Individual differences within and across attentional blink tasks revisited

Gillian Dale · Paul E. Dux · Karen M. Arnell

Abstract When the second of two targets (T2) is presented in close temporal proximity (within 200–500 ms) to the first (T1), the accuracy for reporting T2 is reduced relative to when the targets are separated by longer durations; this effect is known as the *attentional blink* (AB). Two recent studies have shown that individual differences in the magnitudes of the AB are stable both within a single testing session and over time. While one study found a large positive correlation between AB magnitudes when there was an attentional-set/task switch between T1 and T2 and when there was not, the other study found no relationship between the switch and no-switch paradigms. The present study was conducted to clarify this discrepancy by examining the reliability of, and relationships among, individual differences in AB performance on five different versions of the standard dual-target RSVP paradigm, three of which involved an attentional-set/task switch between T1 and T2, and two of which did not. Participants completed all five paradigms, and then returned 7–10 days later to again complete the same paradigms. The performance on all five versions was reliable both within and across testing sessions, demonstrating again that individual differences in AB performance are stable over time. In addition, all five AB versions were significantly intercorrelated, although the strengths of the relationships differed depending on the extent to which the T1 and T2 attentional sets/tasks overlapped. These findings provide evidence that multiple distinct dual-target RSVP tasks do share underlying variability, providing support for the use of different versions of the paradigm in the literature.

G. Dale (✉) · K. M. Arnell
Department of Psychology, Brock University,
500 Glenridge Avenue,
St. Catharines, ON L2S 3A1, Canada
e-mail: gd03fh@brocku.ca

P. E. Dux
The University of Queensland, St Lucia, Queensland, Australia

Keywords Attentional blink · Dual-task performance · Individual differences

When participants are required to report two targets from amongst a stream of distractors in a rapid serial visual presentation (RSVP) stream, accuracy for the second target (T2) is significantly reduced when it is presented within approximately 200–500 ms of the first target (T1). This phenomenon is known as the *attentional blink* (AB) and is of considerable interest to cognitive scientists, as it is thought to reflect a fundamental failure of perceptual consciousness (Raymond, Shapiro, & Arnell, 1992; see Dux & Marois, 2009, for a theoretical review, and MacLean & Arnell, 2012, for a methodological review).

As the study of the AB has grown in popularity, multiple variations on the original paradigm have been developed. As such, a variety of task combinations have now been shown to produce a robust lag-dependent T2 deficit under dual-target RSVP conditions. For example, T1 can require identification of a stimulus on the basis of a feature value (e.g., a red stimulus) or category membership (e.g., a letter as opposed to a digit). Some AB paradigms require participants to detect the presence or absence of a given item for T2 (e.g., "Was the 'X' present or absent?"), whereas others require participants to perform an *n*-alternative forced choice (*n*-AFC) to identify the second target (e.g., "Was an 'X' or 'Y' presented?"). The second target can also be defined on the basis of its physical features or categorical information. Thus, some versions of the AB contain an attentional-set/task switch, such that the T1 task differs from the T2 task (e.g., "What was the red T1 letter, and what black digit followed?"), while others contain no such switch (e.g., "Identify the two red digits"). The stimuli vary, too, with letters, digits, shapes, words, pictures, faces, line drawings, sounds, and compound letters/shapes having all been used as targets and distractors. As such, it is clear that a great deal of variation takes place from study to study in terms of the type of dual-target RSVP task employed.

Researchers have recently begun to examine whether performance on AB paradigms can be predicted by dispositional factors and an individual's performance on other cognitive tasks. For example, individual differences in AB magnitude have been shown to relate to individual scores on cognitive measures such as working memory control (e.g., Arnell, Stokes, MacLean, & Gicante, 2010; Colzato, Spapé, Pannebakker, & Hommel, 2007), the ability to inhibit irrelevant material from entering working memory (Arnell & Stubitz, 2010; Dux & Marois, 2008; Martens & Valchev, 2009), and the breadth of the attentional focus (e.g., Dale & Arnell, 2010). In addition, this deficit has been shown to correlate with dispositional factors such as affective traits (MacLean, Arnell, & Busseri, 2010) and personality (MacLean & Arnell, 2010), as well as physiological measures such as resting electroencephalography (EEG) in the alpha and beta bands (MacLean, Arnell, & Cote, 2012) and spontaneous eye-blink rate—a marker of dopaminergic function (Colzato, Slagter, de Rover, & Hommel, 2011).

However, many tasks and cognitive processes are not related to AB magnitude. Information processing speed (i.e., rapid automatized naming [RAN] times, 4-AFC manual response times to a single stimulus, vocal naming times to a single stimulus, and accuracy on a location probe task), reading speed, and reading comprehension have all been shown to be unrelated to AB magnitude, but to correlate with overall T2 accuracy (Arnell, Howe, Joanisse, & Klein, 2006). Similarly, fluid intelligence (Colzato et al., 2007), digit span (Arnell et al., 2010), visual working memory capacity (Arnell & Stubitz, 2010), and general task-switching ability (Klein, Arend, Beauducel, & Shapiro, 2011) have all been shown to relate to overall target accuracy, but not to the AB deficit. Thus, there appears to be specificity to what predicts AB magnitude, as opposed to general factors (e.g., general intelligence, alertness) underlying the effect (see Table 1 for a list of factors predicting AB magnitude and/or overall target accuracy).

Given the increasing use of the AB as an individual-differences measure, it has become important to establish its reliability. Estimates of reliability provide an upper bound on the degree of relationship that can be expected between two tasks, and it does not make sense to ask what dispositional characteristics predict individual differences in AB magnitude if the deficit itself is not a reliable individual-differences variable. In addition to the reliability of the AB, it is also important to examine how individual differences in AB performance are influenced by the type of target tasks employed. While a wide range of dual-target RSVP paradigms have been used in the literature, there has been limited work assessing whether all of these protocols measure the same underlying attention cost. To our knowledge, currently only three studies have directly examined the issue of stability in performance within and across different dual-target RSVP paradigms.

The first study to examine the reliability of the AB was that of McLaughlin, Shore, and Klein (2001), who examined the relationship between individual differences in performance on two different AB paradigms completed on different days. In one session, participants were tested under conditions in which they were asked to report two letter targets presented within a stream of digit distractors. They then returned four weeks later to complete a skeletal version of this task, such that only the targets and the posttarget pattern masks were presented (see also Duncan, Ward, & Shapiro, 1994). McLaughlin et al. observed that performance on the two different tasks correlated .66, and they concluded that individual AB performance is relatively stable over time and is consistent across different paradigm versions. This result was important, as in the skeletal version the AB targets were not presented in a traditional RSVP stream; thus, this study provided support for the underlying relationship between these two very different paradigms.

While McLaughlin et al. (2001) provided evidence for the reliability of the AB, their study examined the relationship between two *different* AB versions over time, rather than the stability of a single paradigm over time. As such, Dale and Arnell (in press) recently conducted a study designed to more directly examine the reliability of the AB deficit. They administered two different versions of the paradigm: one with a task switch between T1 to T2 ("Report the identity of the lone white T1 letter and report whether an 'X' was present or absent in the items that followed."), and one no-switch version in which the T1 and T2 tasks were the same ("Report the two red letters."). Fourty six participants completed both paradigms within a single testing session, and then returned 7–10 days later to complete the same protocols. The results showed that both versions were reliable within, and across, testing sessions, providing evidence for the reliability of AB performance (Dale & Arnell, in press). In addition, the individual AB magnitude for each paradigm was significantly correlated with that in the other paradigm, within and across testing sessions as well as overall (*r* = .56). These results suggest that, despite the obvious differences, these paradigms were measuring the same underlying process.

Kelly and Dux (2011) also examined the reliability of the AB within and across paradigms in a single testing session (with 43 participants), but found results that conflicted somewhat with those of Dale and Arnell (in press). They administered, twice within the same testing session, four different versions of the standard dual-target RSVP paradigm: categorical, featural, probe, and hybrid protocols. The categorical and featural paradigms were classified as "no-switch" paradigms, whereas the probe and hybrid paradigms were classified as "switch" paradigms. For the categorical paradigm, T1 and T2 were black letters embedded within digit distractors, and for the featural paradigm, T1 and T2 were red letters embedded within black distractors. In both cases,

Table 1 Relationships between measures of the attentional blink (AB) and performance on other cognitive tasks

	Task	AB Magnitude	T1	T2
Arnell, Stokes, MacLean, & Gicante (2010)[*]	Digit span (forward)	✗	✗	✗
	Digit span (backward)	✗	✗	✗
	Fluid intelligence (Raven's)	✗	✓	✓
	WM operation span	✓	✓	✓
Arnell, Howe, Joanisse, & Klein (2006)[*]	Delayed RT	✗	✗	✗
	Location probe	✗	✓	✗
	Manual RT	✗	✓	✓
	Rapid automatized naming	✗	✓	✓
	Vocal RT	✗	✓	✓
Arnell & Stubitz (2010)[*]	WM capacity (K)	✗	✗	✗
	WM filtering efficiency	✓	✗	✗
Colzato, Spapé, Pannebakker, & Hommel (2007)[†]	Fluid intelligence (Raven's)	✗	✓	✓
	WM operation span	✓	✗	✗
Dale & Arnell (2010)[*]	Global precedence	✓	✗	✗
Kawahara & Kihara (2011)[†]	Spatial visual search	✗	—	—
	Temporal visual search	✗	—	—
Klein, Arend, Beauducel, & Shapiro (2011)[±]	0-back task	—	✓	✓
	2-back task	—	✓	✗
	Fluid intelligence (Raven's)	—	—	✓
	Inspection time	—	✓	✗
	Predictive flexibility	—	—	✓
	Reactive flexibility	—	—	✓
	Sternberg search	—	✓	✓
	Visual search	—	✗	✗
Martens & Johnson (2009)[†]	Fluid intelligence (Raven's)	✗	✓	✗
	Letter span	✗	✗	✗
	Matrix span	✗	✗	✗
	Reading span	✗	✗	✗
	Symmetry span	✗	✗	✗

✗, n.s. correlation; ✓, significant correlation; —, correlation not reported. [*] Detection AB task, [†] Categorical AB task, [±] Featural AB task. In all cases, AB magnitude was calculated as long lag T2 accuracy minus short lag T2 accuracy. The exact methods and number of lags differed by study (see the publications in table for more details).

participants were required to report the identity of both targets. For the probe paradigm, T1 was a lone red letter that required identification, and T2 required an X/Y discrimination. In the hybrid paradigm, T1 was also a red digit that required identification, but T2 was one of 18 single black letters that had to be identified. Consistent with the findings discussed above (Dale & Arnell, in press; McLaughlin et al., 2001), all four dual-target RSVP deficits were reliably stable across testing blocks. However, in contrast to Dale and Arnell (in press), whereas individual AB magnitudes for the categorical paradigm were significantly correlated with those for the featural protocol, and individual AB magnitudes for the probe paradigm were significantly correlated with those for the hybrid paradigm, no significant relationships emerged between the switch and no-switch protocols. Importantly, this pattern of results was replicated across different samples.

This study provides evidence for a dissociation between switch and no-switch dual-target RSVP paradigms, suggesting that these protocols may not share the same underlying mechanisms/processes or measure the same AB costs. This is a compelling finding, because it highlights a potential issue with the common assumption that the type of RSVP paradigm used does not influence the results. Indeed, the results of Kelly and Dux (2011) support previous suggestions that the dual-task RSVP paradigms that contain a task set switch from T1 to T2 represent task switch costs, as opposed to the AB per se (e.g., Potter, Chun, Banks, & Muckenhoupt, 1998).

The main purpose of the present study was to examine the relationships among different AB paradigms using the full set of dual-task RSVP paradigms employed by both Dale and Arnell (in press) and Kelly and Dux (2011), but in a considerably larger sample than had been employed in these previous studies.

Whereas Kelly and Dux (2011) showed a dissociation between switch and no-switch paradigms, Dale and Arnell (in press) did not find evidence for this pattern. Indeed, Dale and Arnell (in press) found a relatively high correlation between a switch and a no-switch paradigm. Both Dale and Arnell (in press) and Kelly and Dux (2011) used the same featural paradigm ("report the two red letters") as a no-switch paradigm, but only the latter group of researchers employed a categorical dual-target RSVP paradigm. The switch paradigms also differed in the two studies. In Dale and Arnell (in press), the lone switch paradigm involved a true detection task in which T2 (an "X") was present only on some trials, and a present/absent response was required. In contrast, T2 was presented on every trial for the two switch paradigms used by Kelly and Dux (2011), and participants indicated which of the possible alternatives had been presented. Inclusion of all five paradigms from the two studies would allow us to examine whether the different choices of paradigms were responsible for the different pattern of results in the two studies.

To examine this issue, we used three of the AB paradigms (categorical, probe, and hybrid) from Kelly and Dux (2011), one of the two AB paradigms (detection) from Dale and Arnell (in press), and the featural paradigm used in both studies. To control for the possibility that the different results obtained by the two studies were due to population or lab differences, the present study was conducted simultaneously at both Brock University and The University of Queensland.

Three outcomes were possible. First, we might find that our three switch paradigms correlated and that our two no-switch paradigms correlated, but that the ABs observed in the two groups of paradigms do not relate to each other, as in Kelly and Dux (2011). This would indicate that some underlying process differs, depending on whether the dual-target RSVP protocol requires a switch between T1 and T2, and that switch and no-switch versions of the AB may not be measuring the same dual-task cost. Alternatively, all five versions of the AB might correlate with each other, as suggested by Dale and Arnell (in press), indicating no underlying differences for the ABs in switch and no-switch paradigms. Finally, we might find that the relationships among the five different AB protocols differ depending on the specific switch task that is used (n-AFC identification vs. detection). To preview our results, they supported a combination of the second and third hypotheses, such that while all five AB protocols were significantly correlated, the strengths of the relationships differed depending on the tasks that were used.

Method

Participants

Eighty eight undergraduate student volunteers from Brock University and 64 from The University of Queensland

participated for extra course credit or monetary compensation. Thirty-four of the participants were removed from the subsequent analysis for having T1 or T2|T1 long-lag accuracy that fell below the exclusion thresholds for one or more of the AB tasks (i.e., T1 accuracy less than 30 %, or T2|T1 long-lag accuracy less than three standard deviations from the mean).[1] This resulted in a total of 118 participants. The participants ranged in age from 17 to 41 years (M = 20.03, SD = 3.3), reported normal or corrected-to-normal vision, and reported having learned English before the age of 8. All of the participants completed a 2-h testing session, followed by a second, identical 2-h testing session 7 to10 days later. This sample gave us a power of ~ .8 to detect relationships ≥0.25 (Cohen, 1992).

Apparatus

The computerized tasks were controlled, and the data recorded, using either E-Prime software (at Brock University) or MATLAB with the Psychophysics Toolbox extension (Brainard, 1997; Pelli, 1997; at The University of Queensland), running on PC desktop (Brock) or Mac Mini (Queensland) computers and CRT monitors. In all cases, the participants made responses via manual buttonpresses on the computer keyboard.

Stimuli and design

The participants completed five different versions of dual-target RSVP (all of which are common in the literature): categorical, featural, alternative forced choice (AFC), detection, and hybrid (see Fig. 1). The presentation order was counterbalanced across participants and testing sessions. For all of the five protocols, each trial began with a black fixation cross or square that was presented in the center of a white screen for 1,000 ms. The fixation marker was then replaced by the first item in the RSVP stream. Each stimulus was presented individually on the screen for 100 ms with no blank interstimulus interval (ISI). In all five paradigms, each RSVP stream contained two target items and 11 distractor items, for a total of 13 items. All of the items were presented in Courier New font and subtended approximately 1° of visual angle at an unrestrained viewing distance of 57 cm. T1 always appeared in Position 5, and T2 was separated

---

[1] While most AB studies have made use of one or two different AB paradigms and a single session, our study contained five different AB paradigms administered at two different time points. As such, our participants had ten different opportunities to fall short of criterion on T1 or T2. If a participant did not reach the cutoff threshold for even one of the targets on one of the ten paradigms, their data for all of the paradigms were excluded, so that data from the same participants would be included in all correlations. This resulted in a larger-than-usual proportion of participants being removed from the study.
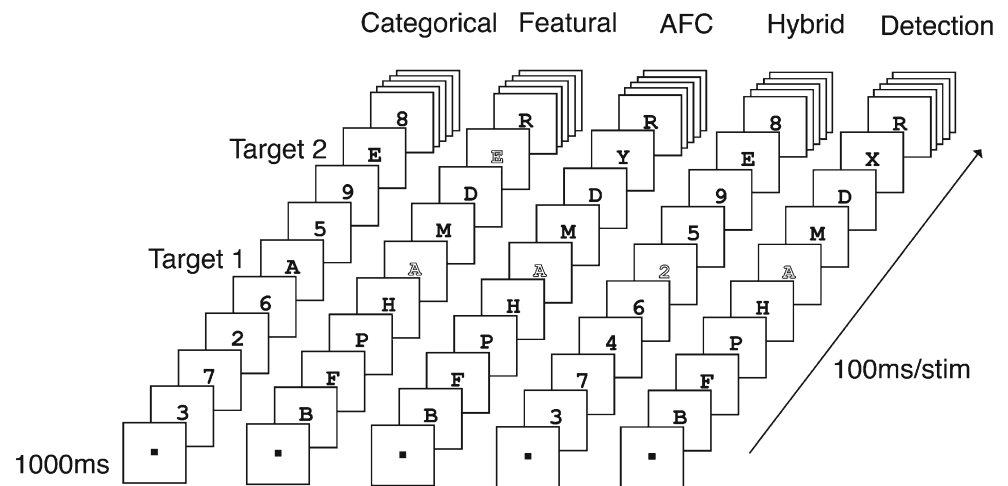
from T1 by a lag of two (200 ms), three (300 ms), five (500 ms), or six (600 ms) items.

Each of the five paradigms began with 12 practice trials, after which participants completed four blocks of 24 trials apiece, for a total of 96 test trials per protocol. Breaks were provided in-between runs and when necessary. Responses were not speeded and were always made at the end of each trial, following onscreen prompts for T1 and then T2.

For the categorical paradigm, both T1 and T2 were single black letters drawn randomly from all of the letters of the alphabet, excluding I, L, O, Q, U, V, X, and Y. The distractors were black digits in the range from 2 to 9. Participants were instructed to report the identity of the two letters presented on each trial, guessing if unsure.

For the featural paradigm, T1 and T2 were the same as in the categorical paradigm, but were presented in red font. The distractors were single black letters drawn from the same pool as T1 and T2, without repetition. Participants were instructed to report the identity of the two red letters, guessing if unsure.

For the AFC protocol, T1 was a single red letter drawn from all of the letters of the alphabet, excluding I, L, O, Q, U, V, X, and Y. T2 was a single black "X" on 50 % of trials and a single black "Y" on 50 % of trials. The distractors were single black letters drawn from the same pool as the T1 letters, without repetition. Participants were instructed to report the identity of the lone red letter and to determine whether the X or Y had appeared.

The detection paradigm was nearly identical to the AFC task, with the exception that T2 was a single black "X" presented on 50 % of the trials, with a random letter replacing the "X" on the other 50 % of trials. Participants were instructed to report the identity of the red letter and to report whether the "X" was present or absent from the RSVP stream. To try to equate false alarms somewhat across participants, each participant was told to indicate that T2 was present only if they were fairly confident that it had appeared on the screen.

For the hybrid paradigm, T1 was a red digit drawn from the numbers 2–9, and T2 was a single black letter drawn from all of the letters of the alphabet, excluding I, L, O, Q, U, V, X, and Y. The distractors were black digits drawn from the numbers 2–9. Participants were instructed to report the identity of the red digit and the identity of the lone letter.

Results

AB performance

For each participant, an overall T1 accuracy score was calculated for each paradigm and session (averaged across lags; see Table 2 for the means). For the categorical and featural versions, T1 accuracy was scored without regard for order errors. A repeated measures analysis of variance (ANOVA) with Paradigm and Session as within-subjects factors was performed on the overall T1 accuracy data. We found significant main effects of paradigm, $F(4, 468) = 81.79$, $p < .001$, and session, $F(1, 117) = 12.74$, $p = .001$. These variables also interacted significantly, $F(4, 468) = 10.05$, $p < .001$, with T1 accuracy increasing slightly from Session 1 to 2 for the categorical protocol, but decreasing slightly for all other paradigms (see Table 2).

Table 2 T1 accuracy by session for all AB tasks

	Session 1		Session 2	
	M	SD	M	SD
Categorical	.89	.08	.91	.08
Featural	.83	.11	.79	.13
AFC	.79	.10	.77	.11
Detection	.83	.11	.79	.13
Hybrid	.75	.14	.74	.15

For the detection paradigm, a measure of T2 sensitivity was obtained by subtracting each participant's overall false alarm rate from their T2 hit rate for each lag. The mean T2 false alarm rates in the detection protocol were .15 (SD = .11) for Session 1 and .14 (SD = .10) for Session 2. The same pattern of results was observed when d' was used for the detection task, instead of hits – false alarms. For all other protocols, the T2 accuracy for each lag was obtained by simply averaging T2 accuracy across trials for each lag. In all cases, T2 accuracy/sensitivity scores were only included for trials in which T1 was identified correctly.

A repeated measures ANOVA with Lag and Session as within-subjects factors was performed on the T2|T1 accuracy/sensitivity scores for each of the five paradigms. In all cases, a significant main effect of lag emerged, indicating the presence of an AB, all ps < .001 (see Fig. 2A–E). We also found a significant main effect of session for all of the protocols, with the exception of the featural paradigm, indicating that T2 performance generally improved over time (i.e., a practice effect), all ps < .001. Finally, for the categorical paradigm, a significant interaction of lag and session was apparent, $F(3, 3515) = 5.94$, $p = .001$, in that the lag effect was reduced in Session 1 as compared to Session 2. However, no such interaction was found for any of the other four AB protocols.

When Lab (Brock, Queensland) was added as a factor to the analyses above, T2 accuracy rates were about 5 % to 10 % higher at Brock for each of the five protocols (all ps < .01), and the lag effect was significantly smaller for Brock participants than for Queensland participants for four of the five protocols (ps < .01 for the Lag × Lab interaction for all protocols except AFC, for which $p = .163$ with the same pattern of means). Therefore, overall, Brock participants had higher T2 accuracy rates and smaller ABs. Nevertheless, the key result was that at both testing sites significant ABs were observed in all of the dual-target RSVP paradigms.

Internal-consistency reliability

The internal-consistency reliability of each paradigm and session was examined by performing a split-half reliability analysis on the data. For each protocol and session, the data were first split into odd and even trials. Then, a measure of AB magnitude was obtained by taking each participant's mean short-lag (2 and 3) T2 accuracy and statistically removing the variability of their long-lag (5 and 6) T2 accuracy, by regressing out the mean long-lag accuracy from the mean short-lag accuracy and saving the residual variability. This was completed for each combination of paradigm, session, and odd/even trials. This method controls for

Fig. 2 T2|T1 accuracy as a function of T1–T2 lag and session for the categorical (A), featural (B), AFC (C), hybrid (D), and detection (E) AB paradigms. Error bars represent the standard error of the difference between Sessions 1 and 2 at each lag (Franz & Loftus, 2012)
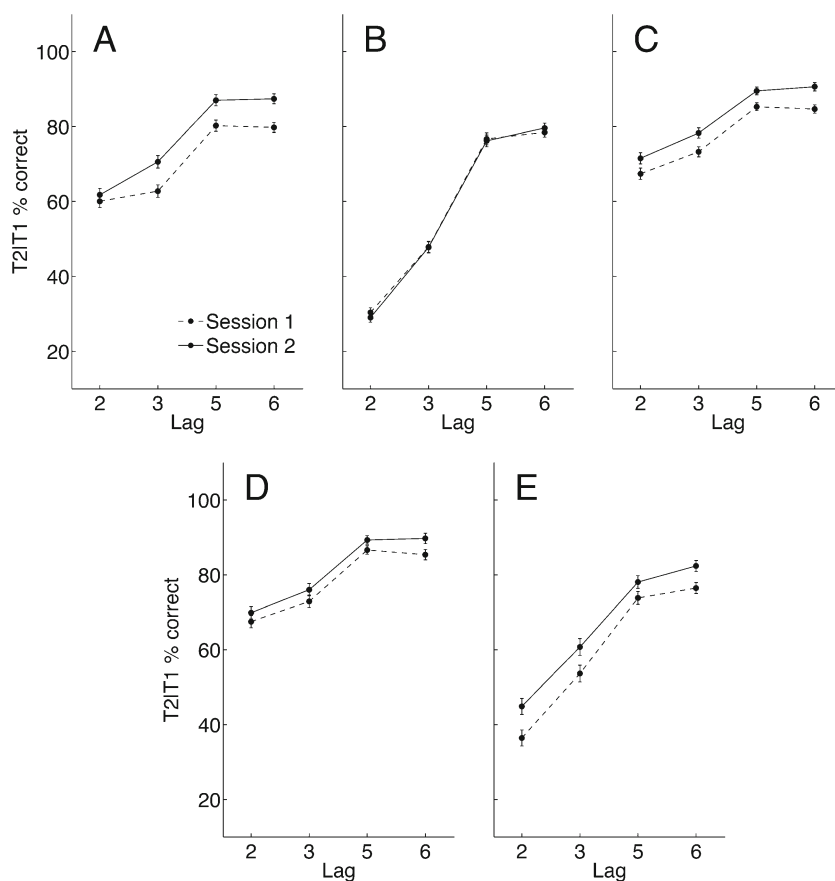
individual differences in T2 accuracy that are unrelated to the AB, without the need for a difference score. Long-minus short-lag difference scores would be correlated with T2 accuracy at both short and long lags, as opposed to the residual, which is correlated only with short-lag accuracy (see MacLean & Arnell, 2012, for a full discussion of estimating an individual's dual-target RSVP deficit magnitude). However, it is worth noting that the same pattern and significance of results was obtained when difference scores were used, although the correlations were slightly lower (as would be expected, given that the difference score also includes long-lag variability).

Performance for the odd trials was then correlated with performance on the even trials (see Table 3). A Spearman–Brown procedure (Spearman, 1904) was used to correct for correlation attenuation due to using only half of the trials, thus allowing for a reliability estimate of the entire paradigm (Nunnally, 1978). The internal-consistency reliability was found to be reasonably high for all five protocols in Session 1 and Session 2, suggesting that these AB tasks were reliable within sessions. Importantly, this pattern of results was consistent across the two labs (Queensland and Brock), and these relationships were significant even when covarying out variability due to the lab.

Test–retest reliability

The test–retest reliability of each of the measures was also examined. Using the same method as described in the previous section, a measure of AB magnitude was obtained for each combination of paradigm and testing session. A correlational analysis was then performed to examine the relationship between this deficit for Sessions 1 and 2 for each protocol. All five paradigms were shown to have acceptable to high test–retest reliability, with numbers that were similar in magnitude to the corrected within-session reliabilities and the test–retest reliabilities found by Dale and Arnell (in press) and Kelly and Dux (2011) (see Table 3 and Fig. 3). Again, this pattern of results

was consistent across the two labs, and all relationships were observed even when controlling for variability due to labs.

Relationships among the tasks

Next, the relationships among the five different AB protocols were examined. First, the measures of AB magnitude were collapsed across session for each paradigm, to obtain an overall AB magnitude score. The correlations among the protocols were then examined (see Table 4, and Fig. 4). All of the paradigms were significantly positively correlated with each other, and in many cases the size of the correlations between protocols approximated the reliabilities of those protocols, suggesting that a task was as related to another task as it was to itself. Contrasts of correlated correlation coefficients (Meng, Rosenthal, & Rubin, 1992) demonstrated that the relationship observed between the detection and categorical paradigms was significantly smaller than the relationship observed between the categorical and featural paradigms and than the relationship between the AFC and detection paradigms ($zs > 2.14$, $ps < .05$). No other pairs of correlations differed significantly.

To further explore these relationships, AB magnitude scores were broken down by session, and within-session relationships among the protocols were examined (see Table 5). The same general pattern of results obtained with the overall data was also observed within each session, with the exception that the categorical and detection paradigms were no longer significantly correlated for Session 1.

When the cross-session relationships among the protocols were examined (i.e., Session 1 data for one paradigm predicting Session 2 data from another paradigm), the same general pattern of results was observed as when the protocols were averaged across sessions. The only exceptions were that three of the relationships (Session 1 categorical with Session 2 detection, Session 1 detection with Session 2 categorical, and Session 1 AFC with Session 2 hybrid) now fell short of significance ($ps > .10$).

Table 3 Internal-consistency reliability, with Spearman–Brown corrected correlations of AB magnitude for each task by session, as well as test–retest reliability of the AB magnitude

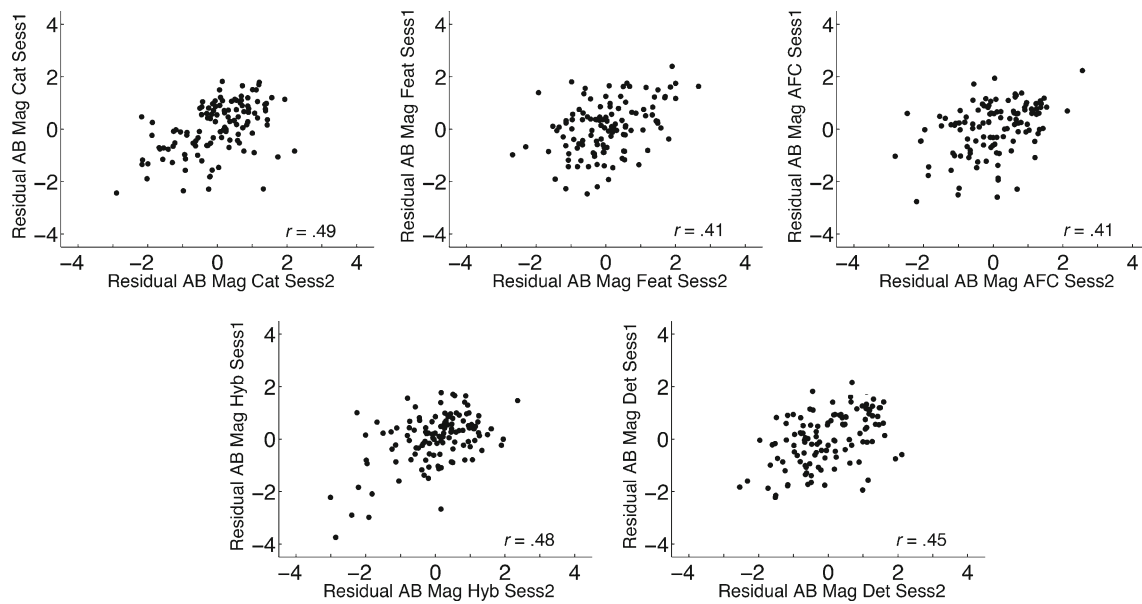|	Session 1		Session 2		Test–Retest r
	Split-Half r	Corrected r	Split-Half r	Corrected r	
Categorical	.37	.54	.37	.54	.49
Featural	.42	.59	.39	.56	.41
AFC	.48	.65	.44	.61	.41
Detection	.38	.55	.38	.55	.45
Hybrid	.61	.76	.57	.73	.48

All $ps < .001$.

Fig. 3 Scatterplots for the test–retest reliability of each task. Attentional blink magnitude (AB Mag) was obtained by taking each participant's mean short-lag (2 and 3) T2|T1 accuracy and statistically controlling for the influence of their mean long-lag (5 and 6) T2|T1 accuracy. Sess1 = Session 1; Sess2 = Session 2; Cat = categorical paradigm; Feat = featural paradigm; AFC = alternative forced choice paradigm; Hyb = hybrid paradigm; Det = detection paradigm

Generally, we found no differences in the patterns of relationships (within or across tasks) from the two labs. Significant positive correlations were found within and across all the paradigms, with the exception of a nonsignificant relationship between the AB deficits observed in the categorical and detection protocols (both labs) and a nonsignificant relationship between the ABs in the categorical and AFC protocols and the hybrid and AFC paradigms (Brock only). Importantly, though, none of the correlations between the Queensland and Brocks labs were statistically different (Fisher z-test $ps > .10$). Furthermore, all relationships remained significant (and largely unchanged) when controlling for variability due to the labs, with the exception of the relationship between the categorical and detection paradigms, which now fell just below significance ($p = .074$).

Factor analysis

Finally, an experimental factor analysis was performed on the overall (collapsed across sessions) dual-target deficit magnitude measures for each of the five paradigms, in order to examine whether certain protocols grouped together statistically. The factors were extracted using principal components analysis, and a varimax rotation with a Kaiser normalization was performed. A one-factor solution was obtained, which explained 52 % of the cumulative variance in the data, with each of the paradigms loading quite highly on the single factor (see Table 6). This indicates that the five protocols were explaining similar variance, and suggests that all five AB

versions may share the same underlying process. However, when a two-factor solution was employed, an additional 18 % of variability was accounted for, and the categorical, featural, and hybrid tasks loaded highly onto the first factor, while the AFC and detection tasks loaded onto the second factor (see Table 7). This suggests that, while all five tasks do share underlying variability, task-specific characteristics do contribute some additional variability. This point will be returned to in the Discussion section.

Table 4 Pearson r correlations among AB magnitude measures

	1	2	3	4
Residuals[a]				
1. Categorical	–			
2. Featural	.43**	–		
3. AFC	.30**	.43**	–	
4. Detection	.21*	.34**	.61**	–
5. Hybrid	.49**	.41**	.32**	.44**
Difference Scores[b]				
1. Categorical	–			
2. Featural	.37**	–		
3. AFC	.27**	.34**	–	
4. Detection	.20*	.22*	.51**	–
5. Hybrid	.49**	.31**	.24**	.41**

* $p < .05$; ** $p < .001$.
[a] Presents the correlations using the residual score
[b] Presents the correlations using difference scores

Springer

Fig. 4 Scatterplots of the relationship between individuals' overall AB magnitudes (Comb = mean across Sessions 1 and 2) for each combination of the five tasks
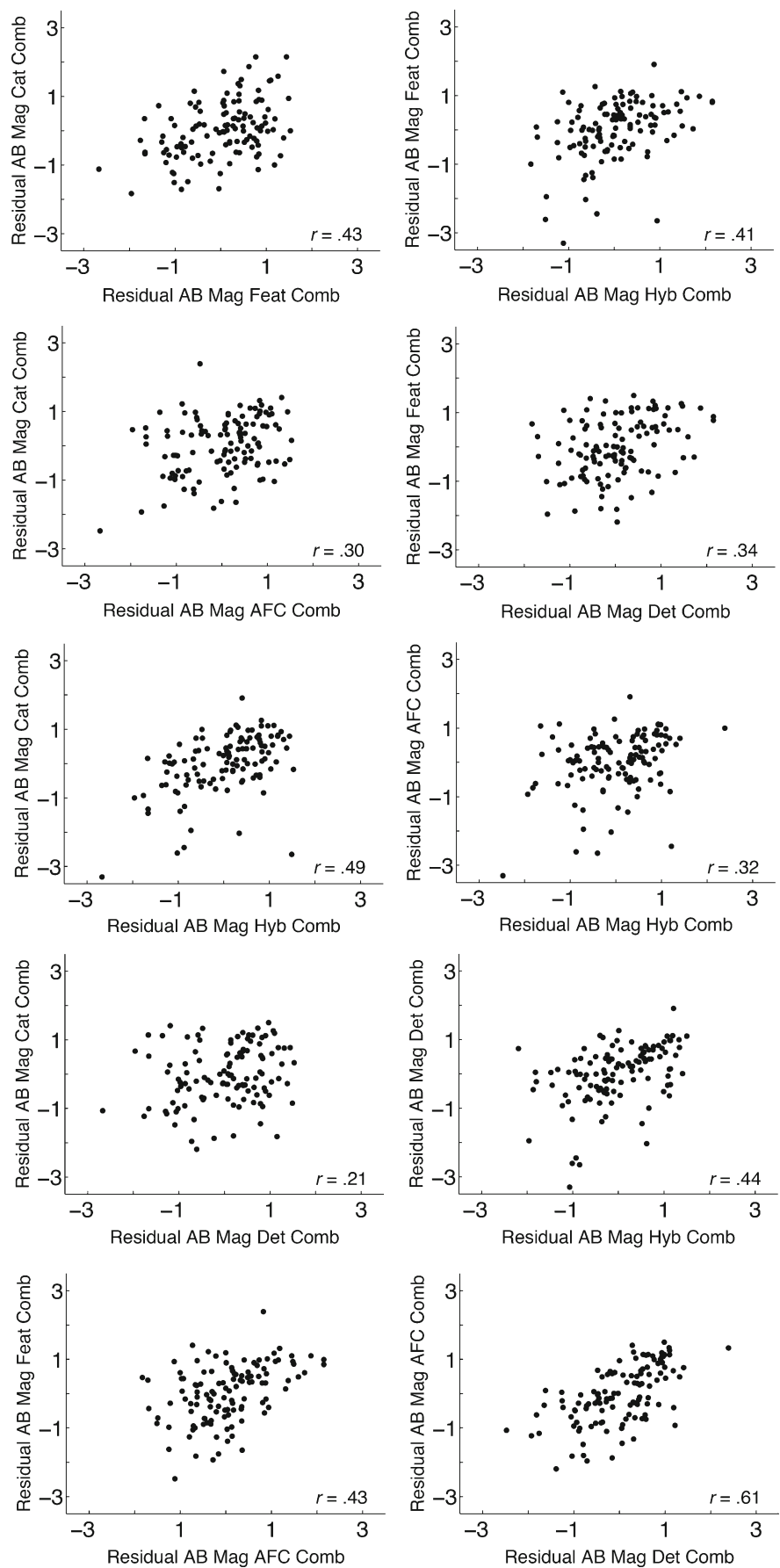
Table 5 Pearson r correlations among AB magnitude measures by session

	1	2	3	4
Session 1				
1. Categorical	–			
2. Featural	.35**	–		
3. AFC	.20*	.34**	–	
4. Detection	.13	.22*	.42**	–
5. Hybrid	.37**	.33**	.20*	.33**
Session 2				
1. Categorical	–			
2. Featural	.34**	–		
3. AFC	.25**	.22*	–	
4. Detection	.24**	.25**	.46**	–
5. Hybrid	.47**	.34**	.33**	.41**

$^* p < .05$; $^{**} p < .001$.

Discussion

Within- and across-session reliability

One of the purposes of this study was to replicate and extend previous findings (i.e., Dale & Arnell, in press; Kelly & Dux, 2011; McLaughlin et al., 2001) regarding the reliability of the AB paradigm, both within a single testing session and over the course of 7–10 days. As hypothesized, we were able to show that AB magnitudes from all five protocol versions had acceptably high internal-consistency reliability for both Sessions 1 and 2, demonstrating the stability of AB performance within testing session and replicating the results of Dale and Arnell (in press) and Kelly and Dux (2011). Additionally, all five AB magnitude estimates showed moderate to high test–retest reliability, demonstrating that dual-target RSVP T2 deficits are stable over the course of at least a week. These findings once again provide evidence for the stability of individual dual-target RSVP deficits.

These findings are of particular importance, given the recent interest in predicting individual differences in AB performance. The relationship between performance on two experimental protocols is naturally restricted by the reliability of these protocols, such that the correlation between them cannot be higher than their reliability estimates (Nunnally, 1978). As such, our findings provide an estimate

Table 6 Factor loadings for a one-factor solution with the five AB tasks, collapsed across session

	Component 1
Categorical	.66
Featural	.72
AFC	.75
Detection	.73
Hybrid	.74

Table 7 Rotated factor loadings for a two-factor solution with the five AB tasks, collapsed across session

	Component	
	1	2
Categorical	.88	.03
Featural	.65	.37
AFC	.23	.85
Detection	.18	.88
Hybrid	.74	.29

of the upper bound for the relationships among AB paradigms and other individual-differences variables, and provide a degree of confidence for using performance on a variety of AB paradigms as individual-differences measures.

Relationships among the AB paradigms

In addition to estimating the reliability of our AB protocols within session and over time, we were particularly interested in examining the relationships among our five different AB magnitude estimates. Previous research had been inconclusive with regard to the relationships among different dual-target RSVP paradigms. A recent study (Kelly & Dux, 2011) suggested that performance on paradigms that involve an attentional-set/task switch between T1 and T2 was distinct from performance on those paradigms that do not have such a switch, whereas another suggested, conversely, that performance on switch and no-switch AB protocols was significantly correlated and reflected the same underlying variability (Dale & Arnell, in press). Due to this discrepancy, it was important to reexamine the relationships among these paradigms.

An examination of the relationships among our dual-target RSVP paradigms revealed statistically significant correlations for AB magnitude among all five protocols overall. This provides support for Dale and Arnell (in press), who showed similar correlations between a detection and a featural AB paradigm, and suggests that, despite the obvious differences, many of the paradigms used to study the AB tap the same underlying variability. This conclusion was also supported by our factor analysis, in which all five paradigms loaded highly on a single common factor that presumably represents the AB. When combined with the results of McLaughlin et al. (2001), these findings allow us to now conclude that all versions of the AB paradigms share common variability, regardless of the presence or absence of task switches from T1 to T2. However, in addition to the variability shared among the protocols, there was also some variability that was not shared, especially for some paradigm pairs. Recall that Kelly and Dux (2011) argued that switch and no-switch protocols tap into distinct underlying cognitive processes. While this position was not supported

by the present data, which showed significant relationships among all of the paradigm pairs, we did find evidence that this switch-versus-no-switch distinction may have influenced the strength of the relationships observed here. The overall correlation found between the seemingly most dissimilar pair of paradigms (categorical and detection) was smaller than that for the protocols with the greatest overlap —the categorical and featural paradigms and the AFC and detection paradigms. This pattern of results suggests that all five ABs do tap the same underlying process to some extent, but that additional task-switch variability is included in the AB estimate for switch paradigms, variability that is unrelated to the AB. The considerable difference between our sample size and that of Kelly and Dux (2011) is potentially the reason why they consistently found no significant relationships between switch and no-switch paradigms—a pattern that they replicated across two experiments.

Over and above the switch/no-switch distinction, the protocols differed on a number of other dimensions, including the similarity between the targets and the distractors, the saliency of the targets, and the number of response alternatives. The protocols may also have differed in terms of factors such as resource consumption, which we did not measure directly here. For example, the categorical and detection paradigms had the weakest relationship with each other and also did not share any superficial similarities in the actual target tasks used. Specifically, the categorical protocol was the only one of the five that did not contain some sort of featural task, and the detection paradigm was the only task that contained a true present/absent judgment. Additionally, the categorical and the hybrid protocols both required participants to identify a T2 letter presented among digits, and they produced one of the strongest relationships. Thus, similarities among the target and distractor items used across the protocols could have influenced the size of the correlations between these different paradigms.

Similarly, the detection and AFC paradigms both contained 2-AFC T2 tasks (the other protocols each had 18 T2 options), and this was the most strongly correlated pair. Interestingly, when a two-factor solution was specified for our factor analysis, the resultant components clearly divided in such a way that protocols that allow for few T2 response alternatives (i.e., the detection and AFC tasks) loaded highly on one factor, whereas protocols that have more T2 (18+) response alternatives (i.e., the categorical, featural, and hybrid paradigms) loaded highly on another, suggesting a distinction between protocols that contain many response alternatives and those that contain few (see Table 7). Therefore, the number of response alternatives may introduce additional variability into the AB that can influence the size of the correlations among different paradigms.

In conclusion, the present results are encouraging, as they provide compelling evidence for the reliability of the AB in multiple dual-target RSVP task versions and for the stability of AB performance over time. These results also demonstrate that multiple distinct paradigms appear to share underlying variability that can confidently be labeled as the AB. This finding has theoretical implications because it suggests that, although task-set reconfiguration between T1 and T2 modulates lag-1 sparing (as was shown by Visser, Bischof, & Di Lollo, 1999), it does not appear to compromise measures of AB magnitude, suggesting that the mechanism underlying the AB is not influenced by task-set reconfigurations. Additionally, some researchers have previously suggested that when there is a task switch from T1 to T2, the resultant dual-task deficit reflects a switch cost rather than the AB (Kelly & Dux, 2011; Potter et al., 1998). In contrast, the present findings provide support for the use of multiple AB paradigm versions across the literature.

References

Arnell, K. M., Howe, A. E., Joanisse, M. F., & Klein, R. M. (2006). Relationships between attentional blink magnitude, RSVP target accuracy, and performance on other cognitive tasks. *Memory & Cognition, 34,* 1472–1483. doi:10.3758/BF03195912

Arnell, K. M., Stokes, K., MacLean, M. H., & Gicante, C. (2010). Executive control processes of working memory predict attentional blink magnitude over and above storage capacity. *Psychological Research, 74,* 1–11. doi:10.1007/s00426-008-0200-4

Arnell, K. M., & Stubitz, S. M. (2010). Attentional blink magnitude is predicted by the ability to keep irrelevant material out of working memory. *Psychological Research, 74,* 457–467. doi:10.1007/s00426-009-0265-8

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433–436. doi:10.1163/156856897X00357

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159. doi:10.1037/0033-2909.112.1.155

Colzato, L. S., Slagter, H. A., de Rover, M., & Hommel, B. (2011). Dopamine and the management of attentional resources: Genetic markers of striatal D2 dopamine predict individual differences in the attentional blink. *Journal of Cognitive Neuroscience, 23,* 3576–3585. doi:10.1162/jocn_a_00049

Colzato, L. S., Spapé, M. M. A., Pannebakker, M. M., & Hommel, B. (2007). Working memory and the attentional blink: Blink size is predicted by individual differences in operation span. *Psychonomic Bulletin & Review, 14,* 1051–1057. doi:10.3758/BF03193090

Dale, G., & Arnell, K. M. (2010). Individual differences in dispositional focus of attention predict attentional blink magnitude. *Attention, Perception, & Psychophysics, 72,* 602–606. doi:10.3758/APP.72.3.602

Dale, G., & Arnell, K. M. (in press). How reliable is the attentional blink? Examining the relationships within and between attentional

blink tasks over time. *Psychological Research*. doi:10.1007/s00426-011-0403-y

Duncan, J., Ward, R., & Shapiro, K. (1994). Direct measurement of attentional dwell time in human vision. *Nature, 369,* 313–315. doi:10.1038/369313a0

Dux, P. E., & Marois, R. (2008). Distractor inhibition predicts individual differences in the attentional blink. *PLoS One, 3,* e3330. doi:10.1371/journal.pone.0003330

Dux, P. E., & Marois, R. (2009). The attentional blink: A review of data and theory. *Attention, Perception, & Psychophysics, 71,* 1683–1700. doi:10.3758/APP.71.8.1683

Franz, V. H., & Loftus, G. R. (2012). Standard errors and confidence intervals in within-subjects designs: Generalizing Loftus and Masson (1994) and avoiding the biases of alternative accounts. *Psychonomic Bulletin & Review, 19,* 395–404. doi:10.3758/s13423-012-0230-1

Kawahara, J.-I., & Kihara, K. (2011). No commonality between attentional capture and attentional blink. *Quarterly Journal of Experimental Psychology, 64,* 991–1008. doi:10.1080/17470218.2010.524304

Kelly, A. J., & Dux, P. E. (2011). Different attentional blink tasks reflect distinct information processing limitations: An individual differences approach. *Journal of Experimental Psychology. Human Perception and Performance, 37,* 1867–1873. doi:10.1037/a0025975

Klein, C., Arend, I. C., Beauducel, A., & Shapiro, K. L. (2011). Individuals differ in the attentional blink: Mental speed and intra-subject stability matter. *Intelligence, 39,* 27–35. doi:10.1016/j.intell.2010.11.004

MacLean, M. H., & Arnell, K. M. (2010). Personality predicts temporal attention costs in the attentional blink paradigm. *Psychonomic Bulletin & Review, 17,* 556–562. doi:10.3758/PBR.17.4.556

MacLean, M. H., & Arnell, K. M. (2012). A conceptual and methodological framework for measuring and modulating the attentional blink. *Attention, Perception, & Psychophysics, 74,* 1080–1097. doi:10.3758/s13414-012-0338-4

MacLean, M. H., Arnell, K. M., & Busseri, M. (2010). Dispositional affect predicts temporal attention costs in the attentional blink

paradigm. *Cognition and Emotion, 24,* 1431–1438. doi:10.1080/02699930903417897

MacLean, M. H., Arnell, K. M., & Cote, K. A. (2012). Resting EEG in alpha and beta bands predicts individual differences in attentional blink magnitude. *Brain and Cognition, 78,* 218–229. doi:10.1016/j.bandc.2011.12.010

Martens, S., & Johnson, A. (2009). Working memory capacity, intelligence, and the magnitude of the attentional blink revisited. *Experimental Brain Research, 192,* 43–52. doi:10.1007/s00221-008-1551-1

Martens, S., & Valchev, N. (2009). Individual differences in the attentional blink: The important role of irrelevant information. *Experimental Psychology, 56,* 18–26. doi:10.1027/1618-3169.56.1.1

McLaughlin, E. N., Shore, D. I., & Klein, R. M. (2001). The attentional blink is immune to masking induced data limits. *Quarterly Journal of Experimental Psychology, 54A,* 169–196. doi:10.1080/02724980042000075

Meng, X.-L., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin, 111,* 172–175. doi:10.1037/0033-2909.111.1.172

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442. doi:10.1163/156856897X00366

Potter, M. C., Chun, M. M., Banks, B. S., & Muckenhoupt, M. (1998). Two attentional deficits in serial target search: The visual attentional blink and an amodal task-switch deficit. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 979–992. doi:10.1037/a0025975

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology. Human Perception and Performance, 18,* 849–860. doi:10.1037/0096-1523.18.3.849

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology, 15,* 72–101.

Visser, T. A. W., Bischof, W. F., & Di Lollo, V. (1999). Attentional switching in spatial and nonspatial domains: Evidence from the attentional blink. *Psychological Bulletin, 125,* 458–469. doi:10.1037/0033-2909.125.4.458