Failures of sustained attention in life, lab, and brain: Ecological validity of the SART

Daniel Smilek, Jonathan S. A. Carriere, J. Allan Cheyne

Department of Psychology, University of Waterloo, Canada

Abstract

The Sustained Attention to Response Task (SART) is a widely used tool in cognitive neuroscience increasingly employed to identify brain regions associated with failures of sustained attention. An important claim of the SART is that it is significantly related to real-world problems of sustained attention such as those experienced by TBI and ADHD patients. This claim is largely based on its association with the Cognitive Failures Questionnaire (CFQ), but recently concerns have been expressed about the reliability of the SART–CFQ association. Based on a review of the literature, meta-analysis of prior research, and analysis of original data, we conclude that, across studies sampling diverse populations and contexts, the SART is reliably associated with the CFQ. The CFQ–SART relation also holds for patients with TBI. We note, however, conceptual limitations of using the CFQ, which was designed as a measure of general cognitive failures, to validate the SART, which was specifically designed to assess sustained attention. To remedy this limitation, we report on associations between the SART and a specific Attention-Related Cognitive Errors Scale (ARCES) and a Mindful Awareness of Attention Scale—Lapses Only (MAAS-LO).

E-mail address: dsmilek@uwaterloo.ca (D. Smilek)

Introduction

The Sustained Attention to Response Task (SART; Robertson, Manly, Andrade, Baddeley, & Yiend, 1997) is widely used as a behavioral measure of sustained attention failures. The SART requires participants to respond to a sequentially presented series of digits (1 through 9) and to withhold a response when an infrequent critical NOGO digit appears (e.g., “3”). The SART has been used to investigate a variety of neuropsychological conditions including traumatic brain injury (TBI; Dockree et al., 2004, Manly et al., 2004, O’Keeffe et al., 2004, Robertson et al., 1997 and Whyte et al., 2006), ADHD (Bellgrove et al., 2006, Bellgrove et al., 2005, Johnson et al., 2007a, Johnson et al., 2007b, Manly et al., 2001 and Mullins et al., 2005), and depression (Smallwood, O’Connor, Sudberry, & Obosawin, 2007). It has also been used to study the neurophysiology of sustained attention, implicating areas such as the anterior cingulate cortex (ACC; Cheyne, Cheyne, Bells, Carriere, & Smilek, 2009) and both dorsomedial and ventromedial prefrontal cortices, which are two areas associated with the default network (Christoff, Gordon, Smallwood, Smith, & Schooler, 2009). The fundamental assumption underlying all of these studies is that performance on the SART is an externally valid measure of an individual's propensity to experience sustained attention failures in everyday life.

The need for external validation of any new neuropsychological tool is a critical component of the neuroscience of attention (see Kingstone et al., 2008 and Kingstone et al., 2003). This need was clearly recognized by the developers of the SART in their initial presentation, in which this issue was addressed at some length (Robertson et al., 1997). The validation of the SART as a measure of failures of sustained attention was based, in that paper, on several positive correlations obtained between SART
Failures of sustained attention in life, lab, and brain: Ecological validity of...

In the present paper we address several empirical and conceptual issues related to the reliability and validity of the SART as a measure of everyday failures of sustained attention. First, we assess the empirical association between the SART and the CFQ by conducting a meta-analysis of published studies of the association between the two measures. Second, we report on original data from a large ($N = 363$) heterogeneous sample assessing the empirical relation between the SART and the CFQ. Finally, we evaluate conceptual arguments for validity for the SART as a measure of sustained attention.

The SART was developed with the intention of providing a brief, reliable, and valid measure of failures of sustained attention (Robertson et al., 1997). Robertson et al. (1997) defined sustained attention as self-sustained (i.e., endogenously managed without external supports), conscious, task-relevant processing during monotonous tasks, which encourage automatic, mindless responding and susceptibility to internal and external distracters that lead to off-task and potentially interfering cognitions. A distinctive feature of the SART is that it requires the automatic response to be the “default” condition, thereby allowing for the development of a habitual response pattern that must be periodically overridden by a conscious executive decision. Hence, the continuation of the habitual response on a NOGO trial is taken as a task-related consequence of a failure of sustained attention, detected by the failure to note the NOGO signal with sufficient rapidity to prevent the habitual response. Thus, the critical, though indirect, attention failure measure yielded by the SART is a count of the failures to withhold a response when presented with a relatively rare (1 in 9) NOGO signal. We have argued that a more direct measure of lapses of attention is a speeding of response time, revealing unconscious automatic responding during frequent GO trials (Cheyne et al., 2006 and Cheyne et al., 2009b). SART errors on NOGO trials are therefore presaged by decreasing reaction times (RTs) in the immediately preceding GO trials (Cheyne et al., 2006, Cheyne et al., 2009a, Cheyne et al., 2009b, Farrin et al., 2003, Manly et al., 1999, Robertson et al., 1997 and Smallwood et al., 2007). The SART therefore provides putative measures of both attention lapses and behavioral attention-related errors during such lapses. Robertson et al. (1997) also provided evidence of good test–retest stability of SART error rates over a period of two weeks ($r = .76$), suggesting individual SART performance is relatively stable over time.

In an effort to demonstrate real-world implications for the SART and to provide evidence of the external validity of the SART, Robertson et al. (1997) examined the relation between the SART and the CFQ, a survey instrument previously developed by Broadbent et al. (1982). The CFQ was based on the pioneering work of (Reason, 1977) and (Reason, 1979) on everyday cognitive errors and action slips. Reason argued, based on analysis of real-world accident and incident reports, that minor attentional errors in routine, overlearned tasks often had far-reaching consequences. Hence, items were selected for the CFQ by sampling a variety of memory, attention, and action slips and errors from a variety of quotidian settings of the sort encountered by people in their everyday lives. It is important to note that the CFQ was designed to sample a broad array of cognitive processes and everyday tasks, largely focusing on attention and memory failures, but also including actions slips (dropping and bumping into things) that might – or might not – result from attention or memory failures. Broadbent and colleagues reported that the CFQ errors for both TBI patients (when CFQ ratings were provided by informants) and controls. A number of studies have subsequently attempted to replicate the original Robertson et al. finding. One recent study by Whyte et al. (2006) reported a failure to replicate the original Robertson finding and, based on a review of the limited literature available, raised questions about the validity of the SART. A list of relevant studies and their key findings with regard to the SART–CFQ relations are shown in Table 1. The studies have been quite diverse in terms of populations sampled, procedures used, and data analysis (see Table 1).

Moreover, some studies carried out their analyses combining different clinical or quasi-clinical groups defined by attentional and/or affective (depression) problems (Farrin et al., 2003 and Van der Linden et al., 2005), or by examining group differences (Manly et al., 1999), whereas others report analysis for such groups separately (Whyte et al., 2006). The former approach will likely inflate correlations (but just in the case the null is false), whereas the latter will depress them because of restricted range. Consistent with the foregoing statistical considerations Manly et al. (1999), Van der Linden et al. (2005) and Farrin et al. (2003) found strong support for the Robertson claim, whereas Whyte and colleagues did not. It is also worth highlighting that small sample sizes can produce unreliable and misleading results because the confidence intervals around the observed correlations would be relatively large. To this point, it is notable that the Whyte et al. study was also one of the studies with a relatively small sample. This explanation cannot, however, account for the other major outlier, a study by Wallace, Kass, and Stanny (2002) which reportedly failed to show a significant correlation between SART performance and the CFQ.

Subsequently, however, Robertson et al. (1997) reported a modest but significant correlation between the CFQ and SART errors for both TBI patients (when CFQ ratings were provided by informants) and controls. A number of studies have subsequently attempted to replicate the original Robertson et al. finding. One recent study by Whyte et al. (2006) reported a failure to replicate the original Robertson finding and, based on a review of the limited literature available, raised questions about the validity of the SART.
In view of the concerns over the external validity of the SART expressed by Whyte et al., 2006, we believed it would be beneficial to conduct a formal meta-analysis of the available studies (Table 1) concerning the SART–CFQ association.

### Methods

The studies listed in Table 1 vary considerably in terms of the clinical status of populations sampled, participant ages, sex composition, education, ethnic composition, and variations of SART testing procedures. Such heterogeneity strongly suggested the use of random effects meta-analysis model (Hunter & Schmidt, 2004). We used the Hunter and Schmidt method as it appears to provide reasonably accurate estimates of effect sizes under conditions of heterogeneity (Field, 2001). The values of \( r \) in the present studies are small to moderate and therefore we used untransformed Pearson product correlation coefficients (or estimates based on statistics provided in the original studies). Z-transformations introduce their own biases (Hunter & Schmidt, 2004) and principally affect the distributional skew of \( r \) at higher values.

Several studies required special treatment. Wallace et al. (2002) reported only a nonsignificant F-ratio, \( F = .01 \). We converted this to a correlation coefficient of \( r = .01 \). Not knowing the direction of the effect we decided to err on the conservative side and...
Results and discussion

The heterogeneity of the methods and sampling of the studies is matched by a significant heterogeneity of results. Nonetheless, the studies generally provide evidence for a positive association between the CFQ and the number of SART commission errors on NOGO trials (Table 1). Our meta-analysis revealed a weighted mean $r$ of .21 ($Z = 2.24, p < .01$) with a 95% confidence interval ranging from .03 to .38. Interestingly, the 95% confidence interval of $r$ encompasses both the original Robertson et al. (1997) value of .27 and the Whyte et al. (2006) results for TBI patients. Based on these results, we conclude that there is a significant association between the CFQ and SART commission errors. Furthermore, given that the studies analyzed included a diverse set of populations and contexts, it follows that these results have considerable generalizability.

Consistent with statistical considerations, the strongest effects (mean $r = .40$) appeared in studies analyzing responses across extreme groups (e.g., groups selected based on CFQ scores: Manly et al., 1999; depressed and non-depressed soldiers: Farrin et al., 2003; burned out and non-burned out teachers: Van der Linden et al., 2005). Conversely, studies finding little or no relation arguably tested the most homogeneous populations (hospital staff, TBI patients only: Whyte et al., 2006; and undergraduates: Wallace et al., 2002). The original Robertson et al. (1997) study, however, appears to have employed a fairly homogeneous sample and obtained an intermediate association.

In order to further bolster the results of our meta-analysis, which was based on an unusually small number of independent studies, we sought to evaluate the SART–CFQ correlation using a large, diverse, sample from the general population. In addition, we were concerned about whether the CFQ was truly appropriate for evaluating the external validity of the SART as a measure of sustained attention failures, given that it is intended as a measure of general cognitive failure, not specific to attention. We therefore also evaluated the relation between the SART and more specific self-report measures of attention failures and attention-related errors.

Study 2: Is the SART a specific measure of sustained attention failure?

We previously developed scales specifically measuring Attention-Related Cognitive Errors (ARCES) and Memory Failures (MFS; Carriere, Cheyne, & Smilek, 2008; see also Cheyne et al., 2006). Both of these included items from the CFQ that were relevant to attention and memory, respectively, as well as new items. A problem encountered and reported by Broadbent et al. (1982) in their initial report was that the CFQ contains items referring to situations (e.g., driving and shopping) that some patients (and likely others; e.g., students) might not commonly experience. Hence, we eliminated any references to driving and shopping situations in the final version of the ARCES. In addition to the ARCES and MFS we also investigated a measure of attention lapses, the Mindful Attention Awareness Scale (MAAS; Brown & Ryan, 2003). To reduce overlap between the ARCES, MFS, and MAAS, we shortened the MAAS to include only items referring specifically to attention lapses (removing items 2 and 6; Cheyne et al., 2006). We hypothesized that the reduced MAAS – a direct measure of attention lapses – should be most closely associated with SART RT, as it is the putative index of mind wandering during SART performance (Robertson et al., 1997), whereas the ARCES – a measure of the behavioral consequences of attention lapses – would be most closely associated with SART errors.

In a relatively large and diverse web-based international sample ($n = 504$), we found all three self-report measures (MAAS, ARCES, and MFS) were correlated with SART error and SART RT as well as with one another (Cheyne et al., 2006). The ARCES–SART error correlation was found to be .32, very close to the mean found for the CFQ–SART error correlation in the present meta-analysis. In addition, and consistent with theory, detailed analysis revealed that the MAAS accounted for the ARCES–SART GO RT correlation, the ARCES accounted for the MAAS–SART error correlation, and the attention measures (MAAS and ARCES) jointly accounted for the correlations of the memory measure (MFS) with both SART GO RTs and SART errors. These results suggest that SART errors may indeed provide a valid measure of specifically attention-related cognitive errors, a conclusion that could not be firmly made based on the CFQ alone. However, given that the relation between SART errors and the ARCES has been demonstrated in only one study to date, it is critical to replicate this relation in another sample.

In the present study we sought to replicate our previous results with the SART, ARCES and MAAS as well as to examine the
associations between the CFQ and these more specific self-report measures of attention failures and attention-related cognitive errors. Subsequent to our earlier report using the MAAS we also removed item 12, as it references lapses when driving, and relabeled the scale as the MAAS-LO, i.e., MAAS-Lapses Only (Carriere et al., 2008). This reduced the 15-item MAAS to 12 items in the MAAS-LO, and made the scale more consistent with the goals set out in our development of the ARCES. Thus, we hypothesized that we would again find stronger relations between the MAAS-LO and SART GO RT (relative to the ARCES) as well as between the ARCES and SART errors (relative to the MAAS-LO). In addition, we sought to examine whether the CFQ would show specificity toward SART errors similar to the ARCES. Finally, this study provided another opportunity for us to evaluate the association between the SART and the CFQ in a large, moderately heterogeneous sample of individuals.

Method

Participants

Participants were randomly selected from a diverse international group of prior respondents to a WWW survey on sleep paralysis. Of 3000 potential participants contacted for the present study, the final sample included 363 participants who voluntarily completed all the necessary questionnaires and the SART, without leaving more than a single response blank for any given questionnaire. This sample included 261 females and 102 males with a mean age of 30.3 (SD = 8.6; females M = 30.6, males M = 29.6).

Measures

The measures included the 12-item ARCES (Carriere et al., 2008), the 12-item MAAS-LO (see Carriere et al., 2008), the 25-item CFQ (Broadbent et al., 1982) and the SART (Robertson et al., 1997). In addition, though not analyzed for the purposes of the present study, participants also completed the Epworth Sleepiness Scale (Johns, 1991) and the short form of the Depression Anxiety Stress Scales (Lovibond & Lovibond, 1995). Within each questionnaire the individual items were presented in a random order, such that no two participants were likely to receive the exact same configuration of items over the course of the study.

The SART employed in the present study is the same as that used in our previous study (Cheyne et al., 2006, Cheyne et al., 2009a and Cheyne et al., 2009b) with two notable exceptions. First, the mask presented following each digit was changed to a double ringed bull’s-eye shape to avoid disproportionate masking of the digit 8 at larger font sizes, which bears a substantial resemblance to the typical SART mask. The outer ring was sized such that it did not overlap with any digits, even at the largest font size, while the inner ring was sized such that it had minimal overlap with digits in any of the four standard font sizes. Second, the intervening number of GO trial digits (digits 1, 2, 4–9) appearing between NOGO trials (the digit 3) was varied from 0 (i.e., sequential NOGO trials) to 16, with each interval being used exactly twice over the course of the task. This range represents the full complement of potential NOGO-to-NOGO intervals for the standard SART (where randomized blocks of nine digits are used, with each digit appearing once per block). This second change also necessitated an increase in the number of SART trials from 225 to 315. All participants received the exact same order of digit presentation when completing the SART, and interval lengths were distributed well over the course of the task.

Procedure

Participants received an informational email inviting them to participate in the study, including a link to the study website. After visiting this website, and upon consenting to participate in the study, participants completed: (1) a short demographic form; (2) each of the above questionnaires, presented in random order; and (3) the SART. At the end of the study participants received a feedback page thanking them for their participation and providing additional information on our research.

Results and discussion

To accommodate the potential for blank responses, participant scores for all questionnaires are based on the mean value of all responses provided by the participant. As well, given the larger number of trials employed in the present SART, the SART error rate is calculated as the proportion of NOGO trials on which a response was made for comparability to the values in Table 1.

Means and SDs are provided for the CFQ, ARCES, MAAS-LO, SART errors, and SART GO RTs in Table 2. There were no significant sex differences for any measures. Pearson product–moment correlations among the cognitive and attentional measures are provided in Table 3. Not surprisingly, given that they share items, the CFQ and ARCES are highly correlated and both are robustly correlated with MAAS-LO scores. All three are moderately correlated with SART errors, with coefficients very similar to the mean r found in the meta-analysis for the CFQ and our previous research with the ARCES and reduced MAAS (Cheyne et al., 2006). Both the CFQ and MAAS-LO, but not the ARCES, were significantly associated with SART GO RT. There were no sex differences for any of the correlations.
There are several specific conclusions to be drawn from the correlation table shown in Table 3. First, consistent with the results of our meta-analysis, we found a significant correlation ($r = .28$, $p < .01$) between SART errors and the CFQ. This correlation is similar to the mean correlation revealed by the meta-analysis and falls squarely within the confidence interval found in the meta-analysis. Thus the results of the present study agree with the results of our meta-analysis. Second, our finding of an association between SART errors and the ARCES replicates our previous work and supports the conclusion that SART errors are in fact a valid measure of sustained attention-related cognitive errors, a conclusion which could not be firmly made on the basis of CFQ total scores alone. Third, the general pattern of correlations is consistent with the CFQ being a more global measure of cognitive failure and the ARCES being a specific measure of attention-related errors, since the CFQ correlates with both SART errors and SART RTs while the ARCES correlates only with SART errors.

The specificity of the ARCES, MAAS-LO, SART errors and SART RTs receives further corroboration from structural equation modeling (SEM) results. In previous work we reported SEM analysis that produced a well-fitting model in which the reduced MAAS predicted SART GO RT independently of ARCES, whereas the ARCES predicted SART errors independently of the MAAS and both mediated the association between the MFS and both SART GO RT and SART errors (Cheyne et al., 2006). Because time constraints prevented the use of the MFS in this study, we created an equivalent memory measure from overlapping items on the CFQ (items 7, 11, 16, 17, 20, 22, and 23). The resulting CFQ-memory scale was significantly correlated with the ARCES, MAAS-LO, SART errors and SART GO RT at $r = .68$, .58, .20, and -.13, respectively, all at $p < .05$.

As in Cheyne et al. (2006), causal paths were constructed from the MAAS-LO to ARCES and from SART GO RT to SART errors to reflect the hypothesized causal role of attention failures on the attention-related cognitive errors. Causal paths were also constructed from the MAAS-LO to SART GO RT and from the ARCES to SART errors consistent with the hypothesized causal role of dispositional attentional factors on behavioral performance on the SART. No paths were provided from the MAAS-LO to SART errors or from the ARCES to SART GO RTs, nor were paths provided from CFQ-memory to either SART GO RT or SART errors as these associations are hypothesized to be explained by the previous causal paths. Significant path coefficients were found, as predicted, for paths between ARCES and SART errors and between the MAAS-LO and SART GO RT (Fig. 1). This theoretically constrained model, eliminating the ARCES–SART GO RT and the MAAS-LO–SART error paths and between
the CFQ-memory and both SART measures, provided very good fit indices, $\chi^2(4) = 2.20, p = .693$, CFI = 1.00, NFI = .997, RMSEA = .00, consistent with previously reported results (Cheyne et al., 2006). For the saturated model, the path coefficients from ARCES to SART GO RT and from the MAAS-LO to SART error were not significant, as predicted. Neither path coefficient from the CFQ-memory to SART measures was significant. We also note that inspection of Fig. 1 reveals stronger path coefficients between the three subjective report measures than between these measures and the SART measures. This result is, however, likely a consequence of the fact that the subjective report measures uniquely share method variance and hence these differences are not theoretically interesting. Indeed, the same observation applies to the relation of SART RT to SART errors which also share method variance. Thus, the weaker coefficients obtained across different measurement methods do not reflect on the validation of the SART as a specific index of attention failures.

![SEM path model with significant paths for self-report measures of attention a lapses (MAAS-LO), Attention-Related Cognitive Errors (ARCES), SART measures: SART GO RTs, SART errors and for the mediation of CFQ-memory and SART measures by the MAAS-LO and ARCES.](image)

To assess the consistency of the results we tested the model further for separate sub-samples, split by sex. First, we tested the model in Fig. 1 for two groups divided by sex for which the paths were unconstrained and free to vary between the two groups. This was a well-fitting model with acceptable goodness-of-fit fit indices: $\chi^2(8) = 5.04, p = .753$, CFI = 1.00, NFI = .994, RMSEA = .00. Next we tested the same model but with paths constrained to be equal for the two groups. That is, the constrained model is assumed to fit both groups equally well. This too was a well-fitting model with acceptable fit indices: $\chi^2(14) = 15.41, p = .351$, CFI = 1.00, NFI = .981, RMSEA = .017. As the models are nested, it is possible to directly compare the models, to assess whether the additional constraints significantly reduced the model fit. The result indicated that the constrained model was not
Failures of sustained attention in life, lab, and brain: Ecological validity of... 

The structural equation model shown in Fig. 1 supports two main conclusions. First, the model results show specificity for both SART measures (SART RT and SART errors) and the subjective report measures (MAAS-LO and ARCES). Specifically, the results support the assumption of the model that increased propensity for experiencing attention lapses (measured by the MAAS-LO) leads to faster SART RTs but does not directly lead to an increase in SART errors. On the other hand an increased propensity for making attention-related errors (measured by the ARCES) leads to increased SART errors but not faster SART RTs. These results highlight the validity and specificity of the SART and also the specificity and utility of the ARCES and the MAAS-LO. Second, the results for the CFQ-memory subscale indicate no need for causal paths between a subset of the CFQ items and either of the SART measures.

**General Discussion**

A review and meta-analysis of prior research investigating the association between the CFQ and SART error scores corroborates the original claim by Robertson et al. (1997). The effect size is, not surprisingly, small and variable across populations and contexts when samples are small and/or homogeneous. Nonetheless, the association appears to hold for diverse populations. Indeed, contrary to criticisms recently raised by Whyte et al. (2006), the CFQ and SART relation seems to hold even for individuals with TBI. We note that Whyte et al.’s (2006) nonsignificant correlation of .11 with 25 participants in the sample is not statistically different from Robertson et al.’s (1997) reported correlation of .44 with 22 participants, $Z = 1.14, p = .13$ (one-tailed). In addition, since the two correlations are not statistically different from each other, we can combine the two correlations by computing the weighted mean correlation. The weighted mean correlation is statically significant, $r = .26$ ($N = 46, p < .04$, one-tailed).

During our review of the literature we noticed that several of the studies reviewed also report results that provide mutual support for the claims of both the CFQ and SART. The Van der Linden et al. (2005) study of teacher burnout is particularly interesting in that it found self-reports of cognitive complaints during the SART task to be significantly related both to burnout status and to SART errors. Thus, people do seem to be sensitive to, and able to report reliably about, problems of sustained attention. In addition, the results of Van der Linden and colleagues are quite consistent with the presenting problems of burnout (e.g., inability to concentrate on reading a newspaper, to keep one’s mind on a complex problem, or to focus during a conversation). These findings are particularly interesting in light of studies showing that attentional complaints (ARCES) are predictors of depression (e.g., Carriere et al., 2008) and that there are SART differences between depressed and non-depressed soldiers (Farrin et al., 2003). Deficits in the ability to sustain attention may therefore be interpreted by those experiencing such deficits as a lack of interest and inability to find meaning in previously engaging tasks, and hence contribute to general dysphoria.

Our review of the literature also, however, revealed several problems of data reporting and interpretation. Conclusions have sometimes been based on inadequate sample sizes and incomplete analyses. For example, Whyte et al.’s (2006) concerns regarding the validity and/or reliability of the SART based on their single small $n$ study seem not to be borne out by the results of the present study or the meta-analysis of previous research. In another example, in their attempts to interpret the stronger correlation between two questionnaires (CFQ and BDI) than between each of these and a behavioral task, Farrin et al. (2003) did not discuss the implications of the effect of shared method variance on the correlations. We also found that researchers often fail to report the actual value of “nonsignificant” parameters. We found a number of correlation tables filled with blanks, en-dashes, or “ns” (see, for example, Table 3 of Robertson et al., 1997). Failure to report effect sizes of whatever size seriously hampers the interpretive efforts of readers and reviewers and seriously compromises quantitative meta-analyses. This problem is particularly serious in small $n$ studies in which effect sizes can be quite substantial and yet not achieve the holy grail of $p < .05$. Such values can still provide important evidence when combined with other data and given appropriate weighting. These are elementary statistical considerations that are all too often ignored in research reports. One small $n$ study can easily “fail to replicate” a previous large $n$ study (or a previous small $n$ study for that matter) simply through the rather ignominious achievement of lack of sufficient statistical power. Small $n$ studies are of course inevitable in many areas and, as our previous remarks should suggest, our purpose is not to discourage such studies or disparage their potential value. Rather, because it is often not feasible to achieve large sample sizes with clinical samples or when conducting neurological assessments, these and ultimately all individual studies must be evaluated in the context of multiple studies including large sample validation studies.

The present results also make a case for the advantage in precision of conclusions achieved by using the ARCES instead of the CFQ when evaluating attention-related cognitive failures. Although we found the CFQ to be reliably associated with SART performance, its lack of specificity limits the conclusions that can reasonably be drawn for more targeted, theoretically oriented research. Indeed, the development of the SART was motivated by the goal of creating a behavioral task that specifically measures attention-related errors as opposed to general cognitive errors. To validate such a targeted behavioral measure it is important to...
use a self-report measure that is of roughly comparable specificity. We provide evidence that the ARCES is a specific and conceptually meaningful measure of attention-related errors distinct from memory related errors (MFQ) and attention lapses (MAAS-LO), and is thus a suitable replacement for the CFQ in studies seeking to measure everyday attention-related cognitive failures.

Finally, we highlight the implications of the present study for interpreting the brain-behavior relations associated with failures of sustained attention. The SART is increasingly being employed in studies assessing the brain areas associated with attention failure. These studies have used a wide range of brain imaging techniques such as EEG-ERP (O’Connell et al., 2008 and Smallwood et al., 2008), fMRI (Christoff et al., 2009), and MEG (Cheyne, Cheyne, et al., 2009). The studies have revealed several areas associated with attention failures, such as the dACC and areas of the PFC that have been linked to the default network. ERP studies are also being employed to evaluate whether SART errors result from inhibition failures or from inattention (O’Connell et al., 2007). The implicit assumption of all of these studies is that the brain areas active prior to an attentional failure in the SART (i.e., a SART error) also reflect brain activity during attention failures in everyday life for both normal and clinical populations. Indeed, the primary goal of the SART was to provide an ecologically valid measure of attention failures that can be used to study normal individuals as well as those with clinical problems such as traumatic brain injury (e.g., Robertson et al., 1997) and attention deficit disorder (e.g., Manly et al., 2001). In the present study, we demonstrate that, contrary to recent criticisms (see Whyte et al., 2006), SART errors are indeed associated with reports of attention failure in everyday life. Such validation provides support for the assumption that brain areas uniquely associated with SART performance also participate in attention failures in everyday life. Given the increasing use of several variants of the SART and related tasks to infer brain states during sustained attention and its failures (e.g., Bellgrove et al., 2004, Cheyne et al., 2009a, Cheyne et al., 2009b, Dockree et al., 2007, Dockree et al., 2005, Fassbender et al., 2006, Hester et al., 2004, Hester et al., 2005, Manly et al., 2001, O’Connell et al., 2007, O’Connell et al., 2008, Robertson et al., 1997 and Zordan et al., 2008) evidence of such ecological validation must be of central concern.

Footnotes

1 Robertson and colleagues employed informants for the patient CFQs because they were obviously concerned that TBI patients might lack insight into the extent of their deficits. It is also important to note that the sample size for patients was much smaller than that for controls. Hence, the significant correlation for patients was, in fact, much larger than that for controls (.44 versus .27). It is possible that effect size even for self-report data from TBI patients was numerically larger than that for controls but not significant given the considerably reduced power for that test in that group. Unfortunately, Robertson and colleagues fail to provide the value of the self-report based correlation and hence it was not possible to include the results based on self-report in our meta-analysis below.

Acknowledgements

This work was supported by a research grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada awarded to DS and a graduate scholarship from NSERC awarded to JSAC. All authors contributed equally to this work.

References


