

Simon says: Reliability and the role of working memory and attentional control in the Simon task

KARL W. U. BORGMANN, EVAN F. RISKO, JENNIFER A. STOLZ, AND DEREK BESNER
University of Waterloo, Waterloo, Ontario, Canada

The *Simon effect* refers to the observation that subjects identify targets (e.g., colors) faster when the irrelevant spatial location of the target corresponds to the location of the response key. Theoretical accounts of the Simon effect typically explain performance in terms of automatic and controlled processes. Furthermore, the relative contributions of automatic and controlled processes are held to change as a function of the proportion of compatible to incompatible trials (*compatibility proportion*). Data are presented demonstrating that the reliability of the Simon effect, indexed by correlating its magnitude within subjects across blocks of trials, varied substantially as a function of the compatibility proportion. When the compatibility proportion was high, so was reliability. When the compatibility proportion was low, reliability was low as well. The results are discussed in terms of the relative reliability of automatic and controlled processes and the role of working memory and attentional control in goal maintenance.

Reliability is a fundamental psychometric property that needs to be determined in the measurement of any theoretically important empirical construct. Furthermore, its assessment is critical to research in many fields of psychology—for example, to studies of semantic processes in visual word recognition (Stolz, Besner, & Carr, 2005), of aging and memory (Buchner & Wippich, 2000; Salthouse & Siedlecki, 2005; Salthouse, Toth, Hancock, & Woodard, 1997), and of aging and cognitive performance (Madden, Pierce, & Allen, 1993). Determining the reliability of a cognitive process over time is thus part of a complete understanding of how that process functions. Curiously, however, the issue of reliability is typically neglected. This neglect likely stems from numerous sources, but an insidious implicit assumption may well be that cognitive processes are reliable because they produce robust and replicable empirical phenomena (e.g., the Stroop effect, semantic priming). This assumption may also, in large part, reflect the widespread belief that many cognitive processes have automatic components that are expected to unfold in a consistent manner. We consider this point below.

Theoretical Utility

Many cognitive paradigms are explained by viewing their underlying components as reflecting some combination of (1) automatic processes and (2) controlled processes. *Automatic* is often taken to mean that processing begins regardless of intent, requires no resources, and is ballistic (i.e., once started, it continues until it is finished; see, e.g., Posner & Snyder, 1975). *Controlled*, in contrast, is often taken to mean that processing is dependent on

both intention and attentional resources. Given these conceptions of automatic versus controlled processing, the former is assumed to unfold in a more consistent manner (i.e., to have lower variability) than the latter. Thus, reliability can be taken as an additional feature that distinguishes an automatic from a controlled process, in which the former yields more reliable performance than the latter. For example, in assessing the reliability of memory measures, Buchner and Wippich (2000) argued that “a memory measure should be reliable to the degree to which the instructions and other characteristics of the task serve to limit the variability in the types of processes underlying task performance” (p. 248). The relative automaticity of the component cognitive processes would be one characteristic that should limit variability in task performance.

Recent work by Stolz et al. (2005) has highlighted the theoretical utility of assessing reliability with respect to the relative automaticity of semantic activation in the context of visual word recognition. It is widely assumed that semantic activation (and its subsequent spread) underlying semantic priming occurs automatically (for reviews, see Neely, 1991; Neely & Kahan, 2001). Stolz et al. investigated the implicit assumption that automatic activation should be reliable by assessing the reliability of semantic priming in response times (RTs) in a series of lexical decision studies that crossed three relatedness proportions (i.e., the proportions of trials in which the prime and the word target were semantically related) and three SOAs. Stolz et al. measured priming (i.e., unrelated RT – related RT) for each subject in Blocks 1 and 2 of their experiments and correlated these values (i.e., unrelated RT – related RT for Block 1 was correlated with

K. W. U. Borgmann, kwuborgm@uwaterloo.ca

unrelated RT – related RT for Block 2). The reliability of the semantic priming effect across the different combinations of relatedness proportions and SOAs varied from essentially 0 to .43. Critically, in the three conditions with a relatedness proportion of .25, which according to Neely (Neely, 1991; Neely & Kahan, 2001) is the proportion most likely to index automatic semantic activation, no reliability was observed. Thus, despite the seemingly widespread (if implicit) assumption that the activation responsible for producing semantic priming should be reliable, it turns out that reliability is strongly influenced by context.

Stolz et al.'s (2005) findings highlight the usefulness of evaluating empirically how consistently cognitive processes unfold as a function of various contexts. The Stolz et al. design (comparing reliabilities across different contexts) has the added benefit of addressing some of the issues related to computing the reliabilities of difference scores. For instance, it cannot be argued that, all else being equal, the finding of low reliability in one condition is simply due to measurement problems if reliability is found in another condition (see also Madden et al., 1993). Because the issue of interpreting null reliability of difference scores is central to the present article, we consider it in more detail below.

A Central Methodological Issue: The Use of Difference Scores

To study the reliability of a cognitive process, one can take an index of that process and correlate performance at two different points in time, as was done by Stolz et al. (2005). This strategy, however, often depends on correlating difference scores, which is considered to be problematic (Williams & Zimmerman, 1996), since the absence of reliability may indicate either low reliability or an inability to detect reliability that is present. Stolz et al. addressed this issue by correlating a second set of difference scores calculated from the same data as a check of their ability to find significant correlations between difference scores, should they be present. Specifically, they correlated changes in performance for each type of trial (i.e., related vs. unrelated) across Blocks 1 and 2 [i.e., related (Block 1) – related (Block 2) was correlated with unrelated (Block 1) – unrelated (Block 2)]. Because they found robust and significant correlations using these difference scores, which were based on the same number of observations per condition and came from the same conditions as the theoretically interesting priming scores, the method of correlating the reliabilities of difference scores per se cannot be the reason for nonsignificant reliability.

Compatibility Proportions and Reliability in Simon (1990)

Another task widely believed to reflect some combination of automatic and controlled processes is the Simon task (Kornblum, Hasbroucq, & Osman, 1990; Lu & Proctor, 1995; Simon, 1990). When a to-be-identified stimulus (e.g., a color patch) is presented to the left or right of a central fixation, subjects are faster and more accurate

when the stimulus location, *despite being irrelevant to the task*, is on the same side as the correct response key. This is commonly referred to as the *Simon effect*, and it is a robust and replicable phenomenon (see Simon's, 1990, review). In addition, the Simon task is an excellent starting point for an investigation of reliability related to automatic and controlled processes, for two reasons: (1) It reflects very basic cognitive processes, and (2) there are two existing accounts that can provide a framework for interpreting the results (cf. Kane & Engle, 2003; Toth et al., 1995).

The Simon effect is argued to arise because two separate internal responses are generated: an automatic spatial correspondence response, and a controlled response based on task instructions (see, e.g., Kornblum et al., 1990; Lu & Proctor, 1995; Simon, 1990). When the automatic and controlled processes produce the same response (i.e., the correct response key corresponds spatially to the stimulus location, a *compatible* trial), subjects respond more quickly than when the automatic and controlled processes produce different responses (i.e., the correct response key does not correspond spatially to the stimulus location, an *incompatible* trial).

The proportion of compatible trials influences both the magnitude and *direction* of the Simon effect. Previous work has established that when the compatibility proportion is high, the Simon effect is large, and when the compatibility proportion is low, the Simon effect reverses direction (Toth et al., 1995).

The manner in which the nominally irrelevant spatial dimension influences performance has been suggested to take the form of an additional "automatic" process (Toth et al., 1995). Using the process dissociation procedure (Jacoby, 1991), Toth et al. argued that the Simon effect is a product of three component processes: (1) an automatic spatial correspondence process; (2) a controlled response process; and (3) an *associative* automatic route, which is sensitive to the local experimental association between the location of the stimulus and the correct response. The latter component is argued to be responsible for the compatibility proportion effects on the magnitude and direction of the Simon effect. According to this account, the contributions to performance of the spatial correspondence and controlled response processes are stable and consistent across different compatibility proportions (see also Kornblum et al., 1990; Lu & Proctor, 1995; Simon, 1990). In contrast, the associative automatic process only contributes to the Simon effect when the compatibility proportion is *not* .50. Furthermore, the direction of that contribution changes as a function of the compatibility proportion. Thus, when the compatibility proportion is low (.25), the influence of the associative automatic process is in the direction opposite that of the automatic spatial correspondence process. Conversely, when the compatibility proportion is high (.75), the contribution of the associative automatic process is in the same direction as the automatic spatial correspondence process.

The effects of compatibility proportion on the magnitude and direction of the Simon effect can also be couched in terms of a modified version of Kane and Engle's (2003)

discussion of the role of working memory and attentional control in producing the Stroop effect. According to Kane and Engle, limitations in working memory and attentional control processes are responsible for Stroop interference, and differences in the magnitude of the Stroop effect at different congruency proportions are a function of the degree to which working memory and attentional control are required to maintain the task goal (i.e., respond to the color, not the word). If we extend this analysis to the Simon task, then subjects require working memory and attentional control in order to maintain the goal of responding on the basis of the character rather than its location.

In Kane and Engle's (2003) account, the congruency proportion affects the ease with which subjects can maintain the task goal. When the proportion of congruent trials is high, with the color and word providing the same response information, it is difficult to maintain the goal of responding to the color. Thus, incongruent word information, although not appearing frequently, will produce robust Stroop interference, which arises primarily from a failure to maintain the appropriate task goal. In contrast, when the congruency proportion is low, the large number of incongruent trials serves as a reminder of the task goal, and the amount of interference is relatively small. Under these low-congruency-proportion conditions, interference is hypothesized to arise primarily from conflict resolution mechanisms. Extended to the present work, the proportion of compatible trials (i.e., the compatibility proportion) is hypothesized to influence performance in the Simon paradigm in a manner parallel to that of the role of the proportion of congruent trials in the Stroop paradigm.

The Present Experiment

Despite the rather large literature, we are unaware of any investigations of the reliability of the Simon effect within subjects across blocks of trials (but see Salthouse et al., 1997, for a different kind of reliability assessment). The present experiment examines the reliability of the Simon effect across blocks of trials as a function of compatibility proportion (i.e., .25 vs. .50 vs. .75). Both the Toth et al. (1995) and Kane and Engle (2003) accounts outlined above can be used to generate predictions concerning this investigation.

Given the Toth et al. (1995) expectation of a positive relation between the relative automaticity of the three component processes and reliability, the following predictions can be made. Since they argue that the associative automatic process contributes only in the .25- and .75-compatible conditions, we should find reliability that reflects this contribution. Specifically, if the associative automatic route is reliable, the .25- and .75-compatible conditions should yield *greater* reliability than the .50 condition, because the .25 and .75 conditions are influenced to a greater extent by the automatic processes. In contrast, if the associative automatic route is unreliable, the .25- and .75-compatible conditions should yield *less* reliability than the .50-compatible condition, because the .25 and .75 conditions are the only ones influenced by this process.

Relevant predictions can also be derived from the Kane and Engle (2003) account. If the goal maintenance processes are reliable, the greatest reliability should be seen when the compatibility proportion is high, and the need to draw upon working memory and attentional control for goal maintenance dominates. In contrast, if the goal maintenance processes are unreliable, reliability should decrease as the need for these processes increases (i.e., as the compatibility proportion increases).

METHOD

Subjects

One hundred undergraduates from the University of Waterloo were paid \$2 each. All reported normal or corrected-to-normal vision.

Design

The experiment consisted of a 2 (compatibility: compatible vs. incompatible) \times 2 (block: 1 vs. 2) \times 3 (compatibility proportion: .25 vs. .50 vs. .75) mixed design. The levels of compatibility and block were manipulated within subjects, whereas compatibility proportion was manipulated between subjects. Trials were considered *compatible* when the target appeared on the same side as the required response and *incompatible* when the target appeared on the side opposite to the required response.

Procedure

The stimuli were presented on a 17-in. color monitor driven by a Pentium computer running E-Prime software (Psychology Software Tools, 2002). The target consisted of an "X" or an "O," measuring 8 mm vertically \times 8 mm horizontally, displayed 38 mm to the left or right of a central fixation point (+). Each target appeared an equal number of times within the experiment. Responses were collected using a standard QWERTY keyboard. Subjects used the *A* and *L* keys to make their responses. Mapping of stimuli to responses was counterbalanced across subjects. The midpoint between the two response keys was aligned with the central fixation.

Subjects were tested individually in a sound-attenuated room. Instructions for the task were displayed visually and relayed verbally by the experimenter. The subjects were told to identify the stimulus by pressing the appropriate key, to ignore the location of the stimulus, and to maintain their focus on the central fixation. They were also instructed to respond as quickly and accurately as possible.

Table 1
Mean Response Times (RTs, in Milliseconds)
and Percent Errors (%E) As a Function
of Compatibility and Compatibility Proportion

Compatibility Proportion	Block 1		Block 2	
	RT	%E	RT	%E
.25-Compatible Condition, <i>n</i> = 60				
Incompatible	425	5.6	421	6.3
Compatible	452	9.3	452	10.7
Simon effect	-27	-3.7	-31	-4.4
.50-Compatible Condition, <i>n</i> = 20				
Incompatible	463	3.2	461	4.9
Compatible	442	2.4	444	3.1
Simon effect	21	0.8	17	1.8
.75-Compatible Condition, <i>n</i> = 20				
Incompatible	465	9.6	465	11.1
Compatible	393	1.1	385	1.3
Simon effect	72	8.5	80	9.8

Each trial began with the presentation of a fixation cross for 500 msec, followed by a target. The target remained on the screen until a response was made. This was followed by an intertrial interval of 500 msec. Eighty practice trials were followed by two blocks of 120 experimental trials. Trial order was determined randomly for each subject.

Compatibility Proportion Manipulation

Sixty subjects were assigned to the .25 compatibility proportion condition, and 20 subjects each to the .50 and .75 compatibility proportion conditions.

RESULTS

An RT analysis was conducted for trials in which the response was accurate. RTs greater than 1,000 msec (0.8% in the .25 condition, 2.4% in the .50 condition, and 1.0% in the .75 condition) were removed prior to analysis (following Hommel, 1994a, 1994b) because irrelevant spatial information is not expected to influence performance at long RTs (see Simon, 1990). Mean RTs and percentage errors are presented in Table 1. Descriptive data are presented in Table 2. Parallel 2 (compatibility: compatible vs. incompatible) \times 2 (block: 1 vs. 2) \times 3 (compatibility proportion: .25 vs. .50 vs. .75) mixed-design ANOVAs were conducted on the mean RT and percentage error data.

RTs

There was no main effect of block ($F < 1$) and no interaction between the effect of block and either compatibility or compatibility proportion ($F_s < 1$). The main effect of compatibility was significant [$F(1,97) = 70.3$, $MS_e = 519.64$, $p < .001$], as was the compatibility \times compatibility proportion interaction [$F(2,97) = 166.27$, $MS_e = 519.64$, $p < .001$]. Subsequent analyses indicated that responses to compatible trials were 29 msec slower than those to incompatible trials in the .25 condition, 19 msec faster in the .50 condition, and 76 msec faster in the .75 condition (all $F_s > 10$, all $p_s < .01$). The block \times compatibility \times compatibility proportion interaction was not significant ($F < 1.1$).

Table 2
Minimum and Maximum Response Times (in Milliseconds)
As a Function of Condition

	Block 1			Block 2		
	Min	Max	Range	Min	Max	Range
.25-Compatible Condition						
Incompatible	338	597	259	328	574	246
Compatible	344	652	308	347	594	247
Simon effect	-98	79	177	-108	31	139
.50-Compatible Condition						
Incompatible	378	582	204	378	569	191
Compatible	376	541	165	363	538	175
Simon effect	-22	69	91	-23	57	80
.75-Compatible Condition						
Incompatible	411	578	167	379	553	174
Compatible	352	451	99	307	453	146
Simon effect	25	152	127	20	150	130

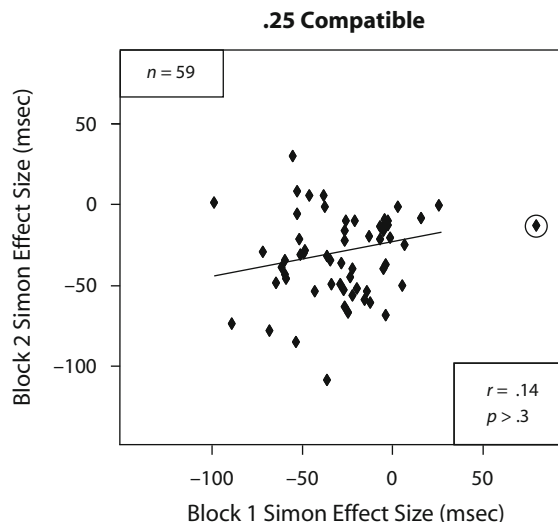


Figure 1. Correlation between the magnitudes of the Simon effect in Blocks 1 and 2 for the .25 compatibility proportion condition. The value identified as an outlier is circled.

Percent Errors

There was a main effect of block [$F(1,97) = 6.1$, $MS_e = 14.39$, $p < .05$], but block did not interact with either compatibility or compatibility proportion ($F_s < 1$). The main effect of compatibility was significant [$F(1,97) = 9.4$, $MS_e = 22.74$, $p < .005$], as was the compatibility \times compatibility proportion interaction [$F(2,97) = 74.7$, $MS_e = 22.74$, $p < .001$]. Consistent with the RT results, subjects made 4.2% more errors on compatible trials than on incompatible trials in the .25 condition, 1.3% fewer errors on compatible trials in the .50 condition, and 9.2% fewer errors on compatible trials in the .75 condition. The block \times compatibility \times compatibility proportion interaction was not significant ($F < 1$).

Reliability of the Simon Effect

The Simon effect (incompatible RT – compatible RT) was calculated for each subject for Block 1 and for Block 2, and these scores were correlated. The outlier procedure based on studentized residual values used by Stolz et al. (2005) resulted in one data point being excluded from the .25 condition (this outlier is circled in Figure 1).

The correlation between the size of the Simon effect in Blocks 1 and 2 was computed for each of the compatibility proportions using Pearson's r ; this yielded a non-significant correlation in the .25 condition ($r = .14$, $p > .3$; see Figure 1) and robust correlations in the .50 and .75 conditions ($r = .56$, $p < .01$, and $r = .65$, $p < .005$, respectively; see Figures 2 and 3).

In order to ensure that these reliabilities were not an artifact of overall slowing as compatibility proportion increased, a z score was calculated for each mean for each subject (Faust, Balota, Spieler, & Ferraro, 1999). The differences of the z scores were then calculated (incompatible – compatible) for each block, and these scores were correlated. This analysis replicated the results observed

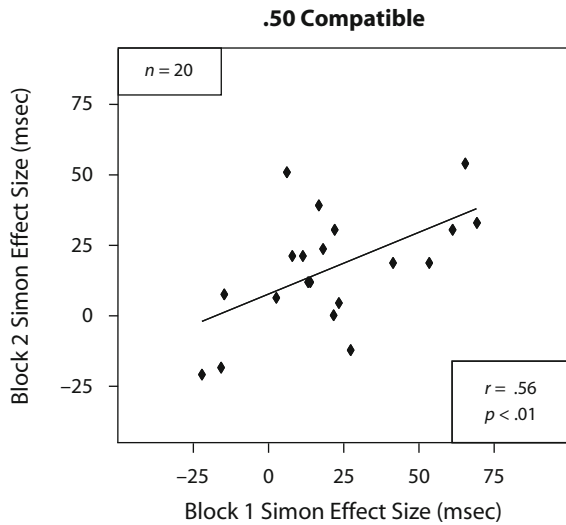


Figure 2. Correlation between the magnitudes of the Simon effect in Blocks 1 and 2 for the .50 compatibility proportion condition.

with the RT data, with a marginal correlation observed in the .25 condition ($r = .25, p = .06$) and robust correlations in the .50 and .75 conditions ($r = .47, p < .05$, and $r = .59, p < .01$, respectively).

Comparing reliabilities across compatibility proportions. The reliability of the Simon effect across blocks was significantly smaller in the .25 condition ($r = .14$) than in either the .50 condition ($r = .56; z = 1.77, SE = 0.277, p < .05$, one-tailed) or the .75 condition ($r = .65; z = 2.22, SE = .277, p < .05$). There was no difference between the .50 and .75 conditions ($z = 0.36, p = .36$).

Summary

This experiment investigated the effect of compatibility proportion on the reliability of the Simon effect. When the compatibility proportion was .25, a *reverse* Simon effect was observed, which was not reliable across blocks when absolute RTs were used ($r = .14$). Marginal reliability was observed in this condition in the *z* score analysis ($r = .25$). In contrast, the .50 condition yielded both a Simon effect and clear evidence of reliability in the direction and magnitude of the effect produced by subjects across blocks ($r = .56$). Likewise, the same held true for the .75 condition, with both a large Simon effect and clear evidence of reliability ($r = .65$).

On Correlating Difference Scores

As noted in the introduction, interpretation of a null correlation of difference scores can be problematic. Specifically, researchers may be concerned that the absence of reliability in the absolute RTs (in the .25 condition) reflects a failure to detect reliability that actually exists. This concern can be addressed in the same manner as it was in Stolz et al. (2005). The equivalent step here was to take a subject's mean incompatible RT in Block 2 and subtract it from his or her mean incompatible RT in Block 1, generating a difference score for the incompatible trials.

The same was done for compatible trials. A correlation was then calculated between these two difference scores. The .25 condition yielded a block difference score correlation of .56 ($p < .001$), the .50 condition a correlation of .52 ($p < .05$), and the .75 condition a correlation of .57 ($p < .01$). Thus, there is no reason to suppose that the nonsignificant reliability observed in RTs for the .25 condition (marginal in the *z* score analysis) was due to the use of difference scores per se.

It should also be noted that it is difficult to account for the absent/marginal reliability in the .25 condition simply on the basis of an inherent difficulty in correlating difference scores, given that significant reliability was observed in the .50 and .75 conditions, using the same procedure with far fewer subjects ($N = 20$ [for .50 and .75] vs. $n = 60$ [for .25]). Furthermore, the magnitude of the Simon effect observed in the .25 condition had the largest range of any of the conditions, thus ruling out the argument that a restriction-of-range problem occurred in the .25 condition (see Figures 1, 2, and 3, as well as Table 2).

Finally, high intercorrelations were observed between component scores across blocks for all compatibility proportions (e.g., compatible RT in Block 1 correlated with compatible RT in Block 2; see Table 3), suggesting that the component scores were not limiting factors in obtaining reliable difference scores. Interpretation of component scores beyond this observation would be of limited utility, however, since they were calculated from raw RT data, and therefore share many common sources of variance.

DISCUSSION

The reliability of the Simon effect was assessed in three groups of subjects using three proportions of compatible trials. We replicated the RT and error patterns found previously with manipulations of compatibility proportion, and also discovered that the reliability of the Simon effect

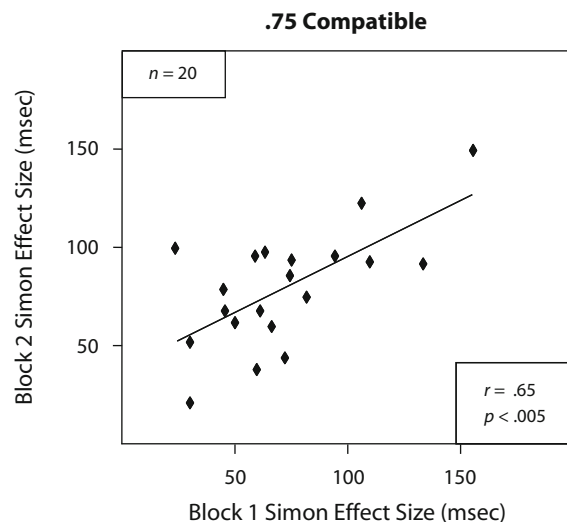


Figure 3. Correlation between the magnitudes of the Simon effect in Blocks 1 and 2 for the .75 compatibility proportion condition.

Table 3
Reliability of Compatibility Condition Response Times
Across Blocks

Scores Correlated	<i>r</i>
.25 Condition, <i>n</i> = 60	
Incompatible Blocks 1 & 2	.83
Compatible Blocks 1 & 2	.72
.50 Condition, <i>n</i> = 20	
Incompatible Blocks 1 & 2	.92
Compatible Blocks 1 & 2	.87
.75 Condition, <i>n</i> = 20	
Incompatible Blocks 1 & 2	.85
Compatible Blocks 1 & 2	.87

increased with increases in the compatibility proportion. This is a novel finding. We turn now to a discussion of the implications of this finding for the theories outlined by Toth et al. (1995) and Kane and Engle (2003).

The pattern of reliabilities found here does not support the predictions we derived from the Toth et al. (1995) account. We suggested that, if the associative automatic route were reliable, the .25- and .75-compatible conditions should have yielded *greater* reliability than the .50 condition, because performance in the former two conditions should be influenced to a greater extent by automatic processes (with both an automatic spatial correspondence and an associative automatic process contributing to the Simon effect). Alternatively, if the associative automatic route were unreliable, then the .25- and .75-compatible conditions should have yielded lower reliability than the .50 condition, since the .25 and .75 conditions are the only ones in which this route was argued to be operative. In contrast, we observed an increase in reliability as the proportion of compatible trials increased. This finding is not consistent with the hypothesis that reliability varies directly with the operation of the associative automatic process.

In contrast, the pattern of reliabilities observed was consistent with Kane and Engle's (2003) account of the Stroop effect as we have extended it to the Simon effect.¹ That is, our extension to this theory stated that as compatibility proportion increased, maintaining the goal of responding on the basis of the character rather than the location should become more difficult. This would, in turn, require subjects to devote more working memory and attentional control resources to the goal of responding to the character rather than to the location as the compatibility proportion increased. Our finding that the reliability of the Simon effect also increased as compatibility proportion increased suggests that the processes associated with maintaining the task goal (e.g., working memory and attentional control) are reliable. It should be noted, however, that although our results are consistent with this view, it has not yet been demonstrated that increases in compatibility proportion are associated with increases in working memory and attentional control demands or increases in their application. Future work should compare the reliability of performance in conditions under which it is known that working memory and attentional control are more operational with reliability in conditions in which they are less operational.

Conclusion

We have shown that the reliability of the Simon effect varies according to the proportion of compatible trials in an experiment. We tentatively suggest that reliability in the present context, at least, may be a function of the extent to which subjects need to apply more controlled processes, such as working memory and attentional control, in order to meet task goals. The present results thus highlight the importance of considering the relation between processes that support skilled performance. More generally, experimental psychologists typically tend to ignore the issues surrounding reliability when building theories. The present results, along with those reported by Stolz et al. (2005), suggest that exploring such issues is likely to be both informative and constraining with respect to such theory development.

AUTHOR NOTE

This work was supported by Grants A0998 and 183905 from the Natural Sciences and Engineering Council of Canada to D.B. and J.A.S., respectively. Correspondence relating to this article may be addressed to K. W. U. Borgmann (e-mail: kwuborgm@uwaterloo.ca).

Note—Accepted by David A. Balota's editorial team.

REFERENCES

- BUCHNER, A., & WIPPICH, W. (2000). On the reliability of implicit and explicit memory measures. *Cognitive Psychology*, **40**, 227-259.
- FAUST, M. E., BALOTA, D. A., SPIELER, D. H., & FERRARO, F. R. (1999). Individual differences in information-processing rate and amount: Implications for group differences in response latency. *Psychological Bulletin*, **125**, 777-799.
- HOMMEL, B. (1994a). Effects of irrelevant spatial S-R compatibility depend on stimulus complexity. *Psychological Research*, **56**, 179-184.
- HOMMEL, B. (1994b). Spontaneous decay of response-code activation. *Psychological Research*, **56**, 261-268.
- JACOBY, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory & Language*, **30**, 513-541.
- KANE, M. J., & ENGLE, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, **132**, 47-70.
- KORNBLUM, S., HASBROUCQ, T., & OSMAN, A. (1990). Dimensional overlap: Cognitive basis for stimulus-response compatibility—A model and taxonomy. *Psychological Review*, **97**, 253-270.
- LU, C.-H., & PROCTOR, R. W. (1995). The influence of irrelevant location information on performance: A review of the Simon and spatial Stroop effects. *Psychonomic Bulletin & Review*, **2**, 174-207.
- MADDEN, D. J., PIERCE, T. W., & ALLEN, P. A. (1993). Age-related slowing and the time course of semantic priming in visual word identification. *Psychology & Aging*, **8**, 490-507.
- NEELY, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264-336). Hillsdale, NJ: Erlbaum.
- NEELY, J. H., & KAHAN, T. A. (2001). Is semantic activation automatic? A critical re-evaluation. In H. L. Roediger III, J. S. Nairne, I. Neath, & A. M. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 69-93). Washington, DC: American Psychological Association.
- POSNER, M. I., & SNYDER, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 55-82). Hillsdale, NJ: Erlbaum.
- PSYCHOLOGY SOFTWARE TOOLS (2002). E-Prime [Software]. Pittsburgh: Author.
- SALTHOUSE, T. A., & SIEDLECKI, K. L. (2005). Reliability and validity of the Divided Attention Questionnaire. *Aging, Neuropsychology, & Cognition*, **12**, 89-98.

- SALTHOUSE, T. A., TOTH, J. P., HANCOCK, H. E., & WOODARD, J. L. (1997). Controlled and automatic forms of memory and attention: Process purity and the uniqueness of age-related influences. *Journals of Gerontology*, **52B**, P216-P228.
- SIMON, J. R. (1990). The effects of an irrelevant directional cue on human information processing. In R. W. Proctor & T. G. Reeve (Eds.), *Stimulus-response compatibility: An integrated perspective* (pp. 31-86). Amsterdam: North-Holland.
- STOLZ, J. A., BESNER, D., & CARR, T. H. (2005). Implications of measures of reliability for theories of priming: Activity in semantic memory is inherently noisy and uncoordinated. *Visual Cognition*, **12**, 284-336.
- TOTH, J. P., LEVINE, B., STUSS, D. T., OH, A., WINOCUR, G., & MEIRAN, N. (1995). Dissociation of processes underlying spatial S-R compatibility: Evidence for the independent influence of what and where. *Consciousness & Cognition*, **4**, 483-501.
- WILLIAMS, R. H., & ZIMMERMAN, D. W. (1996). Are simple gain scores obsolete? *Applied Psychological Measurement*, **20**, 59-69.

NOTE

1. It should be noted that although the extension of Kane and Engle's (2003) theory can account for the pattern of reliabilities reported here, it has no explanation for the reversal of the Simon effect observed in the .25 compatibility proportion condition. In contrast, the Toth et al. (1995) account can accommodate the reversal in the Simon effect.

(Manuscript received July 20, 2005;
revision accepted for publication June 8, 2006.)