Supplemental Material
General Cognitive Ability and the Psychological Refractory Period: Individual Differences in the Mind's Bottleneck

James J. Lee and Christopher F. Chabris

Lee and Chabris (2013). *Psychological Science*.
doi:10.1177/0956797612471540.

Additional Methods

The dual number-comparison task was implemented in PsyScope X Build 57 and run on iMac desktops. The diagonal length of the monitor was 50.8 cm. The stimuli were displayed in the Monaco font with a point size of 36 and a duration of 250 ms. The viewing distance was approximately 70 cm.

Participants familiarized themselves with the task over a total of 106 practice trials before beginning the actual experiment. One preliminary practice block consisted of 10 trials with variable SOA. Two practice blocks consisted of 32 trials with variable SOA. One practice block consisted of 32 trials, all with an SOA of 60 ms, in order to accustom participants to performing the task quickly and accurately without "grouping" responses (withholding the first response and then emitting both responses in a rapid burst). The instructions strongly emphasized prioritizing the first stimulus.

In our previous research, we have found that student volunteers from the Boston area exhibit a very high mean SAT score and little variability (about 40 percent of the total variance among all examinees taking the SAT) (Chabris et al., in press). Our recruitment of participants from extreme groups was therefore intended to overcome low statistical power to detect any baseline g-RT association. We validated the documentation of SAT scores with untimed administrations of a short-form Raven's

Advanced Progressive Matrices and the Vocabulary subtest of the Multidimensional Aptitude Battery II (Figure S1).

Participants performed the number-comparison task and completed the psychometric tests in a session lasting between 2 and 3.5 hours. Up to 3 participants were tested under the supervision of an experimenter in a single session. Each participant was seated at a cubicle obstructing all other participants from view. All participants in a session were members of only one g group. Most participants received a baseline compensation of $40; a few participants performed the study for course credit.

The moderate-g participants were told that they were participating in "Version A," while the high-g participants were told that they were participating in "Version B." After completing all tasks, each participant had to predict the outcome of a die roll. Within each version all participants who correctly predicted the roll were entered in a pool and randomly paired. The participant within the pair who performed better on the sum of mean RT in the number-comparison task and number correct on the IQ tests was mailed a check for an additional $40. Each term in this sum was converted to a Z-score and appropriately signed. Sacrificing accuracy for speed was discouraged by penalizing accuracy below 90 percent with a multiplication of mean RT by $(1+.20x)$, where x is the participant's number of rounded percentage points below 90-percent accuracy. One randomly chosen participant was not paired because the number of pool members was odd, and whether to grant this participant an incentive payment was also decided randomly. This incentivizing of good performance was explained, with a numerical example, to participants at the beginning of the session. All participants, even those performing the study for course credit, were eligible for incentive payments.

The reason for incentivizing participants to perform well *only* relative to other participants in their version (A or B) was to ensure that moderate-g participants, despite tending to have slower RT and lower test scores, were as likely to receive incentive payments as high-g participants. The moderate-g participants were thus given a motivation to perform well and a strong signal that their good performance was of importance to the experiment. It is unlikely that these participants were aware of the fact that higher test scores are associated with faster and less variable RT. Nevertheless, even if any participant was so aware (perhaps having read about it in a psychology textbook), all participants were given a motivation to perform well because they were competing against participants recruited according to the same SAT score criterion.

The salience of accuracy in our incentive instructions may have led all participants to the same target accuracy of roughly 95% (Table S1). Participants with slower drift rates would then have adopted more conservative boundary separations to prevent themselves from making more errors. This may account for the unexpected association between g and boundary separation.

The study advertisements, online listing, and 24-hour-notice email all emphasized the possibility of an additional 40-dollar reward for good performance and the importance of being well rested.

After every fourth block during the study itself, participants were given an indefinitely long break, during which they were allowed to use the restroom, have a cup of water, and so on. The next block was initiated when all participants indicated that they were ready.

Before analyzing the data, we discarded all trials resulting in RT1 or RT2 less than 250 ms. All trials resulting in either RT more than 4 standard deviations from a participant's mean in a given cell (first or second stimulus, SOA, dichotomized numerical distance) were discarded iteratively. These criteria eliminated about 0.5% of all trials in both the moderate- and high-g groups. In the analysis of raw RT, only trials resulting in a correct response were used; when one response within a trial was correct and the other was in error, we only discarded the erroneous response.

The EZ2 diffusion estimation method requires a nonzero number of errors. Whenever a participant made no errors in a given cell, we set the proportion of errors to one half divided by the number of trials in the cell.

One high-g participant's T was estimated to be negative. If the SD of the T estimates is based only on the other participants, this participant's estimate is nearly 11 standard deviations below the sample mean. Examination of the participant's raw data revealed that she performed like a typical high-g participant at the longest SOA, but became an extreme outlier at the second longest SOA with an RT2 variance more than four times greater than that of the next most variable participant. Her pattern of performance across all SOAs was similarly erratic. We excluded this participant from all analyses.

Explicit PRP Model

Let E denote the executive task-scheduling stage preceding C_1 in a dual task. Also let τ denote the summed durations of P_1, E, and C_1—that is, the duration of all stages preceding and constituting the serial bottleneck.

Then we can write the dependence of RT2 on SOA as

$$\text{RT2(SOA)} = \begin{cases} P_1 + E + C_1 + C_2 + M_2 - \text{SOA} & \text{if SOA} + P_2 \leq \tau, \\ P_2 + C_2 + M_2 & \text{if SOA} + P_2 > \tau. \end{cases}$$

Note that E is itself a function of SOA; it has been found that this stage can sometimes be prolonged by shortening the SOA (or withholding instructions regarding which stimulus should be prioritized). Direct estimates of the $P + M$ and C durations are made possible by assuming that corresponding stages of tasks 1 and 2 have the same mean, which is plausible if tasks 1 and 2 are identical. At short SOAs where we may safely assume that $\text{SOA} + P_2 \leq \tau$ always holds, we simply use

$$\begin{aligned} &2 \times \text{RT1(SOA)} - \text{RT2(SOA)} - E - \text{SOA} \\ &= 2(P + E + C + M) - (P + E + 2C + M - \text{SOA}) - E - \text{SOA} \\ &= 2P + 2E + 2C + 2M - P - E - 2C - M + \text{SOA} - E - \text{SOA} \\ &= P + M \end{aligned}$$

to estimate the duration of the non-bottleneck stages. However, before applying this equation, we must be able to estimate E at a given SOA. On the assumption that RT1 at a given SOA is longer than asymptotic RT2 as a result of inserting E between P_1 and C_1, we can estimate E(SOA) by subtracting RT1(SOA) from asymptotic RT2 (that is, RT2 at the two longest SOAs, *not* RT2 at the given SOA). We used this method, averaging the results over the 60- and 120-ms SOAs, to estimate $P + M$ separately for the moderate- and high-g groups.

We also used our estimates of E(SOA) to correct for the carryover of E into RT2 in our estimates of how much the high-g advantage increases as the SOA becomes small. This carryover occurs because the insertion of E pushes forward C_1 and hence C_2. At a short SOA, we used

$$\frac{[\text{RT2}_{\text{low } g}(\text{SOA}) - E_{\text{low } g}(\text{SOA})] - [\text{RT2}_{\text{high } g}(\text{SOA}) - E_{\text{high } g}(\text{SOA})]}{\text{RT2}_{\text{low } g}(\infty) - \text{RT2}_{\text{high } g}(\infty)},$$

again averaging over the 60- and 120-ms SOAs. This correction is desirable because the insertion of E into the first reaction process in a dual task introduces a processing stage that is absent from the equivalent reaction process performed as a single task, thus complicating our attempts to infer the nature of the g-RT correlation observed in single tasks. One such complication is that our quantitative prediction under the seriality hypothesis can be no longer be expressed as a specific number but rather as an algebraic expression involving the difference between the two g groups in the durations of E and C. That is, whereas the ratio above correcting for the carryover of E into RT2 reduces to the number two under the seriality hypothesis, the corresponding ratio without the correction becomes two plus a term that cannot be known *a priori*.

Because it is not obvious how to perform parametric statistical inference with respect to this complicated ratio, we used the BC_a bootstrap to decide between the two contending hypotheses regarding the ratio's value.

Variability in Diffusion Rate Across Trials

The EZ diffusion estimation methods work best when a participant exhibits no variability across trials in the parameters (Wagenmakers, van der Maas, & Grasman, 2007). In the number-comparison task, however, v increases dramatically with numerical distance between reference and stimulus (Dehaene, 2007).

We did not administer enough trials to estimate a full diffusion model with across-trial variability (Ratcliff & Tuerlinckx, 2002). To verify robustness against at least across-trial variability in v, we used EZ2 to estimate each participant's parameter values for each distinct numerical distance from 1 through 4. We included trials from all SOAs from 720 to 960 ms to approach 80 trials per distance, although some participants still exhibited signs of PRP interference at the first few of these SOAs. The mean v among moderate-g participants was .28 when the distance was equal to 1 and increased by $.058 \pm .003$ per distance increment. The corresponding regression line for the high-g participants was displaced upward by $.039 \pm .017$.

We averaged each of the three diffusion parameters over distances. These averaged parameter estimates were highly correlated with the estimates used in the main text, which neglected distance and used only trials where SOA $\geq$ 900 ms; $v, r(68) = .92, p < .001$; $a, r(68) = .88, p < .001$; $T, r(68) = .83, p < .001$. For all three parameters, the line of zero intercept and unit slope fit the two sets of estimates well, although the distance-free estimates of v tended to fall short of the estimates incorporating distance. The two groups did not differ in T, $t(54.2) = .05, p > .95$.

Perceptual Parallelism for Two Visual Tasks

To verify the parallelism of the two visual identifications in our task, we ran an additional experiment varying the contrasts of both stimuli.

We tested nine participants. We employed the PsyScope default for the black (high-contrast) stimuli and the setting $(-5000, -5000, -5000)$ for the gray (low-contrast) stimuli. There were only four SOAs (50 to 950 ms inclusive, increments of 300 ms). All 256 possible combinations of numbers and contrast levels were used 6 times to constitute the real trials.

Reducing the contrast of stimulus 1 slowed RT1 (29 ± 3.6 ms) and also RT2 at the shortest SOA (32 ± 16 ms). Reducing the contrast of stimulus 2 showed a similar effect on RT2 at the two longest SOAs (27 ± 4.9 ms) but no effect at the shortest SOA (-8 ± 13 ms). This pattern confirms that visual identification remains an early parallel stage in our repetition of the same task (Pashler, 1994).

We also found evidence that a reduction in contrast affects T. A 21-ms increase in T calculated from RT1 over all SOAs reached significance, $t(8) = 3.23, p < .02$. Unexpectedly a decrease of .018 in v for RT1 also reached significance, $t(8) = 4.17, p < .007$.

Calculating diffusion parameters for RT2 at distinct SOAs often produced implausible estimates. In particular, two participants showed negative T estimates at the longest SOA. Recall that only one out of seventy-one participants showed this anomaly in the main experiment. After removing the two anomalous participants from the data collected in the ancillary experiment, we found that a reduction in contrast led to a significant 52-ms increase in T, $t(6) = 4.87, p < .003$ and a nonsignificant increase of .036 in v, $t(6) = 1.50, p > .18$. Including the two participants with unstable data did not change this pattern of effect signs and statistical significance. The effect of contrast on v thus showed opposite signs across RT1 and RT2.

Crosstalk Between Tasks

Stimulus-response overlap between tasks 1 and 2 can lead to "crosstalk" (Hommel, 1998; Schubert, Fischer, & Stelzel, 2008). In our dual number-comparison talk, a slowdown attributable to crosstalk occurs when one stimulus is less than 5 and the other is greater than 5; conversely, responses become more rapid when both stimuli bear the same relation to 5. Figure S2 displays the RT means with stimulus congruence as an additional variable. The difference in RT2 between the g groups was very similar at the three shortest SOAs for both congruent and incongruent trials (99 ± 26 and 108 ± 32 ms respectively). The finding that the RT2 difference more than doubles as the SOA becomes small is therefore robust against crosstalk.

Selection Bias

There are two reasons why the g-C correlation is probably not attributable to selection bias (Bareinboim & Pearl, 2012). First, it is rather improbable that speed of central processing in the number-comparison task would affect the likelihood of a person participating in our study. Second, studies of population samples or convenience samples without restriction of range in IQ have found robust associations between g and RT (or, more informatively, diffusion rate).

Figure S1: Validation of SAT scores as a measure of g. We used the short form of Raven's Advanced Progressive Matrices devised by Bors and Stokes (1998) with the addition of items 32 and 36 (two of the most difficult items). Plotted are the jittered number correct. The two groups differed significantly in both RAPM scores, $t(47.5) = 7.16, p \ll .001, d = 1.77$, and MAB-II Vocabulary scores, $t(50.1) = 11.6, p \ll .001, d = 2.85$.

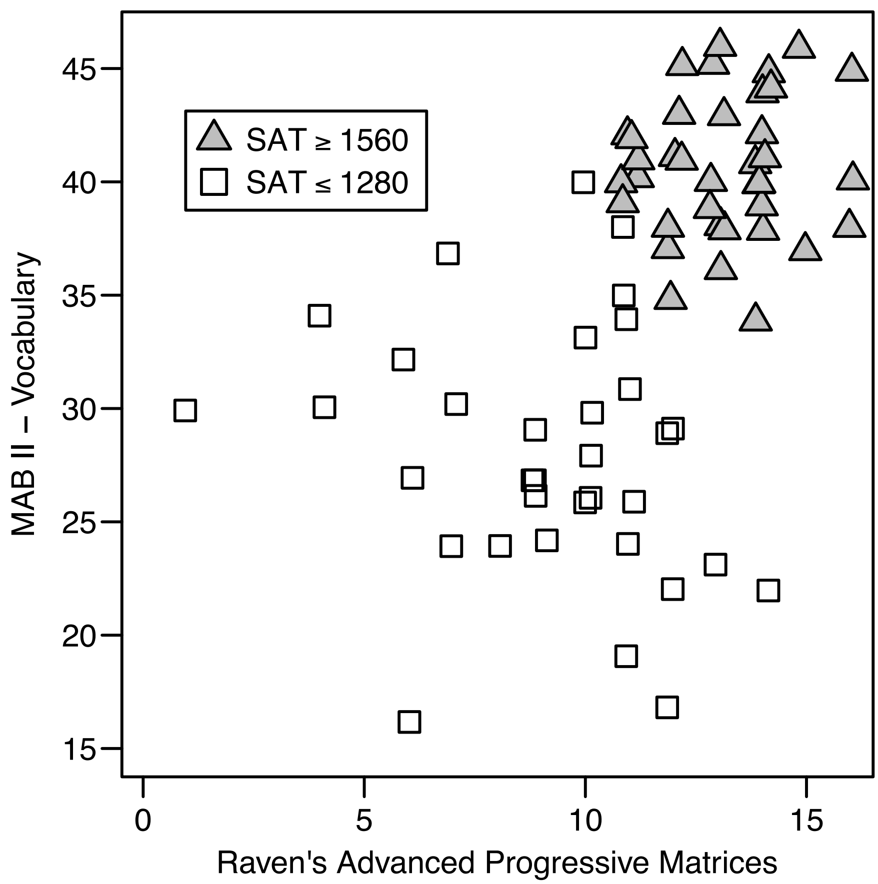

Figure S2: Mean RT as a function of stimulus order, SOA, congruence, and g group. The first and second stimuli within a trial are congruent if they are both less than 5 or both greater than 5.

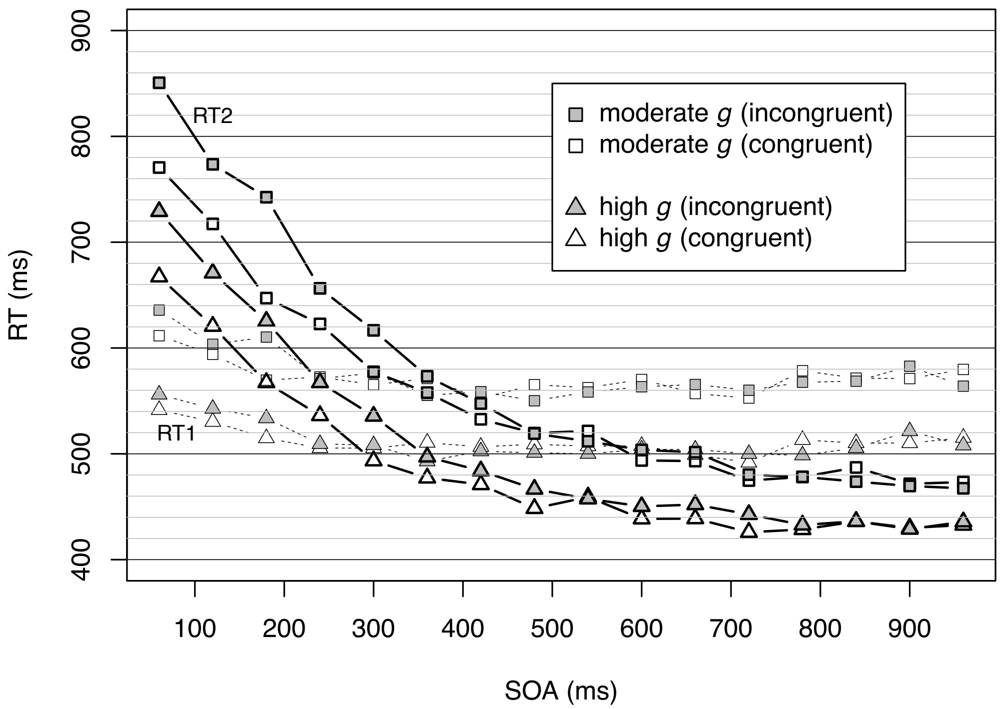

Table S1: Additional Summary Statistics of RT1 and RT2

SOA	RT1 variance		RT1 accuracy		RT2 variance		RT2 accuracy	
	high g	moderate g	high g	moderate g	high g	moderate g	high g	moderate g
60	1.45×10^4	3.51×10^4	.973	.966	2.30×10^4	4.80×10^4	.958	.944
120	1.75×10^4	3.47×10^4	.979	.967	2.31×10^4	4.39×10^4	.960	.953
180	1.56×10^4	3.20×10^4	.968	.964	1.94×10^4	3.99×10^4	.948	.943
240	1.55×10^4	2.69×10^4	.967	.960	1.58×10^4	3.10×10^4	.949	.942
300	1.45×10^4	2.56×10^4	.972	.977	1.46×10^4	2.94×10^4	.956	.949
360	1.57×10^4	2.76×10^4	.970	.969	1.46×10^4	3.23×10^4	.959	.954
420	1.50×10^4	2.32×10^4	.962	.958	1.20×10^4	2.30×10^4	.956	.937
480	1.31×10^4	2.42×10^4	.966	.965	8.95×10^3	2.32×10^4	.962	.958
540	1.46×10^4	2.50×10^4	.968	.976	1.02×10^4	1.93×10^4	.946	.954
600	1.25×10^4	3.05×10^4	.966	.970	8.27×10^3	2.22×10^4	.962	.958
660	1.08×10^4	2.91×10^4	.973	.965	8.27×10^3	2.12×10^4	.963	.957
720	1.32×10^4	2.63×10^4	.974	.970	7.84×10^3	1.27×10^4	.954	.962
780	1.33×10^4	3.00×10^4	.967	.964	6.00×10^3	1.75×10^4	.964	.964
840	1.43×10^4	2.63×10^4	.969	.963	7.27×10^3	1.48×10^4	.958	.960
900	1.48×10^4	2.92×10^4	.959	.955	6.25×10^3	1.21×10^4	.958	.965
960	1.30×10^4	2.34×10^4	.966	.962	7.21×10^3	1.04×10^4	.958	.956

References

Bareinboim, E., & Pearl, J. (2012). Controlling selection bias in causal inference. In *JMLR Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics* (Vol. 22, pp. 100–108).

Bors, D. A., & Stokes, T. L. (1998). Raven's Advanced Progressive Matrices: Norms for first-year university students and the development of a short form. *Educational and Psychological Measurement*, *58*, 382–398.

Chabris, C. F., Lee, J. J., Benjamin, D. J., Beauchamp, J., Glaeser, E. L., Borst, G., ... Laibson, D. I. (in press). Why is it hard to find genes that are associated with social science traits? Theoretical and empirical considerations. *American Journal of Public Health*.

Dehaene, S. (2007). Symbols and quantities in parietal cortex: Elements of a mathematical theory of number representation and manipulation. In P. Haggard, Y. Rossetti & M. Kawato (Eds.), *Attention and performance XXII: Sensorimotor foundations of higher cognition and performance* (pp. 527–574). New York: Oxford University Press.

Hommel, B. (1998). Automatic stimulus-response translation in dual-task performance. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 1368–1384.

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

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin and Review*, *9*, 438–481.

Schubert, T., Fischer, R., & Stelzel, C. (2008). Response activation in overlapping tasks and the response-selection bottleneck. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 376–397.

Wagenmakers, E.-J., van der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin and Review*, *14*, 3–22.