AUTHOR'S PROOF JrnlID 13428_ArtID 784_Proof#1 - 26/07/2016

Behav Res
DOI 10.3758/s13428-016-0784-3

Statistical analyses of the resilience function

Joseph W. Houpt[1] · **Daniel R. Little**[2]

© Psychonomic Society, Inc. 2016

Abstract The extent to which distracting information influences decisions can be informative about the nature of the underlying cognitive and perceptual processes. In a recent paper, a response time-based measure for quantifying the degree of interference (or facilitation) from distracting information termed resilience was introduced. Despite using a statistical measure, the analysis was limited to qualitative comparisons between different model predictions. In this paper, we demonstrate how statistical procedures from workload capacity analysis can be applied to the new resilience functions. In particular, we present an approach to null-hypothesis testing of resilience functions and a method based on functional principal components analysis for analyzing differences in the functional form of the resilience functions across participants and conditions.

Keywords Response time · Information processing · Statistics · Conflict

Introduction

Understanding the time course of decision-making and behavior requires that we are able to make accurate inferences about how information is processed and integrated. Modern approaches to studying information processing aim to differentiate several general properties of information-processing systems. These properties can be categorized as follows: (1) Is information processed in sequence or simultaneously (i.e., in a serial or parallel architecture)? (2) Does the decision stop only after processing all of the information or can the decision terminate prior to that point (i.e., an exhaustive or self-terminating stopping rule)? (3) Is information processed independently or is there an interaction between processing channels? and (4) How does the processing efficiency change with increasing workload (i.e., the workload capacity of information processing)?

In this paper, we focus on a recently defined metric for resilience: How information-processing systems deal with conflicting information (Little et al., 2015, 2016); that is, information from multiple sources, which provides evidence for contrasting responses, actions, or decisions. Resilience, as demonstrated in Little et al., (2015, 2016) and summarized below, is affected by a combination of the four basic properties. For example, the presence of additional information, whether conflicting or not, affects workload (attribute 4 from the previous paragraph). If information is processed dependently, then contrasting information can inhibit processing (attribute 3). These influences and the influences of architecture (attribute 1) and stopping rule (attribute 2) are discussed in detail later.

Like the list of information-processing attributes above, the initial investigation of conflict relied on *qualitative* contrasts between a functional measure of resilience, $R(t)$, derived by Little et al. (2015). The goal of this paper is to introduce a set of quantitative tools for the quantitative assessment of the resilience function and the closely related conflict contrast function. We begin by demonstrating an approach to estimating these functions that has desirable statistical qualities. Next, we derive a null-hypothesis significance test for comparing resilience and conflict contrast

✉ Joseph W. Houpt
joseph.houpt@wright.edu

[1] Department of Psychology, Wright State University, Dayton, OH 45453, USA

[2] School of Psychological Sciences, The University of Melbourne, Melbourne, Australia

UNCORRECTED PROOF

functions to baseline models. Finally, we demonstrate an approach to formal exploratory analysis of the resilience and conflict contrast functions based on functional principal components analysis (fPCA; Ramsay & Silverman, 2005).

The main focus of these analyses is correct response times. The analysis of the error response times is complex; with multiple sources of error, one must consider how each source of information might fail. In some cases, failure of a local process may lead to an error response, whereas in other cases the system may be robust enough to protect itself against failure of any local process. Each of these situations needs to be carefully considered for each processing architecture. Townsend and Altieri (2012) presented such an extension for capacity, and it is possible that an extension might be possible for resilience. However, this is beyond the scope of the present paper.

Houpt and colleagues (Houpt & Townsend, 2012; Houpt et al., 2013) recently introduced statistical tests for a measure of workload capacity termed the capacity coefficient, $C(t)$, and Burns et al. (2013) demonstrated the use of fPCA for comparing among multiple $C(t)$ functions. The resilience function is based on similar functions of observed response times as the capacity coefficient; hence, the same statistical procedures can be leveraged for resilience analysis. The main distinction between the resilience function and the capacity coefficient is the experimental conditions used to obtain the response times that are used in the measure. The resilience function compares response times with congruent information to response times with incongruent information, whereas the capacity coefficient compares response times with congruent information to the sources of information in isolation. Little et al. (2015) show that with conflicting information, the resilience function reflects the speed of processing of the conflicting information. On its own, this measure allows only limited inference about processing architecture, but by contrasting conflicting information of different salience, one can gain substantial information about the underlying processing architecture. Hence, the statistical tools that are introduced here are developed to allow for testing not only resilience but also the difference between resiliency functions, $R_{diff}(t)$ and conflict contrast function.

We first describe the definition and motivation for the resilience and resiliency difference functions and then introduce the statistical tools necessary for testing the various qualitative contrasts between these functions.

Resilience and resiliency difference functions

Consider the question of whether a bat is a mammal or a bird? Although, the answer to this question should be obvious, the fact that bats share some similarity with birds makes this question harder than related questions which do not contain any conflict between biological properties and similarity. For example, is a robin a mammal or a bird? Many basic psychological tasks share an analogous conflict between two sources of information (see Fig. 1). In the categorization task that we use in this paper, a stimulus might contain multiple features some of which satisfy rules for one category and others which satisfy rules for a

UNCORRECTED PROOF

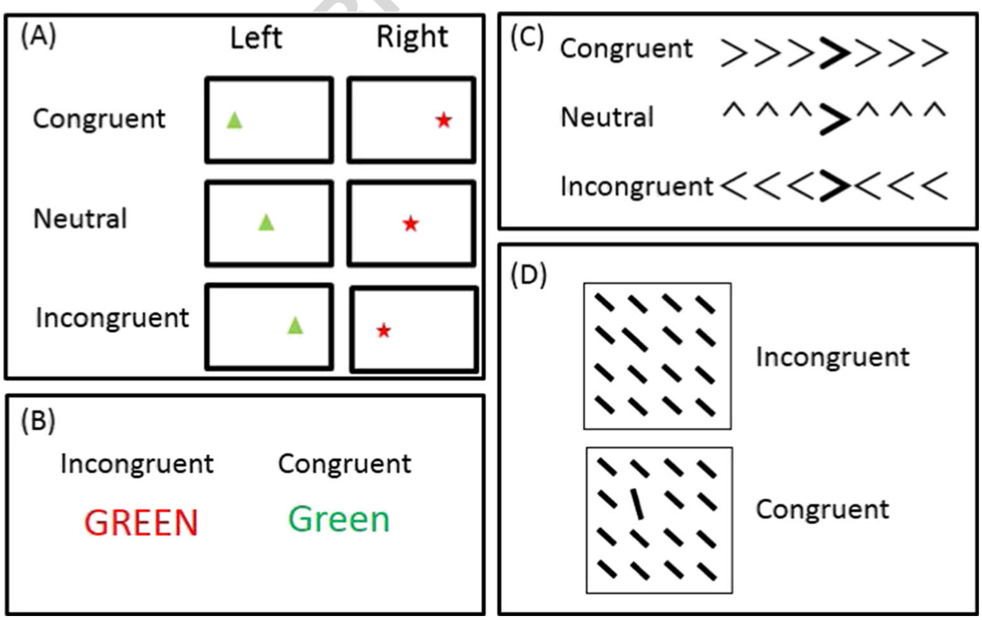

Fig. 1 Examples of tasks containing conflicting information. **a** Simon task: the color of the cue conflicts with its location in the incongruent condition. **b** Stroop task: the color name conflicts with the font color in the incongruent condition. **c** Flanker task: the central target is in conflict with the flanking distractors in the incongruent condition. **d** Oddball Search: the oddball target shares some information with the distractors in the incongruent condition

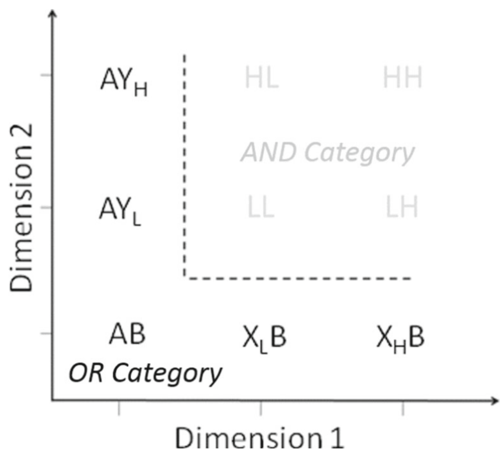

Fig. 2 Schematic illustration of a categorization structure containing conflicting information for some members of the OR category. The stimuli in the upper right quadrant of the space are the members of the AND category since member of this category need to have values greater than the vertical boundary on dimension 1 *and* the horizontal boundary on dimension 2. The remaining stimuli are the members of the OR category since members of this category have a value one dimension 1 less than the horizontal boundary *or* a value on dimension 2 less than the vertical boundary. For the AND category, H and L refer to the high- and low-discriminability dimension values, respectively. Values further from the boundary are easier to categorize. For the OR category, the redundant (AB) stimulus satisfies the OR rule on both dimensions. The remaining OR stimuli are indexed as a combination of one dimension value which satisfies one of the OR rules (either A for dimension 1 or B for dimension 2) and a dimension value which provides evidence for the AND category (X for dimension 1 and Y for dimension 2). The subscripts H and L for the OR category stimuli reflect whether the conflicting information provides evidence for the AND category of high or low discriminability, respectively. For example, the OR stimulus AY_L provides only weak evidence for the AND category on dimension 2 (i.e., because this dimension is close to the horizontal boundary on dimension 2)

different category (Allen and Brooks, 1991; Folstein et al., 2008; Nosofsky, 1991; Nosofsky & Little, 2010). In all of these tasks, the response times (RTs) for the *incongruent* trials, which contain conflicting information, are slower than the RTs for the *congruent* trials, which do not contain conflict information. However, simply finding the RT difference between responses to congruent and incongruent stimuli only allows for limited inference about processing. Our approach is to outline the conditions of congruency and incongruency that allow for strong inferences to be made about information processing. Namely, the resilience analysis demonstrates that varying the salience of the conflicting information allows for a contrast that can differentiate several important theoretical models.[1]

A schematic of a categorization task which contains the type of conflict considered here is shown in Fig. 2. In this task, observers must categorize the nine stimuli, which are created by orthogonally combining the three values on each dimension, into two categories that are defined by an "L-shaped" category. The category formed by the four stimuli in the top-right corner are defined by a conjunctive rule, and this category is consequently termed the AND category. That is, an item's membership in this category requires that it have a value on dimension 1 greater than the value indicated by the vertical bound *and* a value on dimension 2 greater than the value indicated by the horizontal bound. By contrast, the remaining stimuli are defined by a disjunctive rule applied to both dimensions. A decision about this category can be made by noting that an item has a value on dimension 1 less than the value indicated by the vertical bound *or* a value on dimension 2 less than the horizontal bound. This category is consequently termed the OR category.

The four stimuli in the AND category are coded by whether they have either low or high discriminability from the other category (i.e., as defined by distance from the category boundary). In a series of studies, Little and colleagues showed how these stimuli could be used to diagnose whether the processing of both stimulus dimensions occurred either in a serial or parallel fashion or, as a third alternative, pooled into a single processing channel (Blunden et al., 2015; Fific et al., 2010; Little et al., 2011; Little & Lewandowsky, 2012; Little et al., 2013; Moneer et al., in press). In the present paper, however, we focus on the items which belong to the OR category.

The OR category items are coded according to whether their component parts satisfy the disjunctive rule for the OR category, in which case the first dimension is coded *A* and the second dimension is coded *B* (see Fig. 2). Alternatively, one of the components of an OR category stimulus might satisfy only one of the disjunctive rules for the OR category; the other component, however, satisfies the rule for the AND category. We label these items with an X or a Y according to whether they satisfy the vertical or the horizontal rule for the AND category, respectively. Consequently, for most of the OR category items, there is a conflict or incongruency between the dimensions with one dimension providing evidence for the OR category and the other dimension providing evidence for the AND category.

This experimental design can be used an analogue for many tasks which contain conflicting information. Like the conflicting contrast category members (e.g., AY or XB), many tasks contain incongruent conditions that contain stimuli satisfying only one response rule. For example, in the Simon task (Proctor & Vu, 2006; Simon & Rudell, 1967), the location of the cue, which is irrelevant to the response, can be in conflict with the identity of that cue. The color satisfies the rule for determining the left-hand response but the location does not (see Fig. 1, panel a). In

[1]This approach is similar to how initial RT difference approaches to analyzing redundancy gains were extended using more theoretical methods including capacity (Miller, 1982; Townsend & Nozawa, 1995).

UNCORRECTED PROOF

the classic Stroop task (Stroop, 1935), the incongruent stimuli (e.g., the word "red" presented in GREEN) contain one source of information which provides evidence for the correct response (i.e., the color GREEN) and another providing evidence for an incorrect response (the word "red"). The color provides the correct response, but the word itself provides evidence for an incorrect response (see Fig. 1, panel b). In a flanker task, the central target might cue a right hand response but incongruent flankers provide a cue toward an erroneous left hand response (see Fig. 1, panel c) The processing of the distracting flankers interferes with responding and slows RT (Eriksen & Eriksen, 1974). Finally, in visual search, a target can share features with distractors (Duncan & Humphreys, 1989; see Fig. 1, panel d). The unique features signal that an item is a target, but the shared features provide evidence against this decision. Although each of these tasks involve different processes (e.g., with regard to attentional processes; Chajut et al., 2009; Shalev & Algom, 2000), the logical structure of conflict in these tasks is similar.

Little et al. (2015) showed how one could apply the capacity coefficient function to the compare performance on the congruent target, AB, to performance on the pair of incongruent stimuli, e.g., AY and XB (see Fig. 2), that satisfy only one of the disjunctive rules. The capacity coefficient was designed to evaluate the effect of increasing the workload of an information processing by comparing the processing of redundant (i.e., congruent) signals, e.g., AB, to the processing of each of those signals presented in isolation, A and B. When applied to the question of workload, under some basic assumptions (especially assuming independence between the processing channels), there are strong links between the observed capacity and the underlying processing architecture. For instance, unlimited-capacity, independent, parallel, (UCIP) self-terminating models, which predict that processing can terminate as soon as a target is detected predict that the time to process the redundant target should equal the minimum time derived from each of the single targets presented alone. In particular, for a UCIP model, $\log(S_{AB}(t)) = -\log(S_A(t) \times S_B(t))$, or in terms of the cumulative hazard function $(H(t) = -\log[S(t)])$, $H_{AB}(t) = H_A(t) + H_B(t)$. The capacity coefficient function (Eq. 1) compares observed performance with redundant targets to the performance predicted by a UCIP model (i.e., $-\log(S_A(t) \times S_B(t))$).

$$C(t) = \frac{-\log(S_{AB}(t))}{-\log(S_A(t) \times S_B(t))} = \frac{H_{AB}(t)}{H_A(t) + H_B(t)} \quad (1)$$

Consequently, a UCIP model predicts a capacity function of 1 across all t. If we assume that the processing time of the redundant target is unaffected by the presence or absence of a second signal, an assumption termed context invariance (cf. Miller, 1982; Townsend & Eidels, 2011), then serial self-terminating and serial exhaustive models predict capacity functions less than 1 (i.e., limited capacity; Townsend & Nozawa, 1995). By contrast, coactive models that pool information together predict capacity functions that are greater than 1 (i.e., supercapacity; Townsend & Nozawa, 1995; Townsend & Wenger, 2004). Parallel models with non-independent, interactive channels may predict capacity functions which are less than or greater than one depending on whether the interaction is inhibitory or facilitatory, respectively (Eidels et al., 2011; Townsend and Wenger, 2004).

Resilience

The same function can be applied to the present case where there is again a redundant target, AB, but in which the "single targets" are not presented alone but in the presence of conflicting information, AY and XB. Under these conditions, Little et al. (2015) showed that the function does not reflect changes in workload, but instead captures how quickly the conflicting information is processed relative to the target information. We term this function resilience, $R(t)$, to capture the idea that the function tells us something about how the system copes with conflicting information (see Eq. 2).

$$R(t) = \frac{-\log(S_{AB}(t))}{-\log(S_{AY}(t) \times S_{XB}(t))} = \frac{H_{AB}(t)}{H_{AY}(t) + H_{XB}(t)} \quad (2)$$

For example, consider the case in which the stimulus AX is processed in an independent parallel self-terminating fashion. The decision time (for correct decisions) is still determined by the time taken to process dimension A (and likewise, the processing XB only depends on B under the UCIP model); consequently, the derived minimum time and consequently the value of, $R(t)$, remains unchanged under the assumption of UCIP processing. For $R(t)$, the UCIP model can again take on the role of a baseline model for comparison. If the dimensions are processed in a serial fashion, then the distracting information when AY is presented has some probability of being processed before the target information, hence slowing the overall processing time relative to A alone and increasing H_{AY}, or the distracting information when XB is present has some probability of being processed first and H_{XB} increases. This implies that the denominator in Eq. 2 will be smaller than predicted by the UCIP and results in an $R(t)$ function which is greater than 1. However, because the redundant targets do not benefit from statistical facilitation, as with a UCIP model, the numerator will also be smaller, indicating $R(t)$ could also be less than 1.

More generally, if the target information is processed *faster* when distractor information is present, then the derived minimum time might be faster than the redundant

UNCORRECTED PROOF

target processing time, resulting in an $R(t)$ function which is less than 1. If the target information is processed *slower* when distractor information is present, then the derived minimum time might be slower than the redundant target processing time, resulting in $R(t) > 1$. With conflicting or distracting information present in the single target stimuli, the link between architecture and the value of the function is less clear cut than for the capacity coefficient.

Resiliency difference function

The ambiguity in how resilience reflects architecture can be resolved by noting that the discriminability or strength of the conflicting information determines the effect of the conflict on the derived minimum time. In a UCIP model, there is no effect of the conflicting information, but in a serial, self-terminating model, faster-processed conflict information results in a faster derived minimum time than slower processed conflict information. The category space in Fig. 2 effectively manipulates the discriminability of the conflict information by varying the distance from the boundary for items along both the horizontal boundary (e.g., AY_L and AY_H) and the vertical boundary (e.g., $X_L B$ and $X_H B$; see Ashby & Gott, 1988; Fific et al., 2010). The change in the derived minimum time with the discriminability of the distracting item implies that, under the assumption that the discriminability manipulation is effective, that the resiliency functions will be ordered for a serial model with the $R_H(t)$ function being lower than the $R_L(t)$ function. By contrast, a coactive model predicts the opposite ordering: The stronger the evidence for the AND category, the slower the derived minimum time. Consequently, for a coactive model, the $R_H(t)$ should be larger than the $R_L(t)$ function because of the slowed derived minimum time. These relations are shown in Fig. 3 (top panel).

This ordering of resiliency functions suggests that the difference between the resilience function computed from the high and low conflict items can provide a diagnostic of the underlying processing architecture. Little et al. 2015 introduced the resilience difference function, $R_{diff}(t)$, as follows:

$$R_{diff}(t) = R_H(t) - R_L(t) = \frac{H_{AB}(t)}{H_{AY_H}(t) + H_{X_H B}(t))} - \frac{H_{AB}(t)}{H_{AY_L}(t) + H_{X_L B}(t)}. \quad (3)$$

The predictions of this function are shown in Fig. 3 (bottom panel).

A large set of different models can be differentiated based on the value of the $R_{diff}(t)$ function. Consequently, this function can be added to a growing set of theoretical and methodological tools, termed Systems Factorial Technology, which includes, among others, the capacity coefficient (Townsend & Nozawa, 1995), the single-target capacity function (Blaha & Townsend, 2014), the mean interaction contrast, and survivor interaction contrasts (which can be applied to, for example, the factorial combination of discriminabilties in the AND category; Townsend & Nozawa, 1995). Following Houpt and Townsend (2010, 2012), the goal of the remainder of this paper is to introduce methods for providing significance tests for the resilience and resilience difference functions.

Little et al. (2016) presented an alternative form of the resilience difference function known as the conflict contrast function, $CCF(t)$. This function takes advantage of the fact that the ordering of the derived minimum time is preserved even without considering the double target, AB. Consequently, a simple contrast of the RTs for the high and low conflict stimuli can be computed as follows:

$$CCF(t) = \left[H_{AY_L}(t) - H_{AY_H}(t)\right] + \left[H_{X_L B}(t) - H_{X_H B}(t)\right] \quad (4)$$

This function has the benefit of predicting the same qualitative distinctions between the models as shown in Fig. 3

UNCORRECTED PROOF

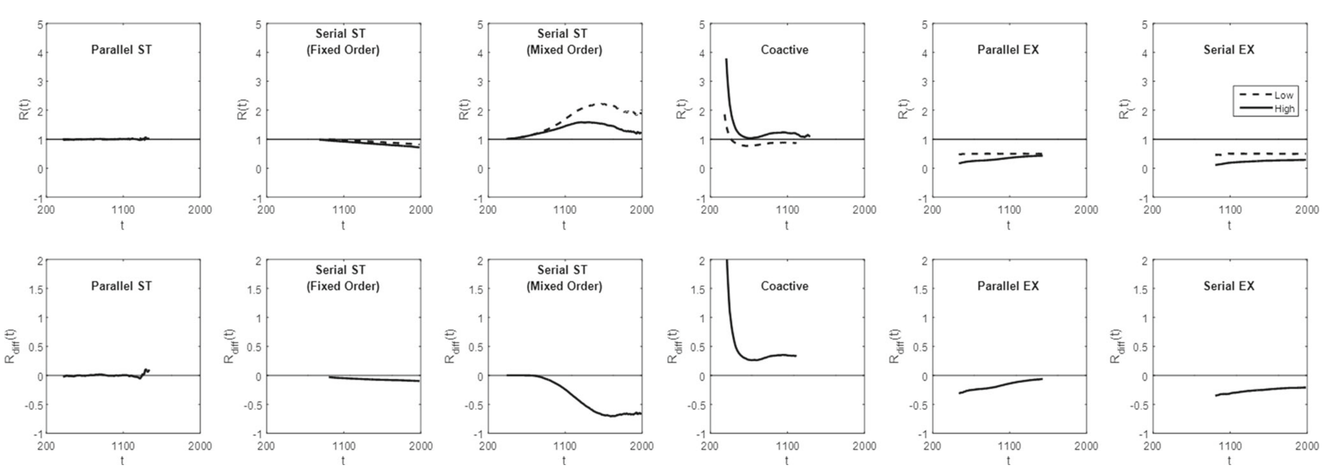

Fig. 3 *Top*: Ordering of resilience functions based on the discriminability of the conflict items. *Bottom*: Resilience difference functions

(bottom panels) but allows for the application of the contrast to tasks where it may not be natural to include a double target (e.g., in the Simon task, see Fig. 1, the incongruent and neutral stimuli can be used as the high and low salience conflict items, respectively). In the following, we also provide the relevant statistics for the $CCF(t)$ function.

Estimation

The first step in developing a hypothesis test for the resilience difference function and the conflict contrast function is to determine the appropriate estimator. One approach would be to bin the observed response times to estimate the probabilities, then sum them to estimate the survivor function, and finally take the natural log to estimate each term (cf. Wenger & Townsend, 2000). Alternatively, we can use the fact that the negative log of the survivor function is equal to the cumulative hazard function, which is in turn the integral of the density divided by the survivor function,

$$-\log S(t) = H(t) = \int_0^t \frac{f(s)}{S(s)}\, ds. \tag{5}$$

To estimate the survivor function for correct response times we can use one minus the empirical cumulative distribution function (ECDF), a well-established estimator (e.g., Parzen, 1962). The basic idea of the ECDF is to estimate the probability that a response time occurs at or before a given time by the proportion of observed correct response times that were faster than that time. Formally,

$$\hat{S}(t) = 1 - \hat{F}(t) = 1 - \frac{1}{n}\sum_{i=1}^{n} I\left(T_i \le t\right) = \frac{1}{n}\sum_{i=1}^{n} I\left(T_i > t\right).$$

Here, n is the total number of observed correct response times used to estimate the ECDF, T_i is one of the observed correct response times, and $I(\cdot)$ is an indicator function which is 1 if the argument is true and zero otherwise.

The next step is to estimate the density. This simplest approach is to use $\hat{f}(t) = 1/n$ whenever t is equal to an observed correct response time and $\hat{f}(t) = 0$ for all other times,

$$\hat{f}(t) = \begin{cases} 1/n & \text{if } t = T_i \text{ for some } i \\ 0 & \text{otherwise.} \end{cases}$$

With this estimator of the density, the integral in Eq. 5 becomes a sum over all of the times $s < t$ at which there was a correct response,

$$\hat{H}(t) = \sum_{T_i < t} \frac{1/n}{\hat{S}(T_i)} = \sum_{T_i} \frac{1}{\sum_{j=1}^{n} I\left(T_j > T_i\right)}. \tag{6}$$

Equation 6 is known as the empirical cumulative hazard function (ECH). The ECH could be used in Eqs. 2, 3 and 4, however if there are incorrect responses or cases in which the participant does not respond in time the ECH will be biased. One approach used by Houpt et al. (2013) is to mitigate that bias by treating time-outs and incorrect response times as censoring, e.g., assuming that if the participant had more time or if they had not already made an incorrect response, they would eventually choose the correct response. This leads to a generalization of the ECH known as the Nelson-Aalen estimator of the cumulative hazard function (NAH, Andersen et al., 1993; Aalen et al., 2008).

The NAH is essentially the same as the ECH, but with the sum in the estimated survivor function, $\hat{S}$ replaced with a sum over *all* response times instead of only correct response times. To clean up the notation a bit, we use bold notation for a set of times with a subscript indicating if there is a bound on that set, e.g., $\mathbf{T}_{\le t}$ is the set of response times less than or equal to t. If we wish to indicate only correct response times, we use the superscript c, e.g., $\mathbf{T}^c_{>t}$ are the correct response times that occurred after t. This allows us to write the NAH as,

$$\hat{H}(t) = \sum_{s \in \mathbf{T}^c_{\le t}} \frac{1}{\sum_{r \in \mathbf{T}} I(r > s)}. \tag{7}$$

The NAH has a number of useful statistical properties (for details, see Andersen et al., 1993; Aalen et al., 2008). It is an unbiased estimator of the true cumulative hazard function.[2] Furthermore, the variance of the difference between the NAH and the true cumulative hazard function is straightforward to calculate. Using $Y(s)$ for $\sum_{r \in \mathbf{T}} I(r > s)$i,

$$\hat{\sigma}^2_H(t) = \sum_{s \in \mathbf{T}^c_{\le t}} \frac{1}{Y^2(s)}.$$

Also, the NAH is a uniformly consistent estimator of the true cumulative hazard function, and the difference between the NAH and the true cumulative hazard function converges in distribution to a zero mean Gaussian process.

Another particularly useful fact is that finite linear combinations of uncorrelated NAHs are again unbiased, uniformly consistent, and the difference between the estimate of the linear combination and the true linear combination is a mean zero Gaussian process with variance (with arbitrary coefficients a_m),

$$\text{Var}\left(\sum_{i=1}^{m} a_m \hat{H}_m(t)\right) = \sum_{i=1}^{m} a_i^2 \hat{\sigma}^2_{H_i}(t). \tag{8}$$

Null-hypothesis testing

With a well-defined estimator for terms in the resilience, we can focus on hypothesis testing. Like Houpt and Townsend (2012), we will stick to differences of cumulative hazard

[2] Technically, this statement and the variance statement are only true for t up to the time the last observed response occurs.

UNCORRECTED PROOF

functions for hypothesis tests rather than ratios. In particular, that means instead of testing $R(t) = 1$, we test if the difference between the numerator and denominator of $R(t)$ is zero. Of course if the ratio between the numerator and denominator is 1 then the difference is zero. We will also focus on null-hypothesis tests for the CCF rather than on the resiliency difference function.

A null-hypothesis test may not always be appropriate for analyzing resilience functions for many of the same reasons null-hypothesis tests are avoided in other contexts. In particular, these tests treat the null-hypothesis differently than other alternatives so the outcome of a null-hypothesis test should not be interpreted as a model comparison. Like all other null-hypothesis tests, these tests cannot offer evidence in favor of the null. If one is interested in model comparison questions, particularly in relative evidence for the null model, the semiparametric Bayesian analysis proposed by Houpt et al. (2016) offers promise, although its application to resilience analyses are beyond the scope of this paper.

The resilience function

Our first step is to encode the null hypothesis of UCIP processing into a statement about the estimators. Under the UCIP model, the processing time survivor function should be the same for A (B) regardless of the context, i.e., $S_{AY}(t) = S_A(t)$ ($S_{XB}(t) = S_B(t)$.

Additionally, if the elements are processed in parallel, then $S_{AB}(t) = S_A(t)S_B(t)$. By taking the negative natural logarithm of both sides, we get,

$$\begin{aligned} H_{AB}(t) &= -\log(S_{AB}(t)) = -\log(S_A(t)) - \log(S_B(t)) \\ &= H_A(t) + H_B(t). \end{aligned}$$

Replacing $H(t)$ with its estimator, we arrive at the null hypothesis in terms of observable quantities:

$$\text{H0: } \hat{H}_{AB}(t) - \hat{H}_{AY}(t) - \hat{H}_{XB}(t) = 0. \tag{9}$$

From the previous section, we know that the limit distribution of each of the terms on the left hand side, and hence their linear combination, is Gaussian. Thus, to get a test statistic distribution, we only need to determine the mean and variance. Because the NAH is an unbiased estimator, under the null hypothesis the expected value of Eq. 9 is zero for all t. Because the data used to estimate each term in Eq. 9 is independent, we can use Eq. 8 to determine the variance,

$$\begin{aligned} &\text{Var}\left[\hat{H}_{AB}(t) - \hat{H}_{AY}(t) - \hat{H}_{XB}(t)\right] \\ &= \text{Var}\left[\hat{H}_{AB}(t)\right] + \text{Var}\left[\hat{H}_{AY}(t)\right] + \text{Var}\left[\hat{H}_{XB}(t)\right]. \end{aligned}$$

This allows us to calculate a statistic for any fixed time[3] t that, under the null-hypothesis, has a standard normal distribution,

$$R' = \frac{\hat{H}_{AB}(t) - \hat{H}_{AY}(t) - \hat{H}_{XB}(t)}{\sqrt{\text{Var}\left[\hat{H}_{AB}(t)\right] + \text{Var}\left[\hat{H}_{AY}(t)\right] + \text{Var}\left[\hat{H}_{XB}(t)\right]}} \xrightarrow{d} \mathcal{N}(0, 1).$$

For testing cases when the entire resilience function is expected to be either above, equal to, or below one for all t, a single test at the largest possible response time (t_m) is most sensible because it uses the largest amount of data. For this reason, in all of the null-hypothesis testing reported below, we use a single z-test at the maximum possible time.

The conflict contrast function

Here again we use the UCIP first-terminating model as the null hypothesis. In terms of the estimated cumulative hazard functions,

$$\text{H0: } \left[\hat{H}_{AY_H}(t) - \hat{H}_{AY_L}(t)\right] + \left[\hat{H}_{X_H B}(t) - \hat{H}_{X_L B}(t)\right] = 0.$$

Again, the limit distribution of each term is Gaussian and the estimators are unbiased and consistent so the limit of the distribution has mean 0. The estimate of the variance is unbiased and consistent, so dividing the difference by the sum of the variances results in limit distribution with unit variance. Together, this implies that, under the null hypothesis,

$$CC' = \frac{\left[\hat{H}_{AY_H}(t) - \hat{H}_{AY_L}(t)\right] + \left[\hat{H}_{X_H B}(t) - \hat{H}_{X_L B}(t)\right]}{\sqrt{\text{Var}\left[\hat{H}_{AY_H}(t)\right] + \text{Var}\left[\hat{H}_{AY_L}(t)\right] + \text{Var}\left[\hat{H}_{X_H B}(t)\right] + \text{Var}\left[\hat{H}_{X_L B}(t)\right]}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Weighting functions

Following Aalen et al. (2008), Houpt and Townsend (2012) also demonstrated the possibility of using weighting functions with the hypothesis test to emphasize different regions of time. One such weighting function is the Harrington-Fleming function,

$$L(t) = S_{\text{KM}}(t)^{\rho} \frac{Y_{AB}(t)\left[Y_{AY}(t) + Y_{XB}(t)\right]}{Y_{AB}(t) + Y_{AY}(t) + Y_{XB}(t)}.$$

[3] Or any time that is chosen based only on information up to that time (formally, any *stopping time*; see Houpt & Townsend, 2012 for details).

UNCORRECTED PROOF

Here, $S(t)$ is left-continuous version of the Kaplan-Meier estimate of the survivor function for the pooled response times, $\hat{S}_{\mathrm{KM}}(t) = \prod_{t_i < t} \left(|\mathbf{T}_{>t_i}| - 1\right) / \left(|\mathbf{T}_{>t_i}|\right)$. With $\Delta N(s)$ indicating the number of correct responses times that occurred at time s,

$$S(t) = \prod_{s \in \mathbf{T}^c_{<t}} \left(1 - \frac{\Delta N(s)}{Y_{AB}(s) + Y_{AY}(s) + Y_{XB}(s)}\right).$$

The parameter ρ can be chosen to emphasize lower response times more (larger ρ) or less (smaller ρ).

When the weighting function is used, the numerator of R' is replaced with,

$$\sum_{s \in \mathbf{T}^{AB,c}_{\leq t}} \frac{L(s)}{Y_{AB}(s)} - \sum_{s \in \mathbf{T}^{AY,c}_{\leq t}} \frac{L(s)}{Y_{AY}(s)} - \sum_{s \in \mathbf{T}^{XB,c}_{\leq t}} \frac{L(s)}{Y_{XB}(s)}.$$

The denominator of R' is replaced with,

$$\sqrt{\sum_{s \in \mathbf{T}^{AB,c}_{\leq t}} \frac{L(s)}{Y^2_{AB}(s)} + \sum_{s \in \mathbf{T}^{AY,c}_{\leq t}} \frac{L(s)}{Y^2_{AY}(s)} + \sum_{s \in \mathbf{T}^{XB,c}_{\leq t}} \frac{L(s)}{Y^2_{XB}(s)}}.$$

Hence, we define the resilience statistic with a weighting function as,

$$R = \frac{\sum_{s \in \mathbf{T}^{AB,c}_{\leq tm}} \frac{L(s)}{Y_{AB}(s)} - \sum_{s \in \mathbf{T}^{AY,c}_{\leq tm}} \frac{L(s)}{Y_{AY}(s)} - \sum_{s \in \mathbf{T}^{XB,c}_{\leq tm}} \frac{L(s)}{Y_{XB}(s)}}{\sqrt{\sum_{s \in \mathbf{T}^{AB,c}_{\leq tm}} \frac{L(s)}{Y^2_{AB}(s)} + \sum_{s \in \mathbf{T}^{AY,c}_{\leq tm}} \frac{L(s)}{Y^2_{AY}(s)} + \sum_{s \in \mathbf{T}^{XB,c}_{\leq tm}} \frac{L(s)}{Y^2_{XB}(s)}}}. \quad (10)$$

An analogous weighting function for the CCF is given by

$$L_C(t) = S(t)^{\rho} \frac{\left[Y_{AY_H}(t) + Y_{X_H B}(t)\right]\left[Y_{AY_L}(t) + Y_{X_L B}(t)\right]}{Y_{AY_H}(t) + Y_{XB_H}(t) + Y_{AY_L}(t) + Y_{XB_L}(t)}.$$

The numerator of CC' is replaced with,

$$\left[\sum_{s \in \mathbf{T}^{AY_H,c}_{\leq t}} \frac{L_C(s)}{Y_{AY_H}(s)} - \sum_{s \in \mathbf{T}^{AY_L,c}_{\leq t}} \frac{L_C(s)}{Y_{AY_L}(s)}\right] + \left[\sum_{s \in \mathbf{T}^{X_H B,c}_{\leq t}} \frac{L_C(s)}{Y_{X_H B}(s)} - \sum_{s \in \mathbf{T}^{X_L B,c}_{\leq t}} \frac{L_C(s)}{Y_{X_L B}(s)}\right].$$

The denominator of CC' is replaced with,

$$\sqrt{\sum_{s \in \mathbf{T}^{AY_H,c}_{\leq t}} \frac{L_C(s)}{Y^2_{AY_H}(s)} + \sum_{s \in \mathbf{T}^{AY_L,c}_{\leq t}} \frac{L_C(s)}{Y^2_{AY_L}(s)} + \sum_{s \in \mathbf{T}^{X_H B,c}_{\leq t}} \frac{L_C(s)}{Y^2_{X_H B}(s)} + \sum_{s \in \mathbf{T}^{X_L B,c}_{\leq t}} \frac{L_C(s)}{Y^2_{X_L B}(s)}}.$$

Likewise, we define the conflict-contrast statistic as,

$$CC = \frac{\left[\sum_{s \in \mathbf{T}^{AY_H,c}_{\leq t}} \frac{L_C(s)}{Y_{AY_H}(s)} - \sum_{s \in \mathbf{T}^{AY_L,c}_{\leq t}} \frac{L_C(s)}{Y_{AY_L}(s)}\right] + \left[\sum_{s \in \mathbf{T}^{X_H B,c}_{\leq t}} \frac{L_C(s)}{Y_{X_H B}(s)} - \sum_{s \in \mathbf{T}^{X_L B,c}_{\leq t}} \frac{L_C(s)}{Y_{X_L B}(s)}\right]}{\sqrt{\sum_{s \in \mathbf{T}^{AY_H,c}_{\leq t}} \frac{L_C(s)}{Y^2_{AY_H}(s)} + \sum_{s \in \mathbf{T}^{AY_L,c}_{\leq t}} \frac{L_C(s)}{Y^2_{AY_L}(s)} + \sum_{s \in \mathbf{T}^{X_H B,c}_{\leq t}} \frac{L_C(s)}{Y^2_{X_H B}(s)} + \sum_{s \in \mathbf{T}^{X_L B,c}_{\leq t}} \frac{L_C(s)}{Y^2_{X_L B}(s)}}}. \quad (11)$$

Because $L(t)$ and $L_C(t)$ are non-negative, measurable processes, the limit distribution of the statistics are unchanged, so $R \xrightarrow{d} \mathcal{N}(0, 1)$ and $CC \xrightarrow{d} \mathcal{N}(0, 1)$ (cf. Aalen et al., 2008, Chapter 3).

Simulation study

In this section we explore performance of the R and CC statistics on simulated data sets for which we know the ground truth. First, we will examine the extent to which reasonably sized samples from model which predicts null effects are represented by derived statistics. In particular, we will test whether the Type I error rates are approximately 0.05 for α at that level (which we use for all simulations below). Second, we will examine the statistical power for two types of effects: the categorical effect of having a model other than the null (i.e., not parallel ST) and the ratio scale effect of moderating the rate of processing when distractors are present in a parallel ST model.

Following Houpt and Townsend (2012), we simulated data assuming the underlying processing time distributions were exponential. Additionally, we tested the statistics on data generated from the Linear Ballistic Accumulator Model (LBA, Brown & Heathcote, 2008). This allowed us to explore the power with more realistic response time distributions as well as explore the effect of higher error rates. Each simulated dataset consisted of 1000 samples. For each simulated data set, we tested power with five different levels

UNCORRECTED PROOF

of ρ ranging from zero (corresponding to a log-rank test) to one (corresponding to Wilcoxon test, cf. Aalen et al., 2008, p. 107).

In theory it is possible to achieve arbitrary precision on estimates of the effects of number of trials, rate factor, model type and ρ, however in practice we are limited by the resources available for running simulations. Although 1000 samples per combination of factors allows for quite high precision, we also applied Bayesian linear regression models to quantify the evidence in favor of, or against, an effect of the factors of interest (cf. Rouder & Morey, 2012).

Exponential model R

For the exponential model, each correct subprocess completion time was sampled from an exponential distribution with rate 0.69 for the targets and 0.93 for the contrast stimuli. For each combination of parallel/serial and exhaustive/first-terminating, the simulated subprocess completion times were combined using the appropriate rule (e.g., the minimum of the subprocess completion times for parallel, first-terminating processing of the redundant targets). We calculated the resilience statistic for each model using $\rho = \{0, .2, .4, .6, .8, 1\}$ and the number of trials per distribution ranging from ten to 150 in increments of ten.

First, as a confirmation that distribution of R converges to a Gaussian relatively quickly, we found that the rate of significant findings for the two-tailed test of R for the parallel, self-terminating for all ρ and all numbers of trials at between 0.030 and 0.067 percent of the generated samples. There was no evidence that increases either in ρ or the number of trials led to increases or decreases in the rate of significance for R (BF= 0.729 and BF= 1.05, respectively).

For the parallel, exhaustive model, the rate of significance increased from 0.32 with ten trials and reached asymptote of nearly 1.0 around 80 trials. Averaged across the number of trials, ρ had nearly no effect. A Bayes factor test comparing linear models of main effect of number of trials and ρ as well as an interaction indicated only the number of trials as an important factor (BF= 51.0 over the next best, which included an interaction and both main effects).

With the data generated from a serial, self-terminating model, the Bayes factor test again indicated only the number of trials as an important factor (BF= 2849 over the next best model). The rate of significance increased linearly from 0.067 with ten trials to 0.55 with 150 trials.

The Bayes factor test also indicated only the number of trials as an important factor for the serial-exhaustive data (BF= 51.7 over the next best model). Like the parallel-exhaustive data, the rate of significance rose from .33 with ten trials to an asymptote of nearly 1.0 with 80 trials.

To test the effect of distractor interference, we also simulated a decreasing rate of processing in each of the channels when they were used together in a parallel, self-terminating model. There was no effect of ρ so the following results are averaged across values of ρ. For small levels of interference (90 % efficiency), power increased linearly but only reached 0.15 by 150 trials. As interference increased, the rate of increase in power as a function of number of trials increased and became less linear due to the upper bound of perfect power. For only ten trials, power was good for the highest levels of interference (0.77 with 30 % efficiency). To achieve power higher than 0.8 for moderate interference (70 % efficiency), at least 110 trials were needed.

Across all of the simulations, the number of trials had a clear effect on the power, with 80 trials per distribution being sufficient for nearly perfect power for parallel and serial exhaustive models, and 110 trials sufficient to detect moderate distractor interference, but more than 150 trial necessary for good power on a serial first-terminating model. There was no indication of an effect of ρ, which may in part be due to the fact exponential random variables have a flat hazard function across time (recall that ρ differentially weights earlier versus later response times in calculating R), although only the parallel, first-terminating model maintains the flat hazard rate when the two sub-processes are combined.

Exponential model CC

For the CC, we tested a range of increases in rates from low to high speed (five levels from 1.2 times to 2.0 times the rate) in addition to testing the effects of varying architecture, stopping-rule, ρ and number of trials.

Across all simulations with the parallel self-terminating model, the 0.048 of the simulation runs were significant. There was evidence for an effect of increasing the number of trials leading to a small increase (3.14×10^{-05} per trial; 95 % HPD =$[1.63 \times 10^{-05}, 4.65 \times 10^{-05}]$) in the number of simulation runs that were significant ($BF = 5.42$ over the next best model, which included rate as a factor as well).

In the parallel exhaustive data, there was evidence for an interaction between rate and number of trials (the increase in power as a function of number of trials increased faster with higher rates) and all of the main effects (BF = 3.73 over the next best model which also included a ρ by rate interaction). Power increased as a function of rate (0.68 per unit, HDI = [0.65, 0.72]), ρ (0.045 per unit, HDI = [0.020, 0.072]) and number of trials (0.0045 per trial, HDI = [0.0042, 0.0047]).

For the serial first-terminating data, all of the two-way interactions were included in the best model, along with main effects, but not the three-way interaction (BF = 7.37 over the next best model, which dropped the ρ by rate

UNCORRECTED PROOF

interaction). Like the parallel exhaustive data, the power increased as a function of number of trials increased faster with higher rates. The larger ρ was, the lower the increase in power as a function of number of trials and as a function of the rate factor. Overall, increases in ρ led to decreases in power (-0.029, HDI $= [-0.046, -0.012]$) while increases in the rate (0.35, HDI $= [0.33, 0.37]$) and number of trials (0.0023, HDI $= [0.0021, 0.0024[$) led to increases in power.

The most likely model for the serial-exhaustive data was the same as for the parallel exhaustive data, an interaction between the rate and number of trials and all three main effects (BF $= 3.73$ over the next best model which added a ρ by rate interaction). The interaction between rate and number of trials had the same qualitative effect as it did for the exhaustive data, an increase in the rate led to a larger increase in power per trial. An increase in rate increased power (0.68, HDI $= [0.65, 0.71]$) as did an increase in the number of trials (0.0045, HDI $= [0.0042, 0.0047]$) and ρ (0.046, HDI $= [0.019, 0.072]$).

The statistic had decent power when the rate of processing in a parallel self-terminating model that was affected by the distractors. For large changes in rate (i.e., the rate with distractors was less than 50 % or more than 200 % of the processing rate without distractors) approximately 40 trials per condition were sufficient to achieve 0.8 power. For moderate changes in rate due to the presence of a distractor (i.e., the rate was between 60 % and 70 % or 140 % and 160 %) approximately 120 trials per condition were necessary to achieve a power of 0.80. For smaller changes (80 % or 125 %) power was approximately 0.50 even with 150 trials per distribution.

LBA model R

To explore the power for R and CC in data that looks more like human response times, and particularly does not have a flat hazard function across time, we also simulated data from the Linear Ballistic Accumulator model (Brown & Heathcote, 2008). We used 0.69 as the mean accumulation rate parameters for the targets, 0.93 as the mean accumulation rate for the contrast stimuli, 0.1 for the standard deviation of the accumulation rate, 0 for the base time and 0.5 for both the incorrect and correct thresholds.

The rate of significance for the parallel self-terminating model was low, ranging between 0.035 and 0.073 across all numbers of trials and values of ρ. The parallel, exhaustive model was significant on nearly every run with ten trials; only non-significant 3 times out of 6000 runs across all ρ values and was significant on every run for 20 or more trials. There was no room for the ρ to have any effect due to the high rate of significance. The serial first-terminating model was significant on only 0.080 of the runs with ten trials but increased to 0.93 with 150 trials and there was no effect of ρ. Like the parallel, exhaustive model, the power was quite high with only ten trials, 0.996 and there were no non-significant runs with 20 or more trials. In the coactive model, power ranged from 0.23 with ten trials to perfect performance, reaching 0.99 by 90 trials per distribution. Again, there was no effect of ρ.

The power to detect that the rate of processing in a parallel self-terminating model was affected by the distractors was nearly identical to that found with the exponential simulation. For large changes in rate 40 trials or fewer per condition were sufficient to achieve 0.8 power. For moderate changes in rate due to the presence of a distractor, approximately 120 trials per condition were necessary to achieve a power of 0.80. For smaller changes, power was approximately 0.50 even with 150 trials per distribution.

The power of the resilience test for the LBA data was generally quite good. Only ten trials per distribution were sufficient for nearly perfect power for parallel and serial exhaustive models, and 40 were trials sufficient to detect high levels of distractor interference. The coactive model had lower power, needing 90 or more trials per distribution to reach power of essentially 1 and performance was worst with the serial first-terminating which only reached 0.93 with 150 trials. Despite the LBA having non-constant hazard rate, there was still no indication of a meaningful effect of ρ.

UNCORRECTED PROOF

Exponential model CC

For the CC, we tested a range of increases in rates from low to high speed (five levels from 1.2 times to 2.0 times the rate) in addition to testing the effects of varying architecture, stopping-rule, ρ and number of trials.

Across all simulations with the parallel self-terminating model, the 0.056 of the simulation runs were significant. The rate of significance was stable across all levels of ρ, numbers of trials and rate increase factor.

In the parallel exhaustive data, there was an increase in the power with an increase in number of trials (from 0.16 to 0.86 averaged across ρ and rate factor) and with an increase in the rate factor (from 0.24 to .90 averaged across the other factors). Additionally, the increase in power as a function of number of trials increased faster with higher rates. There was no evidence of an effect of ρ.

LBA model CC

For the serial first-terminating data, ρ did have an effect: lower ρ values led to higher power and faster increases in power as a function of the other variables with $\rho = 0$ giving the best performance. With $\rho = 0$, the power was 0.15 with ten trials increasing to 0.88 with 150 average across rate factor. Increased rate also increased power, from 0.58 to 0.88 across the levels tested and averaged across number of trials, although it had little additional benefit beyond 1.6. There was again an interaction in that power increased faster across trials with larger rate factors, up to 1.6.

The serial exhaustive data indicated an effect of increasing the number of trials, from 0.32 to 1.0 by 100 trials, and rate factor, from 0.77 to 0.94 for 1.6 and above, but not ρ. Increasing the rate factor again increased the rate at which increasing trials increased power, up to the rate factor of 1.6 after which there was no difference.

The coactive model was easily distinguished with a power of 0.75 with the lowest rate factor and ten trials and 0.98 and above for the rest of the simulated conditions. There was no evidence of an effect of ρ.

fPCA of the resilience function

In some cases, it may be useful to examine the overall shape of a resilience function or conflict contrast function, particularly as it varies across individuals or tasks (for example, the simulated results in Fig. 3 indicate that shape may vary with processing strategy). Recently, Burns et al. (2013) demonstrated the use of functional principle components analysis (fPCA) for extracting important features of the capacity coefficient function. Like the capacity coefficient statistics, we can also adapt the fPCA approach for both resilience and conflict contrast functions.

Summary of fPCA approach

The main idea of functional principle components analysis is exactly the same as the more familiar principle components analysis. Each datum is represented as a linear combination of bases, where the bases are chosen such that variation across data along the first basis is maximized, then each subsequent basis is chosen such that variation is maximized subject to the constraint that the basis is orthogonal to all previously chosen basis. The distinctive feature of fPCA compared to standard PCA is that the bases are *functions* (or infinite dimensional vectors) rather than finite length vectors. Ramsay and coauthors have a series of books on functional data analysis, including fPCA, for the interested reader (Ramsay and Silverman, 2005; Ramsay et al., 2009).

The basic procedure is to first subtract the mean function (averaged across individuals, conditions, etc.; not averaged across time) from each of the collected functions. Next, to find the basis along which the most variation across sample functions occurs, we solve for the weighting function $\xi_1(t)$ that maximizes $\sum_i (\xi_1(t)x_i(t)\,dt)^2$ subject to $\int \xi_1^2(t)\,dt = 1$, where $x_i(t)$ are the resilience (or conflict contrast) functions. The subsequent basis functions are found in a similar manner, ξ_j is chosen to maximize $\sum_i \left(\xi_j(t)x_i(t)\,dt\right)^2$ subject to $\int \xi_j^2(t)\,dt = 1$ and the orthogonality constraint, $\int \xi_j(t)\xi_k(t)\,dt = 0$ for all $k < i$. In practice, the optimization can be over a finite dimensional basis space, such as a b-spline basis, using standard constrained optimization functions. Alternatively, one could represent the full functional as by evaluating each sample at a finite vector of times then use standard PCA techniques.

In theory, one can veridically represent the full variation across the functional data by using as many bases function as there are samples. Normally, fPCA is used to extract just the dimensions on which there is the most variation, so only the first few bases are calculated. For example, all of the resilience functions from an experiment can be represented in the fPCA space as, $R_i(t) = \sum_j f_i^{(j)}\xi_j$ where $f_i^{(j)}$ is the factor score for the ith resilience function on the jth basis. To represent the resilience functions with a low dimensional (e.g., n dimensional) basis, one simply may use the first n principle functions,

$$R_i(t) \approx \sum_{j=1}^{n} f_i^{(j)}\xi_j.$$

Now each resilience function can be represented by the n-dimensional vector $f_i = \left(f_i^{(1)}, f_i^{(2)}, \ldots f_i^{(n)}\right)$. Note that, once this reduced dimensional vector space is used to represent the data, any rigid transformation of the space represents the data equally well, so it is common practice to choose a particular rotation, such as varimax, to represent the data for further analysis (cf. Ramsay & Silverman, 2005, Ch. 8).

Application to empirical data

Little et al. (2011, Experiment 1) measured RTs from four observers for each item in the categorization design shown in Fig. 2. The stimuli in this experiment were schematic lamps which varied in the width of the base (dimension 1) and the curvature of the top piece (dimension 2). The lamps also varied randomly on their design and lamp shade; however, these dimensions were not relevant for the task. Using visual analysis of the SIC (cf. Townsend & Nozawa, 1995) coupled with statistical tests of the mean RTs patterns and parametric modeling, Little et al. (2011) inferred that observers in this task processed the base and top of the lamps in a serial, self-terminating manner.

Little et al. (2013, Experiment 1) also measured RTs from four observers for each of the items in the design shown in

UNCORRECTED PROOF

Table 1 SIC & MIC statistics for Little, Nosofsky & Denton (2011; Exp. 1) and Little, Nosofsky, Donkin & Denton (2013, Exp. 1)

		Negative SIC test		Positive SIC test			MIC test		
Experiment	Observer	value	p	value	p		value	p	
LND2011 Exp 1	1	0.68	0.00	0.19	0.00	serial	−63.64	0.18	serial
	2	0.44	0.00	0.17	0.01	serial	−39.57	0.19	serial
	3	0.29	0.00	0.15	0.02	serial	28.80	0.84	serial
	4	0.45	0.00	0.21	0.00	serial	44.72	0.06	serial
LNDD2013 Exp 1	1	0.07	0.46	0.06	0.54	?	0.00	0.14	serial
	2	0.06	0.47	0.22	0.00	coactive	0.05	0.00	coactive
	3	0.08	0.37	0.09	0.23	?	0.01	0.05	coactive
	4	0.02	0.96	0.20	0.00	coactive	0.06	0.00	coactive

Note: LND2011 = Little, Nosofsky & Denton (2011); LNDD2013 = Little, Nosofsky, Donkin & Denton (2013)

Fig. 2. In this experiment, the stimuli were small Munsell color squares (hue 5R) varying in saturation and brightness. Using the same set of tools, the authors concluded that the best model of the RT data was a coactive processing architecture. The finding that the lamp dimensions were processed in a serial, self-terminating manner and that the brightness and saturation dimensions were processed in a coactive manner corresponds nicely to the long-standing distinction between separable and integral dimensions (Fific et al., 2008; Garner, 1974).

Because the OR category items in this task (with the exception of item AB, see Fig. 2) satisfy the decision rule for the OR category on one of the dimensions but satisfy the decision rule for the AND category on the other dimension, there is a conflict between the two dimensions. For example, for item AY_H, the curvature of the top piece is below the boundary on dimension 2, but the base of the lamp is wider than the value indicated by the boundary on dimension 1. Consequently, for this stimulus, the base provides strong evidence for the AND category, which, for this stimulus, is the incorrect response.

In each of these stimuli, two dimensions are always present, which precludes the use of the workload capacity measure. However, because the values of this incorrect dimension are varied in their discriminability (e.g., from AY_H to AY_L and from X_HB to X_LB), the resilience difference function and conflict contrast function can be used to provide further evidence about the processing architecture. Little et al. (2016) reported that the CCF(t) functions for each observer were negative indicating support for the serial, self-terminating model. Likewise, Little et al. (2016) reported that the CCF(t) functions for each observer were positive indicating coactivity. Here we apply the CC statistic developed above, along with the relevant SIC statistics (see Houpt & Townsend, 2010; Houpt et al., 2013), which have not been reported previously. We also applied the

UNCORRECTED PROOF

Table 2 CC statistics for Little, Nosofsky & Denton (2011; Exp. 1) and Little, Nosofsky, Donkin & Denton (2013, Exp. 1)

	CC test			
Experiment	Observer	value	p	Inference
LND2011 Exp 1	1	−5.96	0.00	serial ST/serial EX/parallel EX
	2	−7.20	0.00	serial ST/serial EX/parallel EX
	3	−3.65	0.00	serial ST/serial EX/parallel EX
	4	−1.33	0.18	parallel ST[a]
LNDD2013 Exp 1	1	11.19	0.00	coactive
	2	6.39	0.00	coactive
	3	9.59	0.00	coactive
	4	6.02	0.00	coactive

Note: LND2011 = Little, Nosofsky & Denton (2011); LNDD2013 = Little, Nosofsky, Donkin & Denton (2013)

[a]In the present case, although we do not reject the null hypothesis, the best inference in this case is the parallel ST model. While we cannot rule out the other models on the failure of this test, the inference can still be useful in conjunction with the results of tests of other aspects of the data (for instance, as in, Table 1)

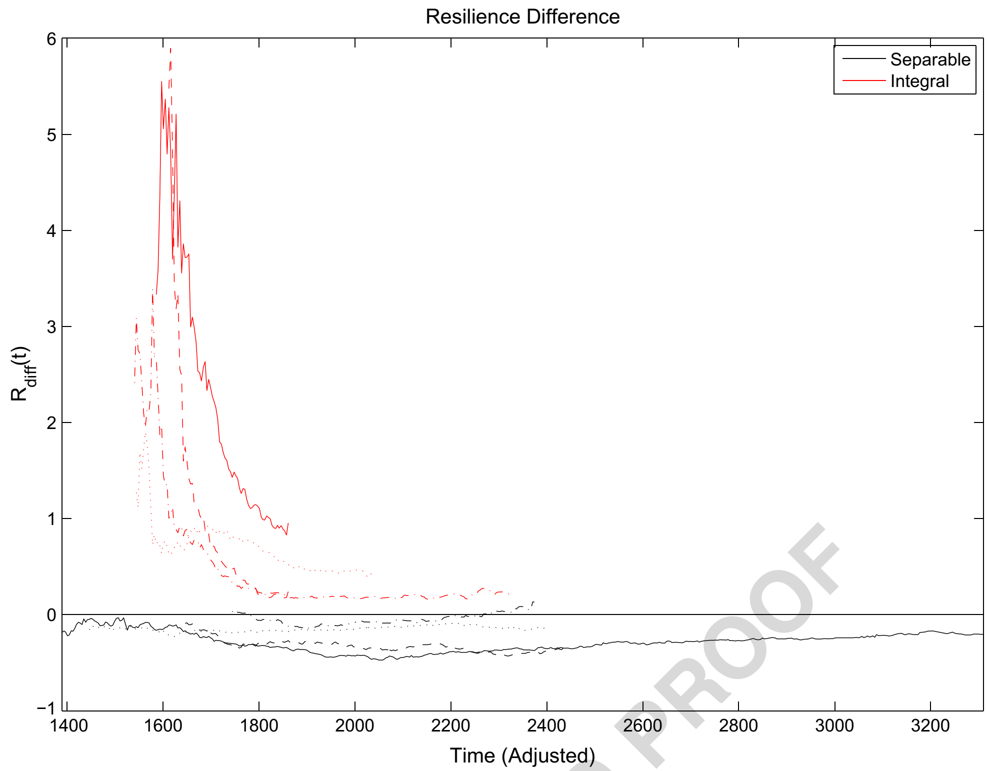

Fig. 4 Resilience difference functions for the data from Little, Nosofsky & Denton (2011; Experiment 1) and Little, Nosofsky, Donkin & Denton (2013; Experiment 1). Each *line* represents a different participant

UNCORRECTED PROOF

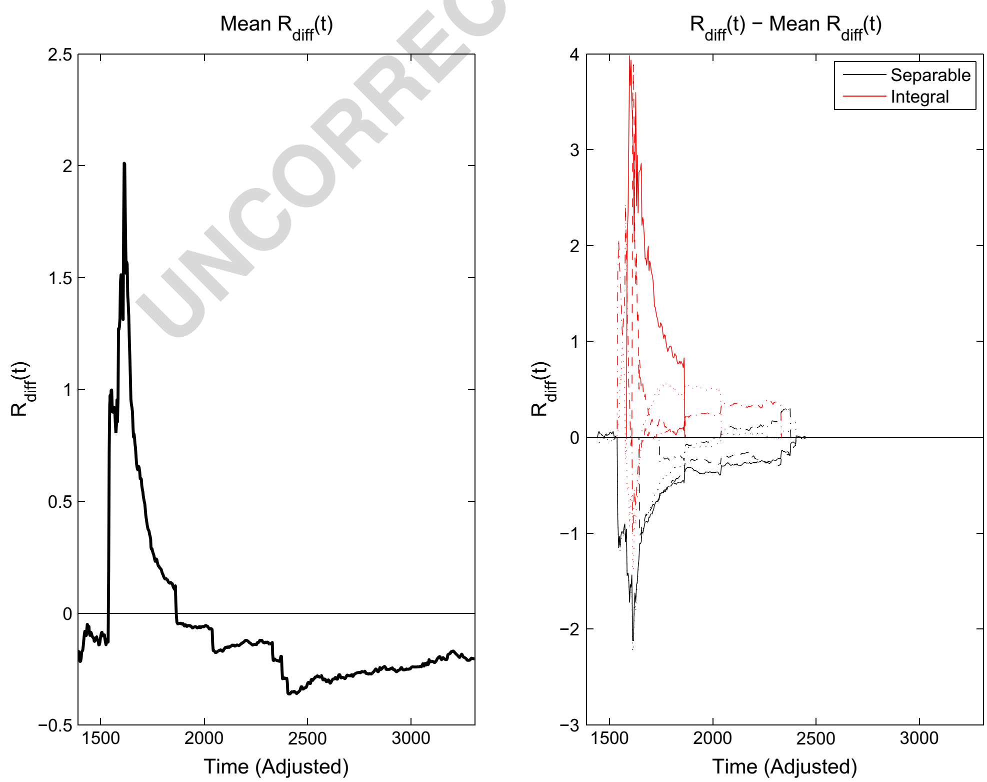

Fig. 5 Left panel: Mean resilience difference function averaged across participants and conditions. Right panel: Mean subtracted resilience difference functions for each participant

Kolmogorov–Smirnov test of stochastic dominance (Houpt et al., 2013) to test whether the AND category data meet the assumption of selective influence necessary for use of the SIC. Stochastic dominance was confirmed for all subjects.

Null-hypothesis tests

Table 1 shows the SIC statistics for Little et al. (2011) and Little et al. (2013). The results of the CC statistic are shown in Table 2. The statistical SIC tests largely agree with the conclusions reported in those papers. The SICs from the separable dimension case (e.g., lamps, Little et al., 2011) demonstrate significant negative and positive deflections from zero consistent with the predicted shape for a serial exhaustive SIC. (Note that the AND category used this tasks necessitates exhaustive processing even from a self-terminating system). The MIC tests for all four observers are not significantly different from zero. For the CC test, three of the observers demonstrate significantly negative CC statistics, indicating that the CCF(t) function is significantly less than 0. For one observer, we failed to reject the null hypothesis that the CCF(t) function was different from 0; although, the CC statistic was negative as expected. A significantly negative CCF(t) function is consistent with serial self-terminating, serial exhaustive, or parallel exhaustive processing. Taken together with the SIC results, the present analyses, to a large extent, agree with Little et al.'s (2011) conclusions of serial self-terminating processing.

For the integral dimensioned stimulus data, the SIC tests are more varied. In two cases, there is a significant positive deflection from zero, consistent with coactive processing. For one of the observers who does not show any significant deflections in the SIC, the MIC is significantly positive supporting an inference of coactivity. We failed to reject the null hypothesis for the remaining observer; though we note that the parametric modeling results favored an inference of coactivity for this observer as well (Little et al., 2013). For this experiment, the CC tests are all significantly positive supporting an inference of coactivity for all observers.

fPCA

We applied the fPCA Resilience Difference analysis to the $Rdiff(t)$ functions from Little et al. (2011) and Little et al. (2013) (see Fig. 4). Recall that in Little et al. (2011), the stimuli were comprised of separable dimensions but in Little et al. (2013), the stimuli were comprised of integral dimensions. As shown, the $Rdiff(t)$ functions are negative for the separable-dimensions data and positive

UNCORRECTED PROOF

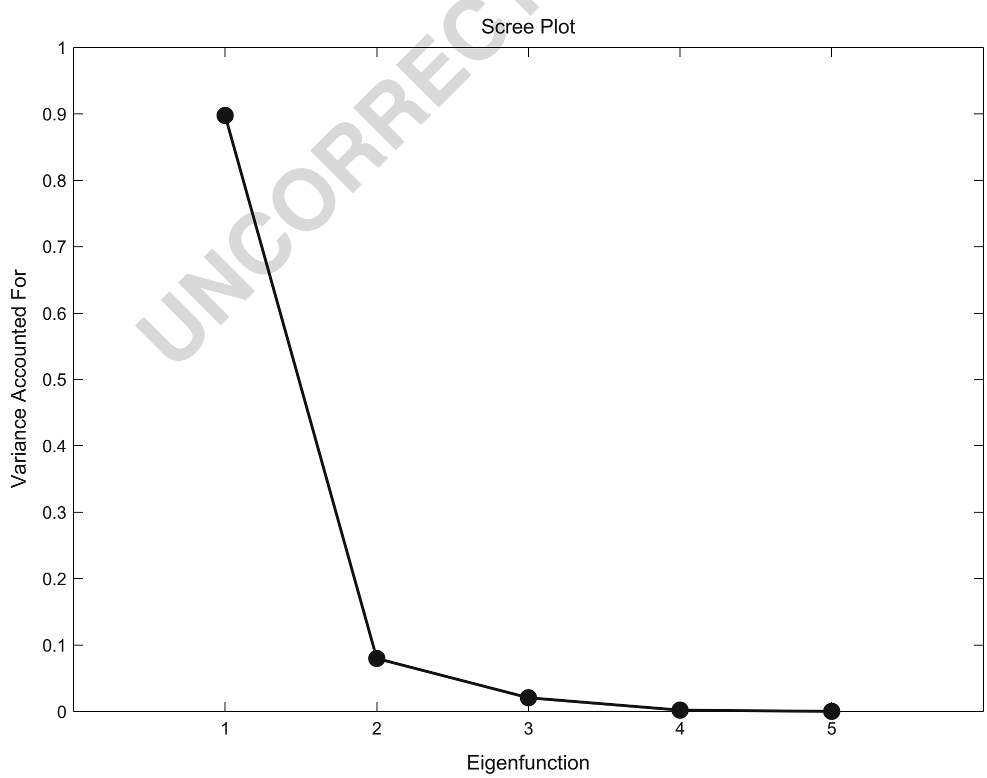

Fig. 6 Percentage of variance accounted by adding each eigenfunction up to 5. The first eigenfunction captures approximately 90 % of the variance across all of the resilience difference functions shown in the right panel of Fig. 5. The second eigenfunction adds approximately an additional 9 % and the rest of the eigenfunctions add only negligible amounts

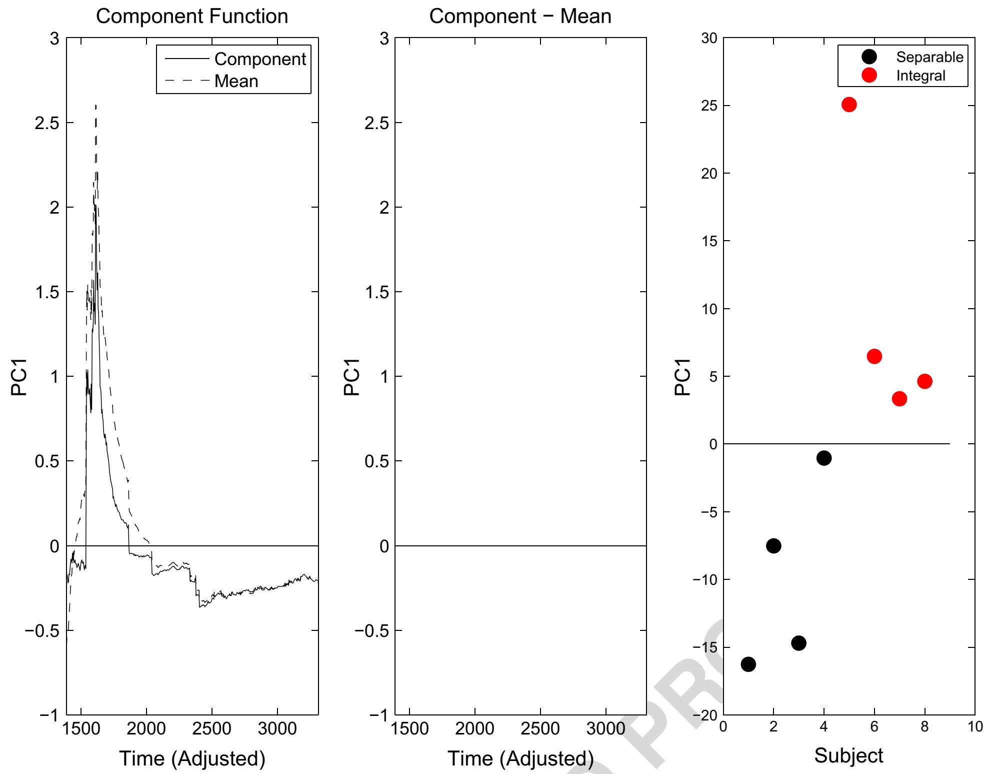

Fig. 7 Left panel: The first functional principle component weighted by the average magnitude of the factor score compared to the mean resilience difference function. Middle panel: The first functional principle component weighted by the average magnitude of the factor score after subtracting the resilience difference function. Right panel: Factor scores for each participant's resilience difference function in both experiments

UNCORRECTED PROOF

for the integral-dimensions data consistent with the inference of serial self-terminating and coactive processing, respectively.

Figure 5 shows the mean resilience difference function and the resilience difference functions after subtracting the mean function. As shown in Fig. 6, most of variation in the resilience difference functions was captured by the first functional principle component and only this component is selected for analysis. The first function principle component, weighted by the average magnitude of the factor score, is shown in Fig. 7 along with the mean function. This function increases at earlier times and then decreases at later times (for positive factor scores; the inverse is true for negative factor scores). The factor scores shown in the right panel of Fig. 7 nicely separate the observers who categorized separable dimensioned stimuli(with negative scores) and the observers who categorized the integral dimensioned stimuli.

Conclusions about data from resilience

The factor weights in from the fPCA provide a low dimensional representation of the resilience difference functions shown in Fig. 4, and consequently, allow a convenient analysis of differences between conditions and participants that does not require qualitative comparison between functions. The factor weights provide further support for the conclusion that integral dimensions are processed differently from separable dimensions. The key insight provided by the resilience difference function is that the integral dimensions are consistent with coactive processing whereas the separable dimensions are consistent with independent channel processing (i.e., serial and self-terminating although other architectures are possible candidates). Consequently, the analyses outlined here (see also Little et al., 2015, 2016) can be added to the growing set of methodological and theoretical analyses termed Systems Factorial Technology (Townsend and Nozawa, 1995).

Discussion

We have demonstrated a means for quantitatively analyzing resilience functions. The form of the resilience is quite similar to the capacity coefficient, and hence we were able to adapt the main tools for analyzing the capacity coefficient. However, despite the similarity in formulation, the resilience and resilience-difference functions are developed for a different set of inferences than the capacity function. We adapted the Houpt and Townsend (2012) null-hypothesis tests for inferences about whether the resilience functions are different from zero, a prediction of the parallel, first terminating model. Directional versions of the Houpt-Townsend test can additionally be used to reject either coactive or serial/parallel-exhaustive models. Following, Burns et al. (2013), we also demonstrated the use of fPCA for exploring differences among the *shapes* of resilience and resilience-difference functions.

Simulations indicated good statistical power of the null-hypothesis tests with reasonable numbers of simulated trials for both exponentially distributed times and response times generated from the LBA model (Brown and Heathcote, 2008). Similar to the findings reported in Houpt and Townsend (2012), we explored variations in the relative weighting across the range of response time and showed that there was not a strong effect on Type-I or Type-II error rates.

Using these new statistical approaches, we reexamined two datasets collected from experiments following the design in Fig. 2. Of the eight observers tested across the two datasets, seven had significantly non-zero CCFs using our new null-hypothesis test. The first experiment used stimuli made up of attributes that are traditionally classified as separable and hence our *a priori* assumption was that the best model would be either independent-serial or independent-parallel. Thus, we expected a negative CCF, which was observed for all observers and the null-hypothesis of a zero CCF (parallel, self-terminating) was rejected for three of the four observers. These findings were further corroborated using the SIC and MIC, other SFT measures of architecture and stopping-rule. The second experiment we analyzed used attributes considered to be integral. Hence, we expected the best model to be coactive, indicated by a positive CCF. This is indeed what we found: all observers had positive CCFs and the null hypothesis of zero CCF was rejected for each. Although the SIC and MIC were less decisive with the second dataset, coactive processing was indicated for three of the four observers.

Further analyses of these data using the fPCA approach indicated that the Resilience-Difference function shapes were distinctive between the integral and separable stimuli. The fPCA indicated that this distinction was most evident in the overall magnitude of the R_{diff} function for earlier response times.

Future directions

While the addition of these analyses is a major improvement over qualitative judgment of resilience analyses, there are potential further improvements. Perhaps most important to many users of resilience analysis is the ability to make both group and individual level inferences. The current suggested approach to aggregating across subjects is to first calculate each individuals resilience (or CCF) statistic, then perform standard null-hypothesis tests on those values. For example, to test whether the participants had a higher CCF with integral stimuli than with separable stimuli, we could have used a t-test on the CC statistics. A hierarchical analysis offers a more principled approach, in particular incorporating the uncertainty of the estimated CCF into tests about group differences. Houpt et al. (2016) recently proposed a hierarchical Bayesian model for estimating cumulative hazard functions and cumulative reverse hazard functions based on a piecewise-exponential model of response times. They have demonstrated success using the model for inferences regarding standard capacity coefficients, so the approach holds promise for resilience analysis as well.

Acknowledgments This work was supported by AFOSR Grant FA9550-13-1-0087 and ARC Discovery Project Grant DP120103120.

UNCORRECTED PROOF

References

Aalen, O.O., Borgan, Ã., & Gjessing, H.K. (2008). *Survival and event history analysis: a process point of view*. New York: Springer-Verlag.

Allen, S.W., & Brooks, L.R. (1991). Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General*, *120*, 3–19.

Andersen, P.K., Borgan, Gill, R.D., & Keiding, N. (1993). *Statistical models based on counting processes*. New York: Springer-Verlag.

Ashby, F.G., & Gott, R.E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory 709 and Cognition*, *14*, 33–53.

Blaha, L.M., & Townsend, J.T. (2014). On the capacity of single-target self-terminating processing. Manuscript under review.

Blunden, A.G., Wang, T., Griffiths, D., & Little, D.R. (2015). Logical-rules and the classification of integral dimensions: Arbitrary dimensions are not necessarily processed coactively. *Frontiers in Psychology*, *5*, 1531.

Brown, S.D., & Heathcote, A. (2008). The simplest complete model of choice response time. Linear ballistic accumulation. Cognitive Psychology.

Burns, D., Houpt, J.W., Townsend, J.T., & Endres, M.J. (2013). Functional principle components analysis of workload capacity functions. *Behavior Research Methods*, *45*, 1048–1057.

Chajut, E., Schupak, A., & Algom, D. (2009). Are spatial and dimensional attention separate? Evidence from Posner, Stroop, and Eriksen tasks. *Memory & Cognition*, *37*, 924–934.

Duncan, J., & Humphreys, G.W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*, 433–458.

Eidels, A., Houpt, J.W., Altieri, N., Pei, L., & Townsend, J.T. (2011). Nice guys finish fast and bad guys finish last: Facilitatory vs. inhibitory interaction in parallel systems. *Journal of Mathematical Psychology*, *55*, 176–190.

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

Fific, M., Nosofsky, R.M., & Townsend, J. (2008). Information-processing architectures in multidimensional classification: A validation of test of the systems factorial technology. *Journal of*

Experimental Psychology: Human Perception & Performance, *34*, 356–375.

Fific, M., Little, D.R., & Nosofsky, R. (2010). Logical-rule models of classification response times: A synthesis of mental-architecture, random-walk, and decision-bound 732 approaches. *Psychological Review*, *117*, 309–348.

Folstein, J.R., Van Petten, C., & Rose, S.A. (2008). Novelty and conflict in the categorization of complex stimuli. *Psychophysiology*, *45*, 467–479.

Garner, W.R. (1974). *The processing of information and structure*. New York: Wiley.

Houpt, J.W., & Townsend, J.T. (2010). The statistical properties of the survivor interaction contrast. *Journal of Mathematical Psychology*, *54*, 446–453.

Houpt, J.W., & Townsend, J.T. (2012). Statistical measures for workload capacity analysis. *Journal of Mathematical Psychology*, *56*, 341–355.

Houpt, J.W., Blaha, L.M., McIntire, J.P., Havig, P.R., & Townsend, J.T. (2013). Systems factorial technology with r. *Behavior Research Methods*, *46*, 307–330.

Houpt, J.W., MacEachern, S.N., Peruggia, M., Townsend, J.T., & Van Zandt, T. (2016). Semiparametric Bayesian approaches to systems factorial technology. Journal of Mathematical Psychology. Disponible sur. http://www.sciencedirect.com/science/article/pii/S0022249616000249.

Houpt, J.W., MacEachern, S.N., Peruggia, M., Townsend, J.T., & Zandt, T.V. (2016). Semiparametric approaches to Systems Factorial Technology. *Journal of Mathematical Psychology*. (in press).

Little, D.R., & Lewandowsky, S. (2012). Multiple-cue probability learning. In Seel, N. (Ed.) *Encyclopedia of the sciences of learning. New York: Springer*.

Little, D.R., Nosofsky, R., & Denton, S.E. (2011). Response time tests of logical rule-based models of categorization. *Journal of Experimental Psychology: Learning Memory & Cognition*, *37*, 1–27.

Little, D.R., Nosofsky, R.M., Donkin, C., & Denton, S.E. (2013). Logical-rules and the classification of integral dimensioned stimuli. *Journal of Experimental Psychology: Learning Memory & Cognition*, *39*, 801–820.

Little, D.R., Eidels, A., Fific, M., & Wang, T. (2015). Understanding the influence of distractors on workload capacity. *Journal of Mathematical Psychology*, *69*, 25–36.

Little, D.R., Eidels, A., Fific, M., & Wang, T. (2016). How do information processing systems deal with conflicting information? differential predictions for serial, parallel, and coactive processing models. Manuscript under revision.

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, *14*, 247–279.

Moneer, S., Wang, T., & Little, D.R. (in press). The processing architectures of whole-object features: A logical rules approach. *Journal of Experimental Psychology: Human Perception & Performance*. Q3

Nosofsky, R.M. (1991). Typicality in logical defined categories: Exemplar-similarity vs rule instantiation. *Memory & Cognition*, *19*, 131–150.

Nosofsky, R.M., & Little, D.R. (2010). Classification response times in probabilistic rule-based category structure: Contrasting exemplar-retrieval and decision-bound models. *Memory & Cognition*, *38*, 916–927.

Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, *33*, 1065–1076.

Proctor, R.W., & Vu, K.-P.L. (2006). *Stimulus-response compatibility principles: Data, theory and application*: CRC Press.

Ramsay, J.O., & Silverman, B.W. (2005). *Functional data analysis*. New York: Springer.

Ramsay, J.O., Hooker, G., & Graves, S. (2009). *Functional data analysis with R and MATLAB*. New York: Springer.

Rouder, J.N., & Morey, R.D. (2012). Default Bayes factors for model selection in regression. *Multivariate Behavioral Research*, *47*(6), 877–903.

Shalev, L., & Algom, D. (2000). Stroop and garner effects in and out of Posner's beam: Reconciling conceptions of selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 997–1017.

Simon, J.R., & Rudell, A.P. (1967). Auditory s-r compability: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology*, *51*, 300–304.

Stroop, J.R. (1935). Studies of interference in serial verbal reaction. *Journal of Experimental Psychology*, *18*, 643–662.

Townsend, J.T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial and coactive theories. *Journal of Mathematical Psychology*, *39*, 321–340.

Townsend, J.T., & Wenger, M.J. (2004). A theory of interactive parallel processing: New capacity measures and predictions for a response time inequality series. *Psychological Review*, *111*, 1003–1035.

Townsend, J.T., & Eidels, A. (2011). Workload capacity spaces: A unified methodology for response time measures of efficiency as workload is varied. *Psychonomic bulletin 797 & review*, *18*(4), 659–681.

Townsend, J.T., & Altieri, N. (2012). An accuracy–response time capacity assessment function that measures performance against standard parallel predictions. *Psychological review*, *119*(3), 500.

Wenger, M.J., & Townsend, J.T. (2000). Basic response time tools for studying general processing capacity in attention, perception, and cognition. *The Journal of General 806 Psychology*, *127*(1), 67–99.

UNCORRECTED PROOF