\mathbf{X}] \cdot \mathrm{Var}[\mathbf{Y}]}}. \qquad (23)$$

With respect to two parallel measures, this reduces to

$$\rho_{XX'} \equiv \mathrm{Corr}[\mathbf{X}, \mathbf{X}'] \equiv \frac{\mathrm{Cov}[\mathbf{X}, \mathbf{X}']}{\mathrm{Var}[\mathbf{X}]} = \frac{\mathrm{Var}[\mathbf{T}]}{\mathrm{Var}[\mathbf{X}]} = 1 - \frac{\mathrm{Var}[\mathbf{E}]}{\mathrm{Var}[\mathbf{X}]}. \qquad (24)$$

This equation illustrates the important fact that, as a general rule, reliability tends to increase with increases in the variance of the true scores and with decreases in the error variance.

Reliability is an important topic within test theory because the correlation between any two measures is limited by the reliability of the measures (e.g., Lord & Novick, 1968). Specifically,

$$\rho_{XY} \leq \sqrt{\rho_{XX'}}, \qquad (25)$$

where $\mathbf{X}$ and $\mathbf{Y}$ are any two measures, and $\mathbf{X}'$ is a measure parallel to $\mathbf{X}$. This inequality indicates that the correlation between $\mathbf{X}$ and $\mathbf{Y}$ must always be smaller than the square root of the reliability of measure $\mathbf{X}$. For example, if the reliability of $\mathbf{X}$ is equal to .64, the correlation of $\mathbf{X}$ with an arbitrary variable $\mathbf{Y}$ cannot exceed the upper bound of $\sqrt{.64} = .8$. Therefore, it is important for a measure $\mathbf{X}$ to have a high reliability when assessing a potential correlation between $\mathbf{X}$ and $\mathbf{Y}$.

Differences of observed scores

Since RT research often focuses on effect sizes, the difference between two random variables is especially important within this context. Fortunately, classical test theory also provides information about the statistical properties of difference scores. For example, it is well-known that the reliability of the difference between two random variables, $\mathbf{D} = \mathbf{X} - \mathbf{Y}$, is notoriously low, as compared with the reliability of their sum, $\mathbf{S} = \mathbf{X} + \mathbf{Y}$. This statistical phenomenon arises because the true values that are associated with $\mathbf{X}$ and $\mathbf{Y}$ will tend to cancel each other out in the difference, $\mathbf{D} = \mathbf{X} - \mathbf{Y}$, at least to the extent that $\mathbf{X}$ and $\mathbf{Y}$ measure the same construct. In the extreme case of identical true score components, $\mathbf{D}$ reflects only the error components of $\mathbf{X}$ and $\mathbf{Y}$ and is thus completely unreliable. By contrast, the reliability of $\mathbf{S} = \mathbf{X} + \mathbf{Y}$ tends to be even higher than the reliability of $\mathbf{X}$ or $\mathbf{Y}$ because the true scores sum rather than canceling each other out.

The reliability of a difference score $\mathbf{D} = \mathbf{X} - \mathbf{Y}$ is generally defined as its correlation with the parallel measure $\mathbf{D}' = \mathbf{X}' - \mathbf{Y}'$

$$\text{Corr}[\mathbf{D}, \mathbf{D}'] = \frac{\text{Cov}[\mathbf{D}, \mathbf{D}']}{\sqrt{\sigma_\mathbf{D}^2 \cdot \sigma_{\mathbf{D}'}^2}}. \tag{26}$$

Assuming $\sigma_X = \sigma_{X'}$ and $\sigma_Y = \sigma_{Y'}$, this is equivalent to

$$\text{Corr}\left[\mathbf{D}, \mathbf{D}'\right] = \frac{\sigma_X^2 \cdot \rho_{XX'} - 2 \cdot \sigma_X \sigma_Y \cdot \rho_{XY} + \sigma_Y^2 \cdot \rho_{YY'}}{\sigma_X^2 + \sigma_Y^2 - 2 \cdot \sigma_X \sigma_Y \cdot \rho_{XY}}. \tag{27}$$

In classical test theory, the additional assumptions $\sigma_X = \sigma_Y$ and $\rho_{XX'} = \rho_{YY'}$ can sometimes be made, although there are cases in which these assumptions are clearly inappropriate (Williams & Zimmerman, 1996). Within RT research, it seems reasonable that both of these assumptions would be true to a good approximation, because differences are usually computed between mean RTs obtained using very similar conditions. With these two additional assumptions, Eq. 27 reduces to

$$\text{Corr}\left[\mathbf{D}, \mathbf{D}'\right] = \frac{\rho_{XX'} - \rho_{XY}}{1 - \rho_{XY}}. \tag{28}$$

Figure 14 illustrates the important consequence of Eq. 28 that the reliability of a difference score $\mathbf{D} = \mathbf{X} - \mathbf{Y}$ is inversely related to the correlation between the two components, $\mathbf{X}$ and $\mathbf{Y}$. It is clear from Eq. 28, for example, that the reliability of the difference score is zero when the correlation between the two components equals their common reliability. As can be seen in the main text, this inverse relation of difference score reliability and component correlation has important implications for the psychometric properties of RT difference scores.

Classical test theory also provides information about the correlation of a difference score with a third variable, sometimes referred to as the validity of the difference score in classical test theory. The correlation of the difference $\mathbf{D}$ with a third variable $\mathbf{Z}$ is

$$\text{Corr}[\mathbf{D}, \mathbf{Z}] = \frac{\text{Cov}[\mathbf{D}, \mathbf{Z}]}{\sqrt{\sigma_\mathbf{D}^2 \cdot \sigma_\mathbf{Z}^2}}. \tag{29}$$
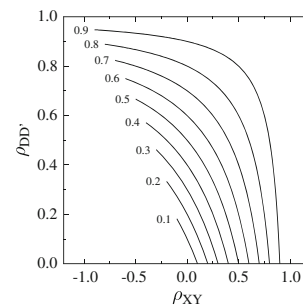
In general, this is

$$\text{Corr}[\mathbf{D}, \mathbf{Z}] = \frac{\sigma_X \cdot \rho_{XZ} - \sigma_Y \cdot \rho_{YZ}}{\sqrt{\sigma_X^2 + \sigma_Y^2 - 2\sigma_X \sigma_Y \rho_{XY}}}. \tag{30}$$

Under the simplifying assumption that $\sigma_X = \sigma_Y$, this reduces to

$$\text{Corr}[\mathbf{D}, \mathbf{Z}] = \frac{\rho_{XZ} - \rho_{YZ}}{\sqrt{2 \cdot (1 - \rho_{XY})}}. \tag{31}$$

Thus, the correlation of the difference score with the third variable depends not only on the correlations of that third

Fig. 14 The reliability, $\rho_{DD'}$, of a difference score, $\mathbf{D} = \mathbf{X} - \mathbf{Y}$, as a function of the correlation between $\mathbf{X}$ and $\mathbf{Y}$, ρ_{XY}. The curve parameter is the reliability of the scores contributing to the difference, $\rho_{XX'} = \rho_{YY'}$. Each curve is bounded on the left and right because $|\rho_{XY}| \leq |\rho_{XX'}| = |\rho_{YY'}|$ (e.g., Lord & Novick, 1968, Equation 3.9.6)

variable with the individual terms involved in computing the difference (i.e., with $\mathbf{X}$ and $\mathbf{Y}$), but also on the correlation of those individual terms with each other. Other things being equal, for example, $\text{Corr}[\mathbf{D}, \mathbf{Z}]$ increases when $\mathbf{X}$ and $\mathbf{Y}$ have a larger positive correlation.

Finally, RT researchers sometimes correlate the sizes of two separate difference scores obtained in different tasks, say $\mathbf{X} - \mathbf{Y}$ and $\mathbf{U} - \mathbf{V}$. Although the correlation of two difference scores is not usually considered in classical test theory, the same underlying model can be used to assess such correlations (e.g., Sriram et al., 2010). Specifically, the correlation of two differences $\mathbf{X} - \mathbf{Y}$ and $\mathbf{U} - \mathbf{V}$ is

$$\text{Corr}[\mathbf{X} - \mathbf{Y}, \mathbf{U} - \mathbf{V}] = \frac{\text{Cov}[\mathbf{X} - \mathbf{Y}, \mathbf{U} - \mathbf{V}]}{\sqrt{\sigma_{X-Y}^2 \cdot \sigma_{U-V}^2}}$$
$$= \frac{\sigma_X \sigma_U \cdot \rho_{XU} - \sigma_X \sigma_V \cdot \rho_{XV} - \sigma_Y \sigma_U \cdot \rho_{YU} + \sigma_Y \sigma_V \cdot \rho_{YV}}{\sqrt{\sigma_X^2 + \sigma_Y^2 - 2\sigma_X \sigma_Y \cdot \rho_{XY}} \cdot \sqrt{\sigma_U^2 + \sigma_V^2 - 2\sigma_U \sigma_V \cdot \rho_{UV}}}. \tag{32}$$

Two special cases of Eq. 32 are especially noteworthy. First, when $\sigma_X = \sigma_Y = \sigma_U = \sigma_V$, Eq. 32 reduces to

$$\text{Corr}[\mathbf{X} - \mathbf{Y}, \mathbf{U} - \mathbf{V}] = \frac{1}{2} \cdot \frac{\rho_{XU} - \rho_{XV} - \rho_{YU} + \rho_{YV}}{\sqrt{(1 - \rho_{XY}) \cdot (1 - \rho_{UV})}}. \tag{33}$$

If, in addition, the four pairwise correlations are equal to one another, $\rho_{XU} = \rho_{XV} = \rho_{YU} = \rho_{YV}$, then this correlation is necessarily zero.

Second, the use of a common term within the two differences (e.g., $\text{Corr}[\mathbf{X} - \mathbf{Z}, \mathbf{Y} - \mathbf{Z}]$) can lead to cases of "spurious correlation" (Pearson, 1896; see also Fuguitt & Lieberson, 1973). For example, in general

$$\text{Corr}[\mathbf{X} - \mathbf{Z}, \mathbf{Y} - \mathbf{Z}]$$
$$= \frac{\sigma_X \sigma_Y \cdot \rho_{XY} - \sigma_Y \sigma_Z \cdot \rho_{YZ} - \sigma_X \sigma_Z \cdot \rho_{XZ} + \sigma_Z^2}{\sqrt{\sigma_X^2 + \sigma_Z^2 - 2\sigma_X \sigma_Z \cdot \rho_{XZ}} \cdot \sqrt{\sigma_Y^2 + \sigma_Z^2 - 2\sigma_X \sigma_Z \cdot \rho_{XZ}}}. \tag{34}$$

Under the simplifying assumption that $\sigma_X = \sigma_Y = \sigma_Z$, this reduces to

$$\text{Corr}[\mathbf{X} - \mathbf{Z}, \mathbf{Y} - \mathbf{Z}] = \frac{1}{2} \cdot \frac{1 + \rho_{XY} - \rho_{XZ} - \rho_{YZ}}{\sqrt{(1 - \rho_{XZ})} \cdot \sqrt{(1 - \rho_{YZ})}}. \quad (35)$$

Note that the correlation of these difference scores is necessarily 1/2 when $\rho_{XY} = \rho_{XZ} = \rho_{YZ}$.

Appendix 2 Derivations of variances and covariances

This appendix presents general equations for the variances and covariances needed to compute the correlations examined in this article, as derived from the IDRT model (Eq. 3).

In all of the cases considered here, the required covariance can be computed using the well-known formula for the (co)variance of two linear combinations of random variables (e.g., Mood, Graybill, & Boes, 1974, p. 179). In general, let $\mathbf{X}_i$, $i = 1,\dots, m$ and $\mathbf{Y}_j$, $j = 1,\dots, n$ be two sets of random variables, possibly correlated. The covariance of two linear combinations of these random variables is

$$\text{Cov}\left[\sum_{i=1}^{m} a_i \mathbf{X}_i, \sum_{j=1}^{n} b_j \mathbf{Y}_j\right] = \sum_{i=1}^{m} \sum_{j=1}^{n} a_i b_j \text{Cov}[\mathbf{X}_i, \mathbf{Y}_j], \quad (36)$$

where a_i and b_j are sets of constants.

Each of the tables in this appendix illustrates this equation. The random variables $\mathbf{X}_i$ and $\mathbf{Y}_j$ correspond to the random variables within IDRT (e.g., $\mathbf{G}$, $\boldsymbol{\Delta}$, $\mathbf{E}$), and these define the rows and columns of the tables. Each cell within a table contains one value of $a_i b_j \text{Cov}[\mathbf{X}_i, \mathbf{Y}_j]$, and the overall covariance is simply the sum of all cells in the table. In all cases, we assume that there is zero covariance between the measurement error term (e.g., $\mathbf{E}$) and any other random variable, so the values in many cells are zero.

Table 5 Cross-product matrix for the covariance of **RT** and **RT′**

RVs in **RT′**	RVs in **RT**			
	G	**Δ**	**R**	**E**
G	$S^2\sigma_G^2$	$BS\sigma_\Delta\sigma_G\rho_{\Delta G}$	$S\sigma_R\sigma_G\rho_{GR}$	0
Δ	$SB\sigma_G\sigma_\Delta\rho_{\Delta G}$	$B^2\sigma_\Delta^2$	$B\sigma_R\sigma_\Delta\rho_{\Delta R}$	0
R	$S\sigma_G\sigma_R\rho_{GR}$	$B\sigma_\Delta\sigma_R\rho_{\Delta R}$	σ_R^2	0
E′	0	0	0	0

The required variances can also be computed as special cases of Eq. 36. In general, the variance of a single linear combination is

$$\text{Var}\left[\sum_{i=1}^{m} a_i \mathbf{X}_i\right] = \sum_{i=1}^{m} \sum_{j=1}^{m} a_i a_j \text{Cov}[\mathbf{X}_i, \mathbf{X}_j] \quad (37)$$

(e.g., Table 4).

Finally, the required correlations can be computed in all cases via Eq. 23, once the variances and covariance of the random variables have been determined.

Variance of **RT**

The variance of **RT** is the sum of the terms in the cross-product matrix shown in Table 4. This sum is

$$\text{Var}[\mathbf{RT}] = S^2\sigma_G^2 + 2BS\sigma_\Delta\sigma_G\rho_{\Delta G} + 2S\sigma_R\sigma_G\rho_{GR} + B^2\sigma_\Delta^2$$
$$+ 2B\sigma_R\sigma_\Delta\rho_{\Delta R} + \sigma_R^2 + \sigma_E^2. \quad (38)$$

As in classical test theory, σ_E^2 is the average error variance across individuals. We assumed that the standard error of the mean for individual k was proportional to that individual's mean RT, $\sigma_E = CV \cdot (S \cdot G_k + B \cdot \Delta_k + R_k)/\sqrt{N}$. Across individuals, then, the average error variance is

$$\sigma_E^2 = \frac{CV^2 \cdot (S \cdot \mu_G + B \cdot \mu_\Delta + \mu_R)^2}{N}. \quad (39)$$

Table 4 Cross-product matrix for the variance of **RT**

RVs in **RT**	RVs in **RT**			
	G	**Δ**	**R**	**E**
G	$S^2\sigma_G^2$	$BS\sigma_\Delta\sigma_G\rho_{\Delta G}$	$S\sigma_R\sigma_G\rho_{GR}$	0
Δ	$SB\sigma_G\sigma_\Delta\rho_{\Delta G}$	$B^2\sigma_\Delta^2$	$B\sigma_R\sigma_\Delta\rho_{\Delta R}$	0
R	$S\sigma_G\sigma_R\rho_{GR}$	$B\sigma_\Delta\sigma_R\rho_{\Delta R}$	σ_R^2	0
E	0	0	0	σ_E^2

Table 6 Cross-product matrix for the covariance of $\mathbf{RT_x}$ and $\mathbf{RT_y}$

RVs in $\mathbf{RT_y}$	RVs in $\mathbf{RT_x}$			
	G	$\boldsymbol{\Delta_x}$	$\mathbf{R_x}$	$\mathbf{E_x}$
G	$S_x S_y \sigma_G^2$	$B_x S_y \sigma_{\Delta_x}\sigma_G\rho_{\Delta_x G}$	$S_y\sigma_{R_x}\sigma_G\rho_{GR_x}$	0
$\boldsymbol{\Delta_y}$	$S_x B_y \sigma_G\sigma_{\Delta_y}\rho_{\Delta_y G}$	$B_x B_y \sigma_{\Delta_x}\sigma_{\Delta_y}\rho_{\Delta_x\Delta_y}$	$B_y\sigma_{R_x}\sigma_{\Delta_y}\rho_{\Delta_y R_x}$	0
$\mathbf{R_y}$	$S_x\sigma_G\sigma_{R_y}\rho_{GR_y}$	$B_x\sigma_{\Delta_x}\sigma_{R_y}\rho_{\Delta_x R_y}$	$\sigma_{R_x}\sigma_{R_y}\rho_{R_x R_y}$	0
$\mathbf{E_y}$	0	0	0	0

Table 7 Cross-product matrix for the variance of $\mathbf{D} = (B_e - B_c) \cdot \mathbf{G} + B_e \cdot \boldsymbol{\Delta_e} - B_c \cdot \boldsymbol{\Delta_c} + \mathbf{E_e} - \mathbf{E_c}$

RVs in	RVs in $\mathbf{D}$				
$\mathbf{D}$	$\mathbf{G}$	$\boldsymbol{\Delta_c}$	$\boldsymbol{\Delta_e}$	$\mathbf{E_c}$	$\mathbf{E_e}$
$\mathbf{G}$	$(B_e - B_c)^2 \sigma_G^2$	$(-B_c)(B_e - B_c)\sigma_{\Delta_c}\sigma_G\rho_{\Delta_c G}$	$B_e(B_e - B_c)\sigma_{\Delta_e}\sigma_G\rho_{\Delta_e G}$	0	0
$\boldsymbol{\Delta_c}$	$(B_e - B_c)(-B_c)\sigma_G\sigma_{\Delta_c}\rho_{\Delta_c G}$	$(-B_c)^2\sigma_{\Delta_c}^2$	$B_e(-B_c)\sigma_{\Delta_e}\sigma_{\Delta_c}\rho_{\Delta_c \Delta_e}$	0	0
$\boldsymbol{\Delta_e}$	$(B_e - B_c)B_e\sigma_G\sigma_{\Delta_e}\rho_{\Delta_e G}$	$(-B_c)B_e\sigma_{\Delta_c}\sigma_{\Delta_e}\rho_{\Delta_c \Delta_e}$	$B_e^2\sigma_{\Delta_e}^2$	0	0
$\mathbf{E_c}$	0	0	0	$\sigma_{E_c}^2$	0
$\mathbf{E_e}$	0	0	0	0	$\sigma_{E_e}^2$

Note. For the cost–benefit analysis using an alternative difference score definition to index facilitation—namely, $\mathbf{D_f} = (B_n - B_f) \cdot \mathbf{G} + B_n \cdot \boldsymbol{\Delta_n} + B_f \cdot \boldsymbol{\Delta_f} + \mathbf{E_n} - \mathbf{E_f}$ —the term $(-B_c)$ is replaced by $(+B_f)$ within all entries of this matrix.

Correlation of $\mathbf{RT}$ and $\mathbf{RT'}$

Table 5 shows the cross-product matrix for two independent measures of the same mean, $\mathbf{RT}$ and $\mathbf{RT'}$, as is needed to compute their correlation (i.e., the reliability of the mean RT measure). The covariance is the sum of terms in this cross-product matrix, which is

$$\text{Cov}\left[\mathbf{RT}, \mathbf{RT}'\right] = S^2\sigma_G^2 + 2BS\sigma_\Delta\sigma_G\rho_{\Delta G} + 2S\sigma_R\sigma_G\rho_{GR}$$
$$+ 2B\sigma_R\sigma_\Delta\rho_{\Delta R} + B^2\sigma_\Delta^2 + \sigma_R^2. \quad (40)$$

$\text{Corr}\left[\mathbf{RT}, \mathbf{RT}'\right]$ can be computed via Eq. 23 from this covariance together with the common variance (Eq. 38).

Correlation of $\mathbf{RT}$ and $\mathbf{Y}$

Within the model, the covariance of these two variables is

$$\text{Cov}[\mathbf{RT}, \mathbf{Y}] = \text{Cov}[S \cdot \mathbf{G} + B \cdot \boldsymbol{\Delta} + \mathbf{R} + \mathbf{E}, \mathbf{Y}]$$
$$= S \cdot \text{Cov}[\mathbf{G}, \mathbf{Y}] + B \cdot \text{Cov}[\boldsymbol{\Delta}, \mathbf{Y}] + \text{Cov}[\mathbf{R}, \mathbf{Y}]$$
$$= S \cdot \sigma_G \cdot \sigma_Y \cdot \rho_{GY} + B \cdot \sigma_\Delta \cdot \sigma_Y \cdot \rho_{\Delta Y} + \sigma_R \cdot \sigma_Y \cdot \rho_{RY}. \quad (41)$$

The correlation $\text{Corr}[\mathbf{RT}, \mathbf{Y}]$ can be computed via Eq. 23 from this covariance and the variances of RT (Eq. 38) and $\mathbf{Y}$ (σ_Y^2).

Correlation of $\mathbf{RT_x}$ and $\mathbf{RT_y}$

Table 6 shows the cross-product matrix for the covariance of $\mathbf{RT_x}$ and $\mathbf{RT_y}$. The covariance is the sum of these terms—namely,

$$\text{Cov}[\mathbf{RT_x}, \mathbf{RT_y}] = S_x S_y \sigma_G^2 + B_x B_y \sigma_{\Delta_x}\sigma_{\Delta_y}\rho_{\Delta_x \Delta_y}$$
$$+ B_x S_y \sigma_{\Delta_x}\sigma_G\rho_{\Delta_x G} + S_y\sigma_{R_x}\sigma_G\rho_{GR_x}$$
$$+ S_x B_y \sigma_G\sigma_{\Delta_y}\rho_{\Delta_y G} + B_y\sigma_{R_x}\sigma_{\Delta_y}\rho_{\Delta_y R_x}$$
$$+ S_x\sigma_G\sigma_{R_y}\rho_{GR_y} + B_x\sigma_{\Delta_x}\sigma_{R_y}\rho_{\Delta_x R_y}$$
$$+ \sigma_{R_x}\sigma_{R_y}\rho_{R_x R_y}. \quad (42)$$

The correlation of $\mathbf{RT_x}$ and $\mathbf{RT_y}$ can be computed via Eq. 23 from this covariance and the variance of each mean RT (Eq. 38).

Table 8 Cross-product matrix for the covariance of $\mathbf{D}$ and $\mathbf{D'}$

RVs in $\mathbf{D'}$	RVs in $\mathbf{D}$				
	$\mathbf{G}$	$\boldsymbol{\Delta_c}$	$\boldsymbol{\Delta_e}$	$\mathbf{E_c}$	$\mathbf{E_e}$
$\mathbf{G}$	$(B_e - B_c)^2 \sigma_G^2$	$(-B_c)(B_e - B_c)\sigma_{\Delta_c}\sigma_G\rho_{\Delta_c G}$	$B_e(B_e - B_c)\sigma_{\Delta_e}\sigma_G\rho_{\Delta_e G}$	0	0
$\boldsymbol{\Delta_e}$	$(B_e - B_c)B_e\sigma_G\sigma_{\Delta_e}\rho_{\Delta_e G}$	$(-B_c)B_e\sigma_{\Delta_c}\sigma_{\Delta_e}\rho_{\Delta_c \Delta_e}$	$B_e^2\sigma_{\Delta_e}^2$	0	0
$\boldsymbol{\Delta_c}$	$(B_e - B_c)(-B_c)\sigma_G\sigma_{\Delta_c}\rho_{\Delta_c G}$	$(-B_c)^2\sigma_{\Delta_c}^2$	$B_e(-B_c)\sigma_{\Delta_e}\sigma_{\Delta_c}\rho_{\Delta_c \Delta_e}$	0	0
$\mathbf{E_c'}$	0	0	0	0	0
$\mathbf{E_e'}$	0	0	0	0	0

Table 9 Cross-product matrix for the covariance of **D** and **Y**

RVs in **D**	**Y**
G	$(B_e - B_c)\sigma_Y\sigma_G\rho_{GY}$
$\mathbf{\Delta_c}$	$(-B_c)\sigma_Y\sigma_{\Delta_c}\rho_{\Delta_c Y}$
$\mathbf{\Delta_e}$	$B_e\sigma_Y\sigma_{\Delta_e}\rho_{\Delta_e Y}$
$\mathbf{E_c}$	0
$\mathbf{E_e}$	0

Variance of **D**

The variance of the RT difference score is needed to study its reliability and its correlation with other variables. Table 7 shows the cross-product matrix needed to evaluate the variance of $\mathbf{D} = (B_e - B_c) \cdot \mathbf{G} + B_e \cdot \mathbf{\Delta_e} - B_c \cdot \mathbf{\Delta_c} + \mathbf{E_e} - \mathbf{E_c}$. The variance is the sum of these terms,

$$\mathrm{Var}[\mathbf{D}] = (B_e - B_c)^2 \cdot \sigma_G^2 + B_e^2 \cdot \sigma_{\Delta_e}^2 + B_c^2 \cdot \sigma_{\Delta_c}^2 + \sigma_{E_e}^2$$
$$+ \sigma_{E_c}^2 + 2(B_e - B_c)B_e\sigma_G\sigma_{\Delta_e}\rho_{\Delta_e G}$$
$$- 2(B_e - B_c)B_c\sigma_G\sigma_{\Delta_c}\rho_{\Delta_c G} - 2B_eB_c\sigma_{\Delta_c}\sigma_{\Delta_e}\rho_{\Delta_c\Delta_e}.$$
$$(43)$$

For the cost–benefit case where facilitation is measured as $\mathbf{D_f} = (B_n - B_f) \cdot \mathbf{G} + B_n \cdot \mathbf{\Delta_n} + B_f \cdot \mathbf{\Delta_f} + \mathbf{E_n} - \mathbf{E_f}$, the variance of facilitation is

$$\mathrm{Var}[\mathbf{D_f}] = (B_e - B_c)^2 \cdot \sigma_G^2 + B_e^2 \cdot \sigma_{\Delta_e}^2 + B_c^2 \cdot \sigma_{\Delta_c}^2 + \sigma_{E_e}^2 + \sigma_{E_c}^2$$
$$+ 2(B_e - B_c)B_e\sigma_G\sigma_{\Delta_e}\rho_{\Delta_e G}$$
$$+ 2(B_e - B_c)B_c\sigma_G\sigma_{\Delta_c}\rho_{\Delta_c G} + 2B_eB_c\sigma_{\Delta_c}\sigma_{\Delta_e}\rho_{\Delta_c\Delta_e}.$$
$$(44)$$

Correlation of **D** and **D′**

Table 8 shows the cross-product matrix associated with the covariance of two independent estimates of an RT difference, **D** and **D′**. The covariance is the sum of the terms in this matrix, which is

$$\mathrm{Cov}\Big[\mathbf{D}, \mathbf{D}'\Big] = (B_e - B_c)^2\sigma_G^2 - 2B_c(B_e - B_c)\sigma_{\Delta_c}\sigma_G\rho_{\Delta_c G}$$
$$+ 2B_e(B_e - B_c)\sigma_{\Delta_e}\sigma_G\rho_{\Delta_e G}$$
$$- 2B_cB_e\sigma_{\Delta_c}\sigma_{\Delta_e}\rho_{\Delta_c\Delta_e} + B_c^2\sigma_{\Delta_c}^2 + B_e^2\sigma_{\Delta_e}^2.$$
$$(45)$$

The correlation $\mathrm{Corr}\big[\mathbf{D}, \mathbf{D}'\big]$ (i.e., difference score reliability) can be computed via Eq. 23 from this covariance and the variance of the difference score (Eq. 43).

Correlation of **D** and **Y**

Table 9 shows the cross-product matrix for the covariance of **D** and **Y**, and the sum of the terms in this matrix is

$$\mathrm{Cov}[\mathbf{D}, \mathbf{Y}] = (B_e - B_c) \cdot \sigma_Y \cdot \sigma_G \cdot \rho_{GY} + B_e \cdot \sigma_Y \cdot \sigma_{\Delta_e} \cdot \rho_{\Delta_e Y} - B_c \cdot \sigma_Y \cdot \sigma_{\Delta_c} \cdot \rho_{\Delta_c Y} \qquad (46)$$
$$= [(B_e - B_c) \cdot \sigma_G \cdot \rho_{GY} + B_e \cdot \sigma_{\Delta_e} \cdot \rho_{\Delta_e Y} - B_c \cdot \sigma_{\Delta_c} \cdot \rho_{\Delta_c Y}] \cdot \sigma_Y.$$

The correlation of **D** and **Y** is thus

$$\mathrm{Corr}[\mathbf{D}, \mathbf{Y}]$$
$$= \frac{(B_e - B_c) \cdot \sigma_G \cdot \rho_{GY} + B_e \cdot \sigma_{\Delta_e} \cdot \rho_{\Delta_e Y} - B_c \cdot \sigma_{\Delta_c} \cdot \rho_{\Delta_c Y}}{\mathrm{Var}[\mathbf{D}]}.$$
$$(47)$$

Correlation of $\mathbf{D_{ce}}$ and $\mathbf{D_{uv}}$

Table 10 shows the cross-product matrix for the covariance of $\mathbf{D_{ce}}$ and $\mathbf{D_{uv}}$. The covariance is the sum of its terms:

$$\mathrm{Cov}[\mathbf{D_{ce}}, \mathbf{D_{uv}}] = (B_e - B_c)(B_v - B_u)\sigma_G^2$$
$$- B_c(B_v - B_u)\sigma_{\Delta_c}\sigma_G\rho_{\Delta_c G}$$
$$+ B_e(B_v - B_u)\sigma_{\Delta_e}\sigma_G\rho_{\Delta_e G}$$
$$+ (B_e - B_c)B_v\sigma_G\sigma_{\Delta_v}\rho_{\Delta_v G}$$
$$- B_cB_v\sigma_{\Delta_c}\sigma_{\Delta_v}\rho_{\Delta_c\Delta_v} + B_eB_v\sigma_{\Delta_e}\sigma_{\Delta_v}\rho_{\Delta_e\Delta_v}$$
$$- (B_e - B_c)B_u\sigma_G\sigma_{\Delta_u}\rho_{\Delta_u G}$$
$$+ B_cB_u\sigma_{\Delta_c}\sigma_{\Delta_u}\rho_{\Delta_c\Delta_c} - B_eB_u\sigma_{\Delta_e}\sigma_{\Delta_u}\rho_{\Delta_e\Delta_u}.$$
$$(48)$$

Table 10 Cross-product matrix for the covariance of $\mathbf{D_{ce}}$ and $\mathbf{D_{uv}}$

RVs in $\mathbf{D_{uv}}$	RVs in $\mathbf{D_{ce}}$				
	G	$\boldsymbol{\Delta_c}$	$\boldsymbol{\Delta_e}$	$\mathbf{E_c}$	$\mathbf{E_e}$
G	$(B_e - B_c)(B_v - B_u)\sigma_G^2$	$(-B_c)(B_v - B_u)\sigma_{\Delta_c}\sigma_G\rho_{\Delta_c G}$	$B_e(B_v - B_u)\sigma_{\Delta_e}\sigma_G\rho_{\Delta_e G}$	0	0
$\boldsymbol{\Delta_v}$	$(B_e - B_c)B_v\sigma_G\sigma_{\Delta_v}\rho_{\Delta_v G}$	$(-B_c)B_v\sigma_{\Delta_c}\sigma_{\Delta_v}\rho_{\Delta_c \Delta_v}$	$B_e B_v\sigma_{\Delta_e}\sigma_{\Delta_v}\rho_{\Delta_e \Delta_v}$	0	0
$\boldsymbol{\Delta_u}$	$(B_e - B_c)(-B_u)\sigma_G\sigma_{\Delta_u}\rho_{\Delta_u G}$	$(-B_c)(-B_u)\sigma_{\Delta_c}\sigma_{\Delta_u}\rho_{\Delta_c \Delta_u}$	$B_e(-B_u)\sigma_{\Delta_e}\sigma_{\Delta_u}\rho_{\Delta_e \Delta_u}$	0	0
$\mathbf{E_v}$	0	0	0	0	0
$\mathbf{E_u}$	0	0	0	0	0

Corr[$\mathbf{D_{ce}}$, $\mathbf{D_{uv}}$] can be computed via Eq. 23 from this co-variance together with the variances of the two difference scores, each of which is given by Eq. 43.

Correlation of $\mathbf{D_i}$ and $\mathbf{D_f}$

Table 11 shows the cross-product matrix for the covariance of $\mathbf{D_i}$ and $\mathbf{D_f}$ for the cost–benefit situation. The covariance of $\mathbf{D_i}$ and $\mathbf{D_f}$ is the sum of the terms in this matrix:

$$
\begin{aligned}
\text{Cov}[\mathbf{D_f}, \mathbf{D_i}] = & (B_i - B_n)(B_n - B_f)\sigma_G^2 + B_i(B_n - B_f)\sigma_{\Delta_i}\sigma_G\rho_{\Delta_i G} \\
& - B_n(B_n - B_f)\sigma_{\Delta_n}\sigma_G\rho_{\Delta_n G} \\
& + (B_i - B_n)B_n\sigma_G\sigma_{\Delta_n}\rho_{\Delta_n G} + B_i B_n\sigma_{\Delta_i}\sigma_{\Delta_n}\rho_{\Delta_i \Delta_n} \\
& - B_n^2\sigma_{\Delta_n}^2 - (B_i - B_n)B_f\sigma_G\sigma_{\Delta_f}\rho_{\Delta_f G} \\
& - B_i B_f\sigma_{\Delta_i}\sigma_{\Delta_f}\rho_{\Delta_f \Delta_i} + B_n B_f\sigma_{\Delta_n}\sigma_{\Delta_f}\rho_{\Delta_f \Delta_n} - \sigma_{E_n}^2.
\end{aligned}
\tag{49}
$$

Corr[$\mathbf{D_f}$, $\mathbf{D_i}$] can be computed via Eq. 23 from this covariance together with the variances of $\mathbf{D_f}$ (Eq. 44) and $\mathbf{D_i}$ (Eq. 43).

Correlation of $\mathbf{RT_c}$ and $\mathbf{D}$

Table 12 shows the cross-product matrix for the covariance of $\mathbf{RT_c}$ and $\mathbf{D}$. The covariance of $\mathbf{RT_c}$ and $\mathbf{D}$ is the sum of these terms:

$$
\begin{aligned}
\text{Cov}[\mathbf{RT_c}, \mathbf{D}] = & S_c(B_e - B_c)\sigma_G^2 + B_c[(B_e - B_c) - S_c]\sigma_{\Delta_c}\sigma_G\rho_{\Delta_c G} \\
& + (B_e - B_c)\sigma_R\sigma_G\rho_{GR} - B_c^2\sigma_{\Delta_c}^2 - B_c\sigma_R\sigma_{\Delta_c}\rho_{\Delta_c R} \\
& + S_c B_e\sigma_G\sigma_{\Delta_e}\rho_{\Delta_e G} + B_c B_e\sigma_{\Delta_c}\sigma_{\Delta_e}\rho_{\Delta_c \Delta_e} \\
& + B_e\sigma_R\sigma_{\Delta_e}\rho_{\Delta_e R} - \sigma_{E_c}^2.
\end{aligned}
\tag{50}
$$

Corr[$\mathbf{RT_c}$, $\mathbf{D}$] can be computed via Eq. 23 from this covariance together with the variances of $\mathbf{RT_c}$ (Eq. 38) and $\mathbf{D}$ (Eq. 43).

Variance of $\overline{\mathbf{RT}}$

To write more compact expressions involving the average of the control and experimental condition mean RTs, we let $\overline{S} = A + (B_c + B_e)/2 + C$.

Table 13 shows the cross-product matrix for the variance of $\overline{\mathbf{RT}}$. The variance of the average is the sum of these terms—namely,

Table 11 Cross-product matrix for the covariance of $\mathbf{D_i}$ and $\mathbf{D_f}$

RVs in $\mathbf{D_f}$	RVs in $\mathbf{D_i}$				
	G	$\boldsymbol{\Delta_i}$	$\boldsymbol{\Delta_n}$	$\mathbf{E_n}$	$\mathbf{E_i}$
G	$(B_i - B_n)(B_n - B_f)\sigma_G^2$	$B_i(B_n - B_f)\sigma_{\Delta_i}\sigma_G\rho_{\Delta_i G}$	$(-B_n)(B_n - B_f)\sigma_{\Delta_n}\sigma_G\rho_{\Delta_n G}$	0	0
$\boldsymbol{\Delta_n}$	$(B_i - B_n)B_n\sigma_G\sigma_{\Delta_n}\rho_{\Delta_n G}$	$B_i B_n\sigma_{\Delta_i}\sigma_{\Delta_n}\rho_{\Delta_i \Delta_n}$	$(-B_n)B_n\sigma_{\Delta_n}^2$	0	0
$\boldsymbol{\Delta_f}$	$(B_i - B_n)B_f\sigma_G\sigma_{\Delta_f}\rho_{\Delta_f G}$	$B_i B_f\sigma_{\Delta_i}\sigma_{\Delta_f}\rho_{\Delta_f \Delta_i}$	$(-B_n)B_f\sigma_{\Delta_n}\sigma_{\Delta_f}\rho_{\Delta_f \Delta_n}$	0	0
$\mathbf{E_f}$	0	0	0	0	0
$\mathbf{E_n}$	0	0	0	$-\sigma_{E_n}^2$	0

Table 12 Cross-product matrix for the covariance of $\mathbf{RT_c}$ and $\mathbf{D}$

RVs in $\mathbf{D}$	RVs in $\mathbf{RT_c}$			
	$\mathbf{G}$	$\mathbf{\Delta_c}$	$\mathbf{R}$	$\mathbf{E_c}$
$\mathbf{G}$	$S_c(B_e - B_c)\sigma_G^2$	$B_c(B_e - B_c)\sigma_{\Delta_c}\sigma_G\rho_{\Delta_c G}$	$(B_e - B_c)\sigma_R\sigma_G\rho_{GR}$	0
$\mathbf{\Delta_c}$	$S_c(-B_c)\sigma_G\sigma_{\Delta_c}\rho_{\Delta_c G}$	$B_c(-B_c)\sigma_{\Delta_c}^2$	$(-B_c)\sigma_R\sigma_{\Delta_c}\rho_{\Delta_c R}$	0
$\mathbf{\Delta_e}$	$S_c B_e\sigma_G\sigma_{\Delta_e}\rho_{\Delta_e G}$	$B_c B_e\sigma_{\Delta_c}\sigma_{\Delta_e}\rho_{\Delta_c\Delta_e}$	$B_e\sigma_R\sigma_{\Delta_e}\rho_{\Delta_e R}$	0
$\mathbf{E_c}$	0	0	0	$-\sigma_{E_c}^2$
$\mathbf{E_e}$	0	0	0	0

$$\sigma_{\overline{RT}}^2 = \overline{S}^2\sigma_G^2 + B_c\overline{S}\sigma_{\Delta_c}\sigma_G\rho_{\Delta_c G} + B_e\overline{S}\sigma_{\Delta_e}\sigma_G\rho_{\Delta_e G} + 2\overline{S}\sigma_R\sigma_G\rho_{GR}$$
$$+ B_cB_e/2\cdot\sigma_{\Delta_c}\sigma_{\Delta_e}\rho_{\Delta_c\Delta_e} + B_c\sigma_{\Delta_c}\sigma_R\rho_{\Delta_c R} + B_e\sigma_{\Delta_e}\sigma_R\rho_{\Delta_e R}$$
$$+ (B_c/2)^2\sigma_{\Delta_c}^2 + (B_e/2)^2\sigma_{\Delta_e}^2 + \sigma_R^2 + 0.5^2\sigma_{E_c}^2 + 0.5^2\sigma_{E_e}^2. \tag{51}$$

Correlation of $\overline{\mathbf{RT}}$ and $\mathbf{D}$

Table 14 shows the cross-product matrix for the covariance of $\overline{\mathbf{RT}}$ and $\mathbf{D}$. Their covariance is the sum of the terms in this cross-product matrix, which is

$$\text{Cov}\left[\overline{\mathbf{RT}}, \mathbf{D}\right] = (B_e - B_c)\overline{S}\sigma_G^2 - B_c\overline{S}\sigma_{\Delta_c}\sigma_G\rho_{\Delta_c G}$$
$$+ B_e\overline{S}\sigma_{\Delta_e}\sigma_G\rho_{\Delta_e G} + 0.5(B_e - B_c)B_c\sigma_G\sigma_{\Delta_c}\rho_{\Delta_c G}$$
$$- 0.5B_c^2\sigma_{\Delta_c}^2 + 0.5B_eB_c\sigma_{\Delta_e}\sigma_{\Delta_c}\rho_{\Delta_c\Delta_e}$$
$$+ 0.5(B_e - B_c)B_e\sigma_G\sigma_{\Delta_e}\rho_{\Delta_e G}$$
$$- 0.5B_cB_e\sigma_{\Delta_c}\sigma_{\Delta_e}\rho_{\Delta_c\Delta_e} + 0.5B_e^2\sigma_{\Delta_e}^2$$
$$+ (B_e - B_c)\sigma_G\sigma_R\rho_{GR} - B_c\sigma_{\Delta_c}\sigma_R\rho_{\Delta_c R}$$
$$+ B_e\sigma_{\Delta_e}\sigma_R\rho_{\Delta_e R} - 0.5\sigma_{E_c}^2 + 0.5\sigma_{E_e}^2. \tag{52}$$

$\text{Corr}\left[\overline{\mathbf{RT}}, \mathbf{D}\right]$ can be computed via Eq. 23 from this covariance together with the variances of $\overline{\mathbf{RT}}$ (Eq. 51) and $\mathbf{D}$ (Eq. 43).

Variance of $\mathbf{\Delta}^*$ and Correlation of $\mathbf{Y}$ and $\mathbf{\Delta}^*$

As is considered in the General Discussion, researchers might be interested in estimating the correlation between an external measure $\mathbf{Y}$ and the difference in task-specific processing times between control and experimental conditions, $\mathbf{\Delta}^* = B_e\mathbf{\Delta_e} - B_c\mathbf{\Delta_c}$. The covariance of these measures is

$$\text{Cov}\left[\mathbf{\Delta}^*, \mathbf{Y}\right] = B_e\sigma_Y\sigma_{\Delta_e}\rho_{\Delta_e Y} - B_c\sigma_Y\sigma_{\Delta_c}\rho_{\Delta_c Y}. \tag{53}$$

In general, the correlation $\rho_{\mathbf{\Delta}^*\mathbf{Y}}$ can be computed via Eq. 23 from this covariance, the variance σ_Y^2, and the variance of $\mathbf{\Delta}^*$, which is

$$\sigma_{\mathbf{\Delta}^*}^2 = B_e^2\sigma_{\Delta_e}^2 - 2B_cB_e\sigma_{\Delta_c}\sigma_{\Delta_e}\rho_{\Delta_c\Delta_e} + B_c^2\sigma_{\Delta_c}^2. \tag{54}$$

Assuming that $\sigma_{\Delta_c}^2 = \sigma_{\Delta_e}^2$, the correlation is

$$\rho_{\mathbf{\Delta}^*\mathbf{Y}} = \frac{B_e\rho_{\Delta_e Y} - B_c\rho_{\Delta_c Y}}{\sqrt{B_e^2 - 2B_cB_e\rho_{\Delta_c\Delta_e} + B_c^2}}. \tag{55}$$

As is depicted in Fig. 12, the observable correlation of Corr[$\mathbf{D}$, $\mathbf{Y}$] is only imperfectly related to the correlation of $\mathbf{\Delta}^*$ and $\mathbf{Y}$. The observable correlation also increases, for example, with the correlation of $\mathbf{G}$ and $\mathbf{Y}$, as can be seen by comparing the three rows of the figure. Within each panel, the exact relation between the observable correlation on the vertical axis and the theoretical one on the horizontal axis

Table 13 Cross-product matrix for the variance of $\overline{\mathbf{RT}}$

RVs in $\overline{\mathbf{RT}}$	RVs in $\overline{\mathbf{RT}}$					
	$\mathbf{G}$	$\mathbf{\Delta_c}$	$\mathbf{\Delta_e}$	$\mathbf{R}$	$\mathbf{E_c}$	$\mathbf{E_e}$
$\mathbf{G}$	$\overline{S}^2\sigma_G^2$	$0.5B_c\overline{S}\sigma_{\Delta_c}\sigma_G\rho_{\Delta_c G}$	$0.5B_e\overline{S}\sigma_{\Delta_e}\sigma_G\rho_{\Delta_e G}$	$\overline{S}\sigma_R\sigma_G\rho_{GR}$	0	0
$\mathbf{\Delta_c}$	$0.5\overline{S}B_c\sigma_G\sigma_{\Delta_c}\rho_{\Delta_c G}$	$0.25B_c^2\sigma_{\Delta_c}^2$	$0.25B_eB_c\sigma_{\Delta_e}\sigma_{\Delta_c}\rho_{\Delta_c\Delta_e}$	$0.5B_c\sigma_R\sigma_{\Delta_c}\rho_{\Delta_c R}$	0	0
$\mathbf{\Delta_e}$	$0.5\overline{S}B_e\sigma_G\sigma_{\Delta_e}\rho_{\Delta_e G}$	$0.25B_cB_e\sigma_{\Delta_c}\sigma_{\Delta_e}\rho_{\Delta_c\Delta_e}$	$0.25B_e^2\sigma_{\Delta_e}^2$	$0.5B_e\sigma_R\sigma_{\Delta_e}\rho_{\Delta_e R}$	0	0
$\mathbf{R}$	$\overline{S}\sigma_G\sigma_R\rho_{GR}$	$0.5B_c\sigma_{\Delta_c}\sigma_R\rho_{\Delta_c R}$	$0.5B_e\sigma_{\Delta_e}\sigma_R\rho_{\Delta_e R}$	σ_R^2	0	0
$\mathbf{E_c}$	0	0	0	0	$0.25\sigma_{E_c}^2$	0
$\mathbf{E_e}$	0	0	0	0	0	$0.25\sigma_{E_e}^2$

Table 14 Cross-product matrix for the covariance of $\overline{\mathbf{RT}}$ and $\mathbf{D}$

RVs in $\overline{\mathbf{RT}}$	RVs in $\mathbf{D}$				
	G	Δ_c	Δ_e	E_c	E_e
G	$(B_e - B_c)\overline{S}\sigma_G^2$	$-B_c\overline{S}\sigma_{\Delta_c}\sigma_G\rho_{\Delta_c G}$	$B_e\overline{S}\sigma_{\Delta_e}\sigma_G\rho_{\Delta_e G}$	0	0
Δ_c	$0.5(B_e - B_c)B_c\sigma_G\sigma_{\Delta_c}\rho_{\Delta_c G}$	$-0.5B_cB_c\sigma_{\Delta_c}^2$	$0.5B_eB_c\sigma_{\Delta_e}\sigma_{\Delta_c}\rho_{\Delta_c\Delta_e}$	0	0
Δ_e	$0.5(B_e - B_c)B_e\sigma_G\sigma_{\Delta_e}\rho_{\Delta_e G}$	$-0.5B_cB_e\sigma_{\Delta_c}\sigma_{\Delta_e}\rho_{\Delta_c\Delta_e}$	$0.5B_e^2\sigma_{\Delta_e}^2$	0	0
R	$(B_e - B_c)\sigma_G\sigma_R\rho_{GR}$	$-B_c\sigma_{\Delta_c}\sigma_R\rho_{\Delta_c R}$	$B_e\sigma_{\Delta_e}\sigma_R\rho_{\Delta_e R}$	0	0
E_c	0	0	0	$-0.5\sigma_{E_c}^2$	0
E_e	0	0	0	0	$0.5\sigma_{E_e}^2$

can also be seen to depend on the correlation of the task-specific processing times in the experimental and control conditions, $\rho_{\Delta_c \Delta_e}$.

Appendix 3 Assumptions and parameter values used to illustrate predictions of IDRT

In general, we made three types of simplifying assumptions in deriving predictions from IDRT, although not every assumption is needed for every analysis. First, we assumed that the residual term $\mathbf{R}$ was not correlated either with the general processing time $\mathbf{G}$ or with any task-specific processing time $\mathbf{\Delta}$. This fits with our conceptualization of the residual term $\mathbf{R}$ as indexing peripheral sensory and motor times, whereas $\mathbf{G}$ and $\mathbf{\Delta}$ reflect more central or cognitive processing time. Second, we assumed equal variances of the task-specific processing times, $\mathbf{\Delta_i}$, for each task i contributing to a given correlation comparison. Our impression is that researchers generally employ tasks that are quite similar to one another, relative to the space of all imaginable cognitive tasks, so all of the tasks would have approximately the same person-to-person variation in task-specific processing time. Third, and for exactly the same reason, we assumed equal correlations of general and task-specific processing times, $\rho_{\Delta_i G}$, across all of the different tasks.

To provide numerical illustrations of the predicted reliabilities and correlations, it was necessary to choose specific parameter values. Although the exact choices were necessarily somewhat arbitrary and each study will have its own values, informal checks with various choices show that all of the qualitative conclusions drawn in the present study hold for any reasonable values. To provide specific illustrations, we chose a set of baseline parameters that would approximately reproduce the data reported by H. D. Kane, Proctor and Kranzler (1997, Table 1). This data set was chosen because the study included several typical conditions whose results agreed well with one another and because the authors reported all of the key characteristics of the data needed to constrain the model (i.e., mean RT, average across participants of the

within-participant standard deviation of RT, and standard deviation across participants of the individual mean RTs). Starting from the baseline parameters derived from H. D. Kane et al.'s study as is discussed next, the effects of individual parameters on reliabilities and correlations could be assessed by introducing changes to the values of these parameters. Table 15 summarizes the baseline parameter values used for all of the computations presented in the figures.

First, we used a baseline overall mean RT of 600 ms, which was the approximate observed value in several of H. D. Kane et al.'s (1997) conditions and which seems to be a typical value for many RT tasks. Of this total mean RT, an average of 200 ms was allocated to the residual term (i.e., peripheral sensory and motor processes; $\mu_R = 200$ ms), and a

Table 15 Default values of parameters

Value	Parameter(s)
100	$A, A_c, A_e, A_u, A_v, A_x,$ and A_y
200	$B, B_b, B_c, B_e, B_n, B_u, B_v, B_x,$ and B_y
100	$C, C_c, C_e, C_u, C_v, C_x,$ and C_y
0.14	CV
0	$\mu_\Delta, \mu_{\Delta_b}, \mu_{\Delta_c}, \mu_{\Delta_e}, \mu_{\Delta_n}, \mu_{\Delta_u}, \mu_{\Delta_v}, \mu_{\Delta_x},$ and μ_{Δ_y}
1	μ_G
200	$\mu_R, \mu_{R_c}, \mu_{R_e}, \mu_{R_u}, \mu_{R_v}, \mu_{R_x},$ and μ_{R_y}
50 and 100	N (number of trials per condition) for RT means and difference scores, respectively
0.2	$\rho_{\Delta G}, \rho_{\Delta_b G}, \rho_{\Delta_c G}, \rho_{\Delta_e G}, \rho_{\Delta_n G}, \rho_{\Delta_u G},$ and $\rho_{\Delta_v G}$
0	$\rho_{\Delta R}, \rho_{\Delta_c R}, \rho_{\Delta_e R}, \rho_{\Delta_n R}, \rho_{\Delta_u R}, \rho_{\Delta_v R}, \rho_{\Delta_x R_y},$ $\rho_{\Delta_y R_x},$ and $\rho_{\Delta_y R_y}$
0	$\rho_{GR}, \rho_{GR_{ce}}, \rho_{GR_{uv}}, \rho_{GR_x},$ and ρ_{GR_y}
0	ρ_{RY}
0.10	$\sigma_\Delta, \sigma_{\Delta_b}, \sigma_{\Delta_c}, \sigma_{\Delta_e}, \sigma_{\Delta_n}, \sigma_{\Delta_u},$ and σ_{Δ_v}
0.20	σ_G
20	$\sigma_R, \sigma_{R_{ce}}, \sigma_{R_{uv}}, \sigma_{R_x},$ and σ_{R_y}
10	σ_Y

Note. Except as noted otherwise in the figure captions or legends, these parameter values were used in computing the reliabilities and correlations displayed in Figs. 1–11.

further 100 ms each was allocated to the sensory and motor processes whose durations would be influenced by **G** (i.e., stages A and C, respectively). Various psychophysiological evidence suggests that totals of approximately 200 ms each would be realistic values for the total durations of the sensory and motor stages (e.g., Carbonnell, Hasbroucq, Grapperon, & Vidal, 2004; Hillyard & Anllo-Vento, 1998; Ritter, Simson, & Vaughan, 1983). The remaining 200 ms was allocated to the central stage B, which also seems plausible for straightforward choice RT tasks with arbitrary stimulus–response mappings. Equation 5 makes it clear that these desired stage durations can be produced with many different combinations of parameter values. Somewhat arbitrarily, we chose $A = 100$, $B = 200$, $C = 100$, and $\mu_G = 1$, so that A, B, and C could be regarded as the number of milliseconds required by each stage for a participant with the average **G** value of $\mu_G = 1$. Correspondingly, μ_Δ was set to zero, so values of **Δ** are increments and decrements that average to zero across individuals.

Second, concerning the trial-to-trial variability of RT that determines σ_E^2 (Eq. 2), we assumed that the standard deviation of RT for a given participant would be $CV = 0.14$ times the participant's mean RT. The value of $CV = 0.14$ was chosen to give an average within-participant standard deviation of RT equal to 85 ms with the mean of 600 ms, because 85 ms was approximately the average standard deviation reported by H. D. Kane et al. (1997).

Third, we chose the remaining baseline parameter values to produce a standard deviation of observed mean RTs (i.e., across individuals) of approximately 90 ms when means were based on 30 trials per participant. H. D. Kane et al. (1997) reported the standard deviations of observed mean RTs for several conditions with different numbers of trials, and adjustment of their reported values to a common value of 30 trials suggests that 90 ms would be approximately correct for all conditions. As is developed in Appendix 2 (e.g., Eq. 38), the standard deviation predicted by the model depends on the variances and intercorrelations of the processing time parameters **G**, **Δ**, and **R**, so again it was possible to produce the desired value (90 ms) in various ways. It seemed plausible to us that the residual and central processing times would be, at most, very weakly correlated, so we chose $\rho_{GR} = \rho_{\Delta R} = 0$. A combination of the remaining parameters that seemed plausible and produced

approximately the desired 90-ms result was $\sigma_G = 0.2$, $\sigma_\Delta = 0.1$, $\sigma_R = 20$ ms, and $\rho_{\Delta G} = 0.2$. In addition, these choices also yielded reliability of individual-participant observed mean RTs to be approximately .95 when based on averages of 30 trials, consistent with the report of Jensen (1985) that mean RT reliability is typically quite high with 30 trials under relatively standard conditions.

Appendix 4 Brinley plots predicted by IDRT

As a preliminary check on the ability of this model to describe individual differences in RT, we investigated its predictions with respect to Brinley plots, which are known to be highly linear, to have slopes greater than one, and to have negative intercepts. Despite some dispute over the exact meaning of Brinley plots, the ability to produce realistic plots is widely regarded as a critical benchmark for models of individual differences in RT (e.g., Bashore, 1994; Cerella, 1994; Faust, Balota, Spieler, & Ferraro, 1999; Fisk & Fisher, 1994; Myerson, Adams, Hale, & Jenkins, 2003; Ratcliff, Spieler, & McKoon, 2000; Van Ravenzwaaij, Brown, & Wagenmakers, 2011).

It is convenient to investigate the form of the Brinley plot predicted by the model summarized in Eq. 3 within the context of a specific example. Specifically, we will consider the relationship between the true mean RTs for young and old participants across a series of tasks in which only the duration of the central stage B varies. The true mean **RT_o** and **RT_y** for the old and young groups can be written as functions of B:

$$\mathrm{E}[\mathbf{RT}_o | B] = (A + C) \cdot \mu_{Go} + (\mu_{Go} + \mu_{\Delta_o}) \cdot B + \mu_{R_o} \quad (56)$$

and

$$\mathrm{E}[\mathbf{RT}_y | B] = (A + C) \cdot \mu_{Gy} + \left(\mu_{Gy} + \mu_{\Delta_y}\right) \cdot B + \mu_{R_y}. \quad (57)$$

Equation 57 can be rearranged to form

$$B = \frac{\mathrm{E}[\mathbf{RT}_y | B] - (A + C) \cdot \mu_{Gy} - \mu_{R_y}}{\mu_{Gy} + \mu_{\Delta_y}}, \quad (58)$$

which can then be substituted into Eq. 56 to obtain

$$\mathrm{E}[\mathbf{RT}_o | B] = (A + C) \cdot \mu_{Go} + \mu_{R_o} + (\mu_{Go} + \mu_{\Delta_o}) \cdot \frac{\mathrm{E}[\mathbf{RT}_y|B] - (A+C) \cdot \mu_{Gy} - \mu_{R_y}}{\mu_{Gy} + \mu_{\Delta_y}} \quad (59)$$

$$= (A + C) \cdot \mu_{Go} + \mu_{R_o} + \frac{\mu_{Go} + \mu_{\Delta_o}}{\mu_{Gy} + \mu_{\Delta_y}} \cdot \{\mathrm{E}[\mathbf{RT}_y|B] - (A+C) \cdot \mu_{Gy} - \mu_{R_y}\}.$$

Equation 59 shows that IDRT does predict linear Brinley plots, consistent with the standard empirical results. Specifically, as the duration of the central stage B varies, the slope of the function relating $E[\mathbf{RT}_o]$ to $E[\mathbf{RT}_y]$ is $\frac{\mu_{Go} + \mu_{\Delta_o}}{\mu_{Gy} + \mu_{\Delta_y}}$, so the slope will be greater than one (as is normally found) as long as the older participants have longer general processing times ($\mu_{Go} > \mu_{Gy}$). Furthermore, the intercept of this function is

$$\alpha = (A + C) \cdot \mu_{Go} + \mu_{R_o} - \frac{\mu_{Go} + \mu_{\Delta_o}}{\mu_{Gy} + \mu_{\Delta_y}} \cdot \left[(A + C) \cdot \mu_{Gy} + \mu_{R_y} \right]. \tag{60}$$

This intercept will be negative, as is normally found, when

$$(A + C) \cdot \mu_{Go} + \mu_{R_o} < \frac{\mu_{Go} + \mu_{\Delta_o}}{\mu_{Gy} + \mu_{\Delta_y}} \cdot \left[(A + C) \cdot \mu_{Gy} + \mu_{R_y} \right]. \tag{61}$$

For example, it can be shown that the intercept is negative if all of the parameters for older adults are proportional to those of the younger adults (i.e., $\mu_{Go} = c \cdot \mu_{Gy}$, $\mu_{\Delta_o} = c \cdot \mu_{\Delta_y}$, and $\mu_{R_o} = c \cdot \mu_{R_y}$, with $c > 1$). As another example, the intercept is negative if $\mu_{\Delta_o} = \mu_{\Delta_y} = 0$ and the age effect is proportionally larger on μ_G than on μ_R (i.e., $\mu_{Go} / \mu_{Gy} > \mu_{R_o} / \mu_{R_y}$).

References

Barutchu, A., Crewther, S. G., Fifer, J., Shivdasani, M. N., Innes-Brown, H., Toohey, S., … Paolini, A. G. (2011). The relationship between multisensory integration and IQ in children. *Developmental Psychology, 47*(3), 877–885. doi:10.1037/a0021903

Bashore, T. R. (1994). Some thoughts on neurocognitive slowing. *Acta Psychologica, 86*(2–3), 295–325. doi:10.1016/0001-6918(94)90006-X

Beauducel, A., & Brocke, B. (1993). Intelligence and speed of information processing: Further results and questions on Hick's paradigm and beyond. *Personality and Individual Differences, 15*, 627–636. doi:10.1016/0191-8869(93)90004-M

Blanton, H., Jaccard, J., Gonzales, P. M., & Christie, C. (2006). Decoding the implicit association test: Implications for criterion prediction. *Journal of Experimental Social Psychology, 42*, 192–212. doi:10.1016/j.jesp.2005.07.003

Brown, T. L. (2011). The relationship between Stroop interference and facilitation effects: Statistical artifacts, baselines, and a reassessment. *Journal of Experimental Psychology. Human Perception and Performance, 37*(1), 85–99. doi:10.1037/a0019252

Brown, S., & Heathcote, A. (2005). A ballistic model of choice response time. *Psychological Review, 112*, 117–128. doi:10.1037/0033-295X.112.1.117

Carbonnell, L., Hasbroucq, T., Grapperon, J., & Vidal, F. (2004). Response selection and motor areas: A behavioural and electrophysiological study. *Clinical Neurophysiology, 115*(9), 2164–2174. doi:10.1016/j.clinph.2004.04.012

Cattell, J. McK. (1890). Mental tests and measurements. *Mind, 15*, 373–380. doi:10.2307/1411718

Cerella, J. (1985). Information processing rates in the elderly. *Psychological Bulletin, 98*(1), 67–83. doi:10.1037/0033-2909.98.1.67

Cerella, J. (1994). Generalized slowing in Brinley plots. *Journals of Gerontology, 49*(2), 65–71.

Chapman, L. J., Chapman, J. P., Curran, T. E., & Miller, M. B. (1994). Do children and the elderly show heightened semantic priming? How to answer the question. *Developmental Review, 14*(2), 159–185. doi:10.1006/drev.1994.1007

Chen, J., Myerson, J., Hale, S., & Simon, A. (2000). Behavioral evidence for brain-based ability factors in visuospatial information processing. *Neuropsychologia, 38*(4), 380–387. doi:10.1016/S0028-3932(99)00095-0

Cherbuin, N., & Brinkman, C. (2006). Efficiency of callosal transfer and hemispheric interaction. *Neuropsychology, 20*, 178–184. doi:10.1037/0894-4105.20.2.178

Chiang, A., & Atkinson, R. C. (1976). Individual differences and interrelationships among a select set of cognitive skills. *Memory & Cognition, 4*, 661–672. doi:10.3758/BF03213232

Cooper, L. A., & Podgorny, P. (1976). Mental transformations and visual comparison processes: Effects of complexity and similarity. *Journal of Experimental Psychology. Human Perception and Performance, 2*, 503–514. doi:10.1037//0096-1523.2.4.503

Corballis, M. C. (2002). Hemispheric interactions in simple reaction time. *Neuropsychologia, 40*, 423–434. doi:10.1016/S0028-3932(01)00097-5

Davis, R. (1959). The role of "attention" in the psychological refractory period. *Quarterly Journal of Experimental Psychology, 11*, 211–220. doi:10.1080/17470215908416313

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology. General, 122*, 371–396. doi:10.1037//0096-3445.122.3.371

Der, G., & Deary, I. J. (2006). Age and sex differences in reaction time in adulthood: Results from the United Kingdom health and lifestyle survey. *Psychology and Aging, 21*, 62–73. doi:10.1037/0882-7974.21.1.62

Dickman, S. J., & Meyer, D. E. (1988). Impulsivity and speed-accuracy tradeoffs in information processing. *Journal of Personality and Social Psychology, 54*, 274–290. doi:10.1037//0022-3514.54.2.274

Donders, F. C. (1868/1969). Over de snelheid van psychische processen. [On the speed of mental processes.] (W. G. Koster, Trans.). In W. G. Koster (Ed.), *Attention and performance II* (pp. 412–431). Amsterdam: North Holland. doi:10.1016/0001-6918(69)90065-1

Eaton, W. O., & Ritchot, K. F. M. (1995). Physical maturation and information-processing speed in middle childhood. *Developmental Psychology, 31*, 967–972. doi:10.1037//0012-1649.31.6.967

Eckert, M. A., Keren, N. I., Roberts, D. R., Calhoun, V. D., & Harris, K. C. (2010). Age-related changes in processing speed: Unique contributions of cerebellar and prefrontal cortex. *Frontiers in Human Neuroscience, 4*(10), 1–14. doi:10.3389/neuro.09.010.2010

Ellermeier, W., Eigenstetter, M., & Zimmer, K. (2001). Psychoacoustic correlates of individual noise sensitivity. *Journal of the Acoustical Society of America, 109*, 1464–1473. doi:10.1121/1.1350402

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*, 143–149. doi:10.3758/BF03203267

Everitt, B. S., & Hand, B. J. (1981). *Finite mixture distributions*. London: Chapman & Hall.

Exposito, J., & Andres-Pueyo, A. (1997). The effects of impulsivity on the perceptual and decision stages in a choice reaction time task. *Personality and Individual Differences, 22*, 693–697. doi:10.1016/S0191-8869(96)00259-0

Eysenck, H. J. (1986). The theory of intelligence and the psychophysiology of cognition. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 3, pp. 1–34). Hillsdale: Erlbaum.

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience, 14,* 340–347. doi:10.1162/089892902317361886

Faust, M. E., Balota, D. A., Spieler, D. H., & Ferraro, F. R. (1999). Individual differences in information-processing rate and amount: Implications for group differences in response latency. *Psychological Bulletin, 125,* 777–799. doi:10.1037//0033-2909.125.6.777

Fisher, D. L., & Glaser, R. A. (1996). Molar and latent models of cognitive slowing: Implications for aging, dementia, depression, development, and intelligence. *Psychonomic Bulletin & Review, 3*(4), 458–480. doi:10.3758/BF03214549

Fisk, A. D., & Fisher, D. L. (1994). Brinley plots and theories of aging: The explicit, muddled, and implicit debates. *Journals of Gerontology, 49*(2), 81–89.

Fjell, A. M., Ostby, Y., & Walhovd, K. B. (2007). Intraindividual variability in cognitive performance and its neural foundation–an important approach in studying the developing and aging mind. In M. A. Lange (Ed.), *Leading-edge psychological tests and testing research* (pp. 193–200). Hauppauge: Nova Science.

Forstmann, B. U., Jahfari, S., Scholte, H. S., Wolfensteller, U., Van den Wildenberg, W. P. M., & Ridderinkhof, K. R. (2008). Function and structure of the right inferior frontal cortex predict individual differences in response inhibition: A model-based approach. *Journal of Neuroscience, 28,* 9790–9796. doi:10.1523/JNEUROSCI.1465-08.2008

Fuguitt, G. V., & Lieberson, S. (1973). Correlation of ratios or difference scores having common terms. *Sociological Methodology, 5,* 128–144. doi:10.2307/270835

Gignac, G. E., & Vernon, P. A. (2004). Reaction time and the dominant and non-dominant hands: An extension of Hick's law. *Personality and Individual Differences, 36*(3), 733–739. doi:10.1016/S0191-8869(03)00133-8

Godefroy, O., Lhullier, C., & Rousseaux, M. (1994). Reliability of reaction time measurements in brain-damaged patients. *Journal of the Neurological Sciences, 126,* 168–171. doi:10.1016/0022-510X(94)90268-2

Graham, J. M. (2006). Congeneric and (essentially) tau-equivalent estimates of score reliability: What they are and how to use them. *Educational and Psychological Measurement, 66*(6), 930–944. doi:10.1177/0013164406288165

Greenwald, A. G., McGhee, D. E., & Schwartz, J. K. L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74*(6), 1464–1480. doi:10.1037//0022-3514.74.6.1464

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*(2), 197–216. doi:10.1037/0022-3514.85.2.197

Hartley, A. A. (2001). Age differences in dual-task interference are localized to response-generation processes. *Psychology and Aging, 16*(1), 47–54. doi:10.1037//0882-7974.16.1.47

Hasbroucq, T., Kornblum, S., & Osman, A. M. (1988). A new look at reaction time estimates of interhemispheric transmission time. *Cahiers de Psychologie Cognitive, European Bulletin of Cognitive Psychology, 8,* 207–221.

Helmbold, N., Troche, S., & Rammsayer, T. (2007). Processing of temporal and nontemporal information as predictors of psychometric intelligence: A structural-equation-modeling approach. *Journal of Personality, 75*(5), 985–1006. doi:10.1111/j.1467-6494.2007.00463.x

Hillis, W. D. (1993). Why physicists like models, and why biologists should. *Current Biology, 3*(2), 79–81.

Hillyard, S., & Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. *Proceedings of the National Academy of Sciences, 95*(3), 781–787. doi:10.1073/pnas.95.3.781

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Personality and Social Psychology Bulletin, 31,* 1369–1385. doi:10.1177/0146167205275613

Huang, L., Mo, L., & Li, Y. (2012). Measuring the interrelations among multiple paradigms of visual attention: An individual differences approach. *Journal of Experimental Psychology. Human Perception and Performance, 38*(2), 414–428. doi:10.1037/a0026314

Hunt, E. (1978). Mechanics of verbal ability. *Psychological Review, 85*(2), 109–130. doi:10.1037//0033-295X.85.2.109

Iacoboni, M., & Zaidel, E. (2000). Cross-uncrossed difference in simple reaction times to lateralized flashes: Between- and within-subjects variability. *Neuropsychologia, 38,* 535–541. doi:10.1016/S0028-3932(99)00121-9

Iacoboni, M., & Zaidel, E. (2004). Interhemispheric visuo-motor integration in humans: The role of the superior parietal cortex. *Neuropsychologia, 42,* 419–425. doi:10.1016/j.neuropsychologia.2003.10.007

Indermühle, R., Troche, S. J., & Rammsayer, T. H. (2011). Personality and the psychological refractory period: No evidence for an extraversion- or intelligence-related effect. *Canadian Journal of Behavioural Science, 43*(3), 214–221. doi:10.1037/a0024156

Jensen, A. R. (1985). Methodological and statistical techniques for the chronometric study of mental abilities. In C. R. Reynolds & V. L. Willson (Eds.), *Methodological and statistical advances in the study of individual differences* (pp. 51–116). New York: Plenum Press.

Jensen, A. R. (1992). The importance of intraindividual variation in reaction time. *Personality and Individual Differences, 13,* 869–881. doi:10.1016/0191-8869(92)90004-9

Jensen, A. R. (1993). Spearman's g: Links between psychometrics and biology. *Annals of the New York Academy of Sciences, 702*(1), 103–129. doi:10.1111/j.1749-6632.1993.tb17244.x

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

Jensen, A. R., & Munro, E. (1979). Reaction time, movement time, and intelligence. *Intelligence, 3*(2), 121–126. doi:10.1016/0160-2896(79)90010-2

Jepma, M., Wagenmakers, E.-J., & Nieuwenhuis, S. (2012). Temporal expectation and information processing: A model-based analysis. *Cognition, 122,* 426–441. doi:10.1016/j.cognition.2011.11.014

Jonides, J., & Mack, R. (1984). On the cost and benefit of cost and benefit. *Psychological Bulletin, 96,* 29–44. doi:10.1037//0033-2909.96.1.29

Just, M. A., & Carpenter, P. A. (1985). Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. *Psychological Review, 92,* 137–172. doi:10.1037//0033-295X.92.2.137

Kane, M. J., May, C. P., Hasher, L., Rahhal, T., & Stoltzfus, E. R. (1997a). Dual mechanisms of negative priming. *Journal of Experimental Psychology. Human Perception and Performance, 23*(3), 632–650.

Kane, H. D., Proctor, B. E., & Kranzler, J. H. (1997b). Reliability and validity of a non-verbal measure of the speed and efficiency of long-term memory retrieval. *Personality and Individual Differences, 22,* 127–129. doi:10.1016/S0191-8869(96)00179-1

Karwoski, T. F., & Schachter, J. (1948). Psychological studies in semantics: III. Reaction times for similarity and difference. *Journal of Social Psychology, 28,* 103–120. doi:10.1080/00224545.1948.9921759

Kauranen, K. J., & Vanharanta, H. V. (2001). Relationship between extremities in motor performance. *Perceptual and Motor Skills, 92*(1), 11–18. doi:10.2466/PMS.92.1.11-18

Keating, D. P., & Bobbitt, B. (1978). Individual and developmental differences in cognitive-processing components of mental ability. *Child Development, 49,* 155–167. doi:10.2307/1128604

Keye, D., Wilhelm, O., Oberauer, K., & Van Ravenzwaaij, D. (2009). Individual differences in conflict-monitoring: Testing means and

covariance hypothesis about the Simon and the Eriksen Flanker task. *Psychological Research, 73,* 762–776. doi:10.1007/s00426-008-0188-9

Kieras, D. E., & Meyer, D. E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human Computer Interaction, 12,* 391–438.

Kirby, N. H., & Nettelbeck, T. (1989). Reaction time and inspection time as measures of intellectual ability. *Personality and Individual Differences, 10,* 11–14. doi:10.1016/0191-8869(89)90171-2

Larson, M., & Clayson, P. (2011). The relationship between cognitive performance and electrophysiological indices of performance monitoring. *Cognitive, Affective, & Behavioral Neuroscience, 11*(2), 159–171. doi:10.3758/s13415-010-0018-6

Lindsay, D. S., & Jacoby, L. L. (1994). Stroop process dissociations: The relationship between facilitation and interference. *Journal of Experimental Psychology. Human Perception and Performance, 20*(2), 219–234. doi:10.1037/0096-1523.20.2.219

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading: Addison-Wesley.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization.* Oxford: Oxford University Press.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109,* 163–203. doi:10.1037//0033-2909.109.2.163

MacLeod, J. W., Lawrence, M. A., McConnell, M. M., Eskes, G. A., Klein, R. M., & Shore, D. I. (2010). Appraising the ANT: Psychometric and theoretical considerations of the attention network test. *Neuropsychology, 24,* 637–651. doi:10.1037/a0019803

Maloney, E. A., Risko, E. F., Preston, F., Ansari, D., & Fugelsang, J. (2010). Challenging the reliability and validity of cognitive measures: The case of the numerical distance effect. *Acta Psychologica, 134*(2), 154–161. doi:10.1016/j.actpsy.2010.01.006

McClelland, J. L. (1979). On the time relations of mental processes: A framework for analyzing processes in cascade. *Psychological Review, 86,* 287–330. doi:10.1037/0033-295X.86.4.287

McConnell, M., & Shore, D. (2011). Mixing measures: Testing an assumption of the attention network test. *Attention, Perception, & Psychophysics, 73*(4), 1096–1107. doi:10.3758/s13414-010-0085-3

Miles, J. D., & Proctor, R. W. (2012). Correlations between spatial compatibility effects: Are arrows more like locations or words? *Psychological Research, 76*(6), 777–791. doi:10.1007/s00426-011-0378-8

Miller, E. M. (1994). Intelligence and brain myelination: A hypothesis. *Personality and Individual Differences, 17,* 803–832. doi:10.1016/0191-8869(94)90049-3

Miller, J. O., & Ulrich, R. (2003). Simple reaction time and statistical facilitation: A parallel grains model. *Cognitive Psychology, 46,* 101–151. doi:10.1016/S0010-0285(02)00517-0

Miller, J. O., Van der Ham, F., & Sanders, A. F. (1995). Overlapping stage models and reaction time additivity: Effects of the activation equation. *Acta Psychologica, 90,* 11–28. doi:10.1016/0001-6918(95)00028-S

Molenaar, P. C. M., & Van der Molen, M. W. (1986). Steps to a formal analysis of the cognitive-energetic model of stress and human performance. *Acta Psychologica, 62,* 237–261. doi:10.1016/0001-6918(86)90090-9

Mood, A. M., Graybill, F. A., & Boes, D. C. (1974). *Introduction to the theory of statistics* (3rd ed.). New York: McGraw-Hill.

Moritz, S., & Andresen, B. (2004). Reduced negative priming in schizotypy: Failure to replicate. *British Journal of Clinical Psychology, 43*(2), 211–215. doi:10.1348/014466504323088060

Moritz, S., & Mass, R. (1997). Reduced cognitive inhibition in schizotypy. *British Journal of Clinical Psychology, 36*(3), 365–376. doi:10.1111/j.2044-8260.1997.tb01244.x

Myerson, J., Adams, D. R., Hale, S., & Jenkins, L. (2003). Analysis of group differences in processing speed: Brinley plots, Q-Q plots, and other conspiracies. *Psychonomic Bulletin & Review, 10*(1), 224–237. doi:10.3758/BF03196489

Myerson, J., Hale, S., Chen, J., & Lawrence, B. (1997). General lexical slowing and the semantic priming effect: The roles of age and ability. *Acta Psychologica, 96,* 83–101. doi:10.1016/S0001-6918(97)00002-4

Myerson, J., Robertson, S., & Hale, S. (2007). Aging and intraindividual variability in performance: Analyses of response time distributions. *Journal of the Experimental Analysis of Behavior, 88,* 319–337. doi:10.1901/jeab.2007.88-319

Navon, D., & Miller, J. O. (2002). Queuing or sharing? A critical evaluation of the single-bottleneck notion. *Cognitive Psychology, 44,* 193–251. doi:10.1006/cogp.2001.0767

Nettlebeck, T. (1980). Factors affecting reaction time: Mental retardation, brain damage, and other psychopathologies. In A. T. Welford (Ed.), *Reaction times* (pp. 355–401). New York: Academic Press.

Neubauer, A. C., Riemann, R., Mayer, R., & Angleitner, A. (1997). Intelligence and reaction times in the Hick, Sternberg and Posner paradigms. *Personality and Individual Differences, 22,* 885–894.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Math = male, me = female, therefore math ≠ me. *Journal of Personality and Social Psychology, 83*(1), 44–59. doi:10.1037/0022-3514.83.1.44

Nosek, B. A., & Smyth, F. L. (2007). A multitrait-multimethod validation of the implicit association test: Implicit and explicit attitudes are related but distinct constructs. *Experimental Psychology, 54*(1), 14–29.

Ollman, R. T. (1973). Simple reactions with random countermanding of the "go" signal. In S. Kornblum (Ed.), *Attention and performance IV* (pp. 571–581). New York: Academic.

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 cognition* (pp. 41–82). Hillsdale: Erlbaum.

Pashler, H. E. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 116,* 220–244. doi:10.1037//0033-2909.116.2.220

Pashler, H. E., & Johnston, J. C. (1989). Chronometric evidence for central postponement in temporally overlapping tasks. *Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology, 41,* 19–45.

Pearson, K. (1896). Mathematical contributions to the theory of evolution. On a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proceedings of the Royal Society of London, 60*(1), 489–498. doi:10.1098/rspl.1896.0076

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

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59–108. doi:10.1037//0033-295X.85.2.59

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review, 111,* 333–367. doi:10.1037/0033-295X.111.2.333

Ratcliff, R., Spieler, D., & McKoon, G. (2000). Explicitly modeling the effects of aging on response time. *Psychonomic Bulletin & Review, 7,* 1–25. doi:10.3758/BF03210723

Reed, T. E., & Jensen, A. R. (1991). Arm nerve conduction velocity (NCV), brain NCV, reaction time, and intelligence. *Intelligence, 15,* 33–47. doi:10.1016/0160-2896(91)90021-5

Richeson, J. A., & Shelton, J. N. (2003). When prejudice does not pay: Effects of interracial contact on executive function. *Psychological Science, 14,* 287–290. doi:10.1111/1467-9280.03437

Ritter, W., Simson, R., & Vaughan, H. G., Jr. (1983). Event-related potential correlates of two stages of information processing in physical and semantic discrimination tasks. *Psychophysiology, 20,* 168–179. doi:10.1111/j.1469-8986.1983.tb03283.x

Sanders, A. F. (1998). *Elements of human performance: Reaction processes and attention in human skill.* Hillsdale: Erlbaum.

Schmiedek, F., Lovden, M., & Lindenberger, U. (2009). On the relation of mean reaction time and intraindividual reaction time variability. *Psychology and Aging, 24*(4), 841–857. doi:10.1037/a0017799

Schmiedek, F., Oberauer, K., Wilhelm, O., Süß, H. M., & Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology. General, 136,* 414–429. doi:10.1037/0096-3445.136.3.414

Schulte, T., Pfefferbaum, A., & Sullivan, E. V. (2004). Parallel interhemispheric processing in aging and alcoholism: Relation to corpus callosum size. *Neuropsychologia, 42,* 257–271. doi:10.1016/S0028-3932(03)00155-6

Schweizer, K. (1989). Relating reaction time components and intelligence. *Personality and Individual Differences, 10*(7), 701–707. doi:10.1016/0191-8869(89)90116-5

Seashore, S. H., & Seashore, R. H. (1941). Individual differences in simple auditory reaction times of hands, feet and jaws. *Journal of Experimental Psychology, 29*(4), 342–345. doi:10.1037/h0061571

Sheppard, L. D., & Vernon, P. A. (2008). Intelligence and speed of information-processing: A review of 50 years of research. *Personality and Individual Differences, 44*(3), 535–551. doi:10.1016/j.paid.2007.09.015

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84,* 127–190. doi:10.1037//0033-295X.84.2.127

Sigman, M., & Dehaene, S. (2006). Dynamics of the central bottleneck: Dual-task and task uncertainty. *PLoS Biology, 4,*(7) e220. doi:10.1371/journal.pbio.0040220

Simonen, R. L., Videman, T., Battie, M. C., & Gibbons, L. E. (1995). Comparison of foot and hand reaction times among men: A methodologic study using simple and multiple-choice repeated measurements. *Perceptual and Motor Skills, 80*(3), 1243–1249.

Smith, E. E. (1968). Choice reaction time: An analysis of the major theoretical positions. *Psychological Bulletin, 69,* 77–110. doi:10.1037/h0020189

Smith, M. C. (1969). The effect of varying information on the psychological refractory period. *Acta Psychologica, 30,* 220–231. doi:10.1016/0001-6918(69)90052-3

Smith, P. L. (1995). Psychophysically principled models of visual simple reaction time. *Psychological Review, 102,* 567–593. doi:10.1037//0033-295X.102.3.567

Smith, P. L. (2000). Stochastic dynamic models of response time and accuracy: A foundational primer. *Journal of Mathematical Psychology, 44,* 408–463. doi:10.1006/jmps.1999.1260

Smith, P. L., Ratcliff, R., & Wolfgang, B. J. (2004). Attention orienting and the time course of perceptual decisions: Response time distributions with masked and unmasked displays. *Vision Research, 44,* 1297–1320. doi:10.1016/j.visres.2004.01.002

Smith, G. A., & Stanley, G. (1983). Clocking g: Relating intelligence and measures of timed performance. *Intelligence, 7,* 353–368. doi:10.1016/0160-2896(83)90010-7

Smulders, F. T. Y., & Meijer, E. H. (2008). Extraversion and performance: A cognitive-energetical approach. *Personality and Individual Differences, 44*(2), 475–486. doi:10.1016/j.paid.2007.09.012

Sriram, N., Greenwald, A. G., & Nosek, B. A. (2010). Correlational biases in mean response latency differences. *Statistical Methodology, 7*(3), 277–291. doi:10.1016/j.stamet.2009.10.004

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica, 30,* 276–315. doi:10.1016/0001-6918(69)90055-9

Sternberg, S. (2001). Separate modifiability, mental modules, and the use of pure and composite measures to reveal them. *Acta Psychologica, 106,* 147–246. doi:10.1016/S0001-6918(00)00045-7

Stolz, J. A., Besner, D., & Carr, T. H. (2005). Implications of measures of reliability for theories of priming: Activity in semantic memory is inherently noisy and uncoordinated. *Visual Cognition, 12,* 284–336.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643–662. doi:10.1037//0096-3445.121.1.15

Stuss, D. T., Pogue, J., Buckle, L., & Bondar, J. (1994). Characterization of stability of performance in patients with traumatic brain injury: Variability and consistency on reaction time tests. *Neuropsychology, 8*(3), 316–324. doi:10.1037//0894-4105.8.3.316

Tipper, S. P. (1985). The negative priming effect: Inhibitory priming by ignored objects. *Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology, 37,* 571–590.

Tombu, M., & Jolicœur, P. (2003). A central capacity sharing model of dual-task performance. *Journal of Experimental Psychology. Human Perception and Performance, 29,* 3–18. doi:10.1037//0096-1523.29.1.3

Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology, 39,* 321–359. doi:10.1006/jmps.1995.1033

Ulrich, R., & Wing, A. M. (1991). A recruitment theory of force-time relations in the production of brief force pulses: The parallel force unit model. *Psychological Review, 98,* 268–294. doi:10.1037//0033-295X.98.2.268

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review, 108*(3), 550–592. doi:10.1037//0033-295X.108.3.550

Van Ravenzwaaij, D., Brown, S., & Wagenmakers, E.-J. (2011). An integrated perspective on the relation between response speed and intelligence. *Cognition, 119,* 381–393. doi:10.1016/j.cognition.2011.02.002

Vernon, P. A. (1990). An overview of chronometric measures of intelligence. *School Psychology Review, 19*(4), 399–410.

Vernon, P. A., & Mori, M. (1992). Intelligence, reaction times, and peripheral nerve conduction velocity. *Intelligence, 16*(3–4), 273–288. doi:10.1016/0160-2896(92)90010-O

Versace, F., Mazzetti, M., & Codispoti, M. (2008). The temporal stability of the effects induced by the cued reaction time task. *Assessment, 15,* 145–152. doi:10.1177/1073191107308106

Vickers, D. (1970). Evidence for an accumulator model of psychophysical discrimination. *Ergonomics, 13,* 37–58. doi:10.1080/00140137008931117

Waechter, S., Stolz, J. A., & Besner, D. (2010). Visual word recognition: On the reliability of repetition priming. *Visual Cognition, 18*(4), 537–558. doi:10.1080/13506280902868603

Wagenmakers, E.-J., & Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological Review, 114,* 830–841. doi:10.1037/0033-295X.114.3.830

Way, T. C., & Gottsdanker, R. (1968). Psychological refractoriness with varying differences between tasks. *Journal of Experimental Psychology, 78,* 38–45. doi:10.1037/h0026154

Wiers, R. W., Van Woerden, N., Smulders, F. T. Y., & De Jong, P. J. (2002). Implicit and explicit alcohol-related cognitions in heavy and light drinkers. *Journal of Abnormal Psychology, 111,* 648–658. doi:10.1037//0021-843X.111.4.648

Williams, L. E., Light, G. A., Braff, D. L., & Ramachandran, V. S. (2010). Reduced multisensory integration in patients with schizophrenia on a target detection task. *Neuropsychologia, 48*(10), 3128–3136. doi:10.1016/j.neuropsychologia.2010.06.028

Williams, R. H., & Zimmerman, D. W. (1996). Are simple gain scores obsolete? *Applied Psychological Measurement, 20*(1), 59–69. doi:10.1177/014662169602000106

Wilson, G. F., & O'Donnell, R. D. (1986). Steady state evoked responses: Correlations with human cognition. *Psychophysiology, 23,* 57–61. doi:10.1111/j.1469-8986.1986.tb00594.x

Wood, G., Willmes, K., Nuerk, H.-C., & Fischer, M. H. (2008). On the cognitive link between space and number: A meta-analysis of the SNARC effect. *Psychology Science Quarterly, 50*(4), 489–525.

Yellott, J. I., Jr. (1971). Correction for fast guessing and the speed-accuracy tradeoff in choice reaction time. *Journal of Mathematical Psychology, 8,* 159–199. doi:10.1016/0022-2496(71)90011-3