Journal of Mathematical Psychology xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

Review

A nonparametric technique for analysis of state-trace functions

Aaron S. Benjamin *, Michael L. Griffin, Jeffrey A. Douglas

University of Illinois at Urbana-Champaign, United States

HIGHLIGHTS

- State-trace analysis is an important and evolving technique in psychology.
- It adjudicates between single-variable and multivariable latent process theories.
- It relaxes assumptions about the measurement qualities of psychological data.
- PIRST is totally nonparametric, relying on isotonic regression and permutation tests.
- PIRST successfully recovers the latent structure of simulated data.
- PIRST also reveals limitations in the informativeness of potential data sets.

ARTICLE INFO

Article history:
Received 27 June 2018
Received in revised form 20 March 2019
Available online xxxx

ABSTRACT

State-trace analysis provides a direct and transparent way of evaluating a question that is central to many studies of cognitive function: do one or two latent processes underlie performance on a particular task? This evaluation is made using a state-trace plot, which is a bivariate plot of two dependent variables over a *dimensional* variable that provides the basis for the hypothesized dissociation, and a *trace* variable, which enables the examination over a range of levels of performance. State-trace analysis has been used successfully in research on perception (Mccarley & Grant, 2008), attention (Verhaeghen & Cerella, 2002), memory (Dunn, 2008), and categorization (Newell, Dunn, & Kalish, 2010). However, inferential techniques for evaluating whether a given state-trace plot yields evidence for one versus two latent processes have only recently started to appear in the literature. Here we develop PIRST (*Permuted Isotonic Regression for State-Trace*), a fully nonparametric algorithm based on isotonic regression that can be applied to a state-trace plot and used to characterize the amount of evidence in support of one hypothesis or the other. The technique is benchmarked using simulated data, and is shown to recover the true underlying latent structure under conditions of adequate measurement. Finally, we compare PIRST to an extant technique for state-trace analysis (CMR) using ROC analysis and evaluate their respective strengths and weaknesses for diagnosing latent structure.

Contents

A deep question underlying many theoretical and empirical endeavors in psychology is: does one latent process underlie performance or are multiple processes necessary? Yet, as posed, this question is poorly matched to the typical tools brought to bear on its adjudication. This article reviews the *state-trace* approach to the design of experiments and analysis of data and discusses how this approach is well suited to this commonly posed psychological question. We then introduce a novel technique for the evaluation of state-trace results and demonstrate its applicability to simulated data.

Consider the application of the state-trace procedure and analyses to the problem of understanding the origin of recognition memory judgments. In a recognition memory task, people are

* Corresponding author.
E-mail address: asbenjam@illinois.edu (A.S. Benjamin).

https://doi.org/10.1016/j.jmp.2019.03.006
0022-2496/

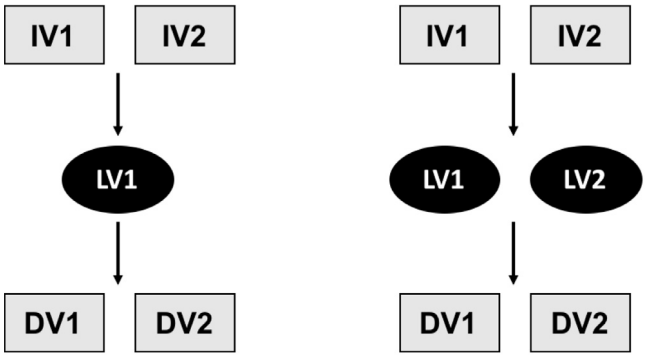

Fig. 1. Competing latent-variable theories within a state-trace analysis. Left panel: Single-process theories have only a single latent variable through which the independent variables in an experiment exert their effect on dependent variables. Right panel: Multi-process theories have more than one latent variable in this role.

exposed to a set of stimuli during a learning phase and on a later test are asked to select those previously studied stimuli out of a larger set. According to the *dual-process* view of recognition (e.g., Jacoby, 1991; Mandler, 1980; Wixted, 2007), the decision to select a stimulus as previously studied reflects some combination of *familiarity*—a sense of "pastness" without specific information about the context of its occurrence – and *retrieval* – a memory for the specific episode during which the stimulus was studied. Theories differ in how those two processes are assessed, as well as how they are combined to form a single judgment, but those differences are not relevant to the questions pursued here. Evidence in support of dual-process theories typically consists of a finding in which two different dependent variables respond differentially to an independent variable. For example, a theorist might conclude that recognition memory involves multiple processes because dividing attention during study reduces scores more on one type of memory test than on another. Dual-process theories are contrasted with *single-process* theories, in which only a single variable contributes to the recognition judgment (see also Newell, Dunn, and Kalish (2010), for an example in the related domain of categorization). These theories also take a variety of forms, but at heart they share the assumption that multiple dependent variables all reflect a common latent variable. The contrasting positions are depicted in Fig. 1 as competing latent-variable theories.

The psychometric tools that would be needed to rigorously compare models of these types are rarely used, and the data collected in typical recognition memory experiments would typically be insufficient for such models. The most common approach is to shoehorn the question into a linear model and ask whether interactions are present between two independent variables of interest. So, one might conclude that recognition memory involves multiple processes because, as in our earlier example, a manipulation of divided attention during study reduces scores more on one type of test than another.

The problems with this approach are, somehow, both well known and yet not widely recognized. Every experimentalist worries about the interpretative difficulties associated with "floor" and "ceiling" effects. That unease reflects an appreciation for the deep hypocrisy associated with applying linear statistical models to probability-scale data. Yet many experiments have at their core an interpretation of an interaction that relies on an interval-level interpretation of the measurement scale. That is, it must be believed that an increase of 10 units is the same, regardless of whether that increase is from 0 to 10, 45 to 55, or 90 to 100. That belief is transparently inconsistent with concerns over measurement floors and ceilings.

This state of affairs has been well known within experimental psychology at least since 1978, when Geoffrey Loftus published a paper pointing out that an explicit theoretical mapping between measured variables and theoretical (latent) variables is a necessary step in interpreting interactions (Loftus, 1978). A recent review of the problem suggests that the lessons of this important article have not spread widely throughout the field (Wagenmakers, Krypotos, Criss, & Iverson, 2012). Interactions that can appear or disappear depending on the nature of this mapping are sometimes called *removable* interactions, for reasons that should now be apparent.

Part of the problem is that researchers so desperately want to draw conclusions about dissociations: that some manipulations affects *this* more than *that*. Developing a fully fledged theory that maps between measured and latent variables requires substantial theoretical development and expertise in applying such models to human data, as Wagenmakers et al. (2012, Figure 5) demonstrated with the diffusion model of Ratcliff (1978). Such an application may stretch the willingness or ken of many experimental psychologists. Others are even more pessimistic, pointing out that the reparameterization of manifest variables into potentially linear latent ones is only allowed under restrictive assumptions about the relationship between measurement scales and statistics (Michell, 1986).

Thankfully, there exists an alternative approach, though it is not well known and not widely used. In the next section, we outline the logic of state-trace design and analysis of experiments. State-trace logic can be seen as a case of order-constrained inference (e.g., Barlow, Bartholomew, Bremner, & Brunk, 1972; Iverson, 2006; Regenwetter & Cavagnaro, 2018), applied to the common but specific problem of inferring one versus multiple processes. In this paper, we base our discussion and procedure around designs in which the variables are all manipulated within-subjects, as is common in the cognitive paradigms in which the state-trace procedure is currently being used. However, the logic of the test allows broad application to a variety of experimental designs.

1. The state-trace approach

The centerpiece of the state-trace approach is the state-trace plot, which is a bivariate map of two measures in an experimental design. Fig. 2 shows three possible state-trace plots. In the left panel, typical data from a 2×2 design are shown. One of the factors in the design is represented as the two axes of the graph, and the other as the two points on the plot. If the slope of the line connecting these two points is not precisely 1, then there is evidence for an interaction. (Whether that interaction is deemed "significant" or not depends, of course, on the sampling error associated with estimation.)

In state-trace analysis, one assumes that the data are interpretable at an ordinal and not an interval level. This relaxation means that the data are no longer interpretable in a linear statistical framework. In the new framework, the exact slope of the line connecting our two points is immaterial, since we have no way of inferring that an increase on one dimension is larger or smaller than an increase on the other dimension. In other words, there is no way of validly interpreting an "interaction" in a 2×2 design under ordinal assumptions unless it is a cross-over (disordinal) interaction (Loftus, 1978). So the left panel of Fig. 2 is uninformative with respect to the question of whether the two measures derive from a common latent variable or not.

Instead of relying on a theoretically laden interpretation of the measurement scale, a state-trace experiment includes a second, *trace*, variable, which is used to empirically map out the contour

Please cite this article as: A.S. Benjamin, M.L. Griffin and J.A. Douglas, A nonparametric technique for analysis of state-trace functions. Journal of Mathematical Psychology (2019), https://doi.org/10.1016/j.jmp.2019.03.006.

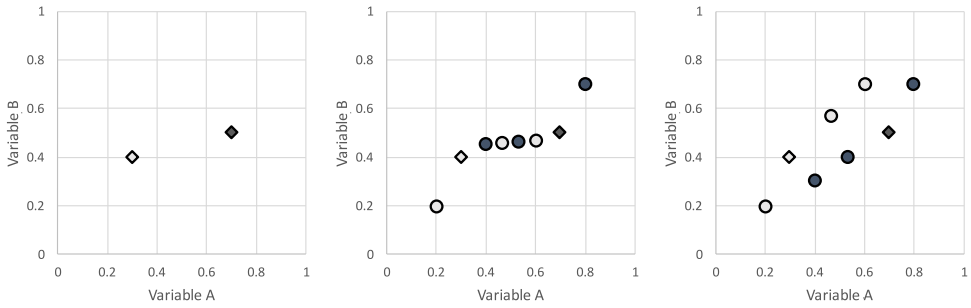

Fig. 2. Examples of theoretical state-trace plots. The diamonds are the same data points across the three plots. Left panel: a typical 2 × 2 experiment for which state-trace analysis cannot be applied. Middle panel: A fleshed-out state-trace plot indicating the action of a single latent process. Right panel: A fleshed-out state-trace plot indicating the action of multiple latent processes.

of the two conditions over a range of different performance levels. This empirical contour replaces assumptions about the measurement scale, and can be used in a straightforward way to determine whether the two original data points lie on the same function or not. In the middle panel of Fig. 2, it can be seen that, upon sweeping out the range of performance in each of the two conditions represented in the left graph, there is a single monotone function that relates the two variables to one another, and that all the condition variable does is reveal one versus another portion of that common function. In contrast, in the right graph, it can be seen that the additional data reveal that a common monotone function cannot connect all of the data points. It appears as though one condition lies on a totally separate function than the other. Such data would be convincing evidence for the influence of multiple latent variables.

There have been numerous treatments of state-trace theory, ranging from the highly technical (e.g., Bamber, 1979; Dunn & James, 2003) to introductory ones geared for experimental psychologists (Loftus, Oberg, & Dillon, 2004; Newell & Dunn, 2008). A thorough overview is provided in Dunn and Kalish (2018). The lynchpin of the relationship between a state-trace plot and a single latent variable formulation like the one shown in the right portion of Fig. 1 is the fact that *any combination of circumstances that leads performance to be higher in one condition than another for one dependent variable must also do so for the other dependent variable*. In the limit, what this means is that a plot of DV1 versus DV2, over the trace variable, must be monotonic. In practice, the functional constraint is weak monotonicity, allowing patterns in which performance rises in one condition and does not fall in the other. If, as in the right panel of Fig. 2, only a nonmonotonic function can connect all of the points on the plot, then this circumstance is violated and evidence for multiple latent variables is present. For a more detailed discussion of this particular motivating logic, the reader is referred to Loftus et al. (2004).

Analysis of state-trace functions. Of course, things would be simple if data always revealed themselves so straightforwardly as to allow visual detection of the pattern in a state-trace function. Optimistically, Loftus et al. (2004) proposed that one assess the rank-order correlation among the full set of data and reject the single-LV model if that value is less than 1. In the presence of measurement noise, this is obviously too stringent a criterion. And the traditional tool in the psychologist's toolbox for reducing noise – averaging over subjects – poses an additional complication: averaging does not necessarily preserve monotonicity (Prince, Brown, & Heathcote, 2012).[1]

[1] It is worth remembering that averaging over items within a subject has the same potential to introduce artifacts (Estes, 1956; Sidman, 1952). Though it is certainly an important point for future development, we do not pursue it further here.

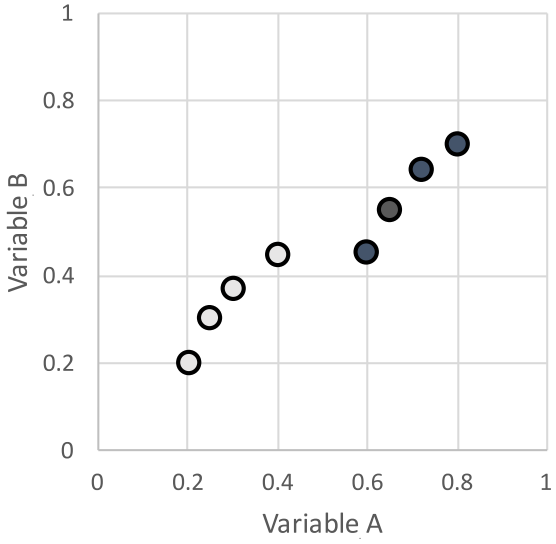

Fig. 3. A state-trace plot that is nondiagnostic with respect to latent structure.

A unique difficulty in assessing state-trace functions is that sometimes a monotone function is possible under completely uninformative conditions. A monotonic function can be drawn between all the points of Fig. 3, but it is clear from visual inspection that the separation of the conditions on both dimensions from one another renders it impossible to evaluate whether a single or multiple functions underlie this particular set of points. This pattern illustrates how important it is for the points among a state-trace plot to overlap on both dimensions. This peculiar aspect of state-trace plots – that the data can render themselves completely unrevealing – is something that we keep in mind in our approach presented here.

One recent approach involves the parametric evaluation of binomial parameters for each datum within the state-trace plot and the comparison of different ordering models upon those parameters (Prince et al., 2012). An important advance in this technique is the inclusion within the suite of models under examination a *no-overlap* model that indicates nondiagnosticity of the plot. However, the model awaits generalization to data that are not suited to a binomial model (such as subjective ratings, saccade amplitudes, or response times; cf. Mccarley & Grant, 2008; Verhaeghen & Cerella, 2002).

One approach that holds much promise is isotonic regression. An isotonic regression line is a monotonic function that, like a linear regression, minimizes the squared distance between the data and the fitted function. It has been suggested for use in state-trace analysis by Newell and Dunn (2008), has been used in

Please cite this article as: A.S. Benjamin, M.L. Griffin and J.A. Douglas, A nonparametric technique for analysis of state-trace functions. Journal of Mathematical Psychology (2019), https://doi.org/10.1016/j.jmp.2019.03.006.

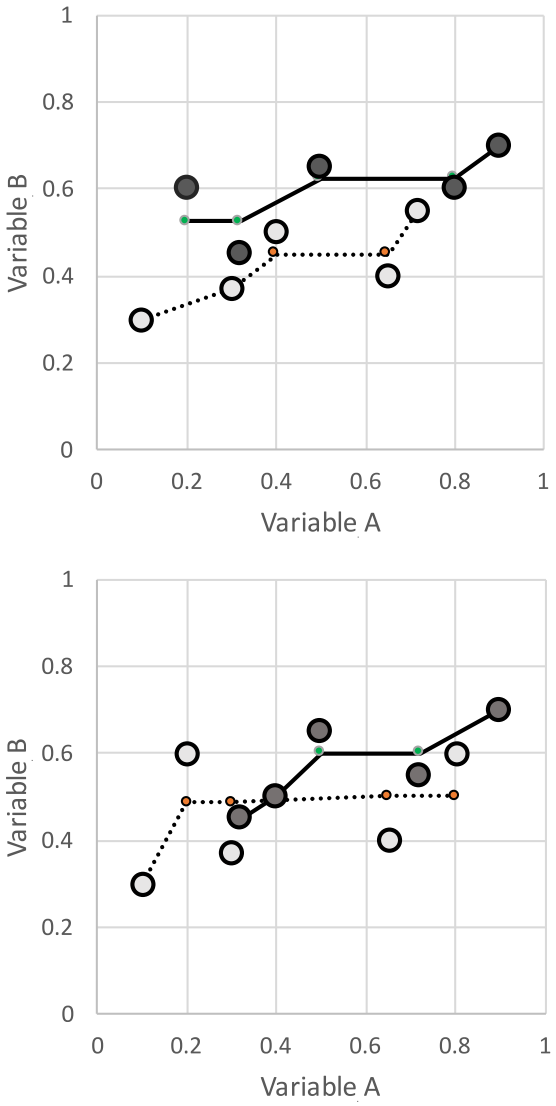

Fig. 4. Isotonic regression applied to state-trace plots. Data are shown as large circles; regression lines as small circles connected by lines. Top panel: isotonic regression applied to the original data. Bottom panel: Isotonic regression applied to one specific permutation of the data.

practice by Kalish, Dunn, Burdakov, and Sysoev (2016) and was developed independently here.

Kalish et al. (2016) compared a nested set of models using isotonic regression. This is also the starting point for the technique we introduce here. Our technique avoids the need for parametric formulations for different dependent variables by pursuing a purely nonparametric means of testing. Since the core theoretical position of state-trace analysis is a relaxation of the traditional parametric assumptions for distributions of scores, a nonparametric means of inferentially evaluating state-trace data provides a high degree of compatibility. We call our technique *Permuted Isotonic Regression for State-Trace* (PIRST).

PIRST combines isotonic regression with a nonparametric approach to permuting points between experimental conditions. The test statistic comes from a comparison of the error of the fit of the regression model to the original data to a sampling distribution of error from the permuted data sets. The permutation algorithm critically relies on the following idea: *If the underlying data-generating function is truly monotone, then the shuffling of condition labels should not matter: all the condition variable does is alter what portion of the function we are sampling from.* Consequently, when the fit quality from the original data set lies near the middle of the sampling distribution, then the preponderance of evidence supports the conclusion that a single function – and thus only a single latent variable – is warranted.[2] In the absence of noise, any permutation of the condition labels will lead to the same fit error as the original (true) assignment of labels.

In contrast, if the data derive from two separate latent functions, then random permutation of the condition labels will reduce the quality of the fit, and the error from the fit to the original data will be lower than the bulk of the sampling distribution. In fact, in the absence of noise, any permutation of condition labels will by definition lead to greater fit error than the original assignment.

The test statistic is thus the proportion of permutations that lead to greater error in the fit of isotonic regression than the original fit to the true data. This value ranges meaningfully from 0.5 to 1.0, but noise can perturb it below 0.5. Values near 0.5 indicate a preponderance of evidence in favor of a single process; values near 1.0 indicate that the evidence favors multiple processes.

The steps of this approach are outlined in more detail in the following section.

We also directly compare this technique with the *coupled monotone regression* (CMR) test recently proposed by Kalish et al. (2016) (see also Dunn & Kalish, 2018) for state-trace analysis. Isotonic regression is also central to their approach, though their technique has the appealing quality of minimizing error on both x- and y-dimensions of the state-trace plot when fitting the regression equation.

In CMR, the multi-process model fits isotonic regressions to the two *state* conditions individually, conditional upon any prespecified partial order constraints. The second model includes an ordering constraint on the variables: the order of one must match the order for the other. This latter model corresponds to the optimal fit of the CMR algorithm, minimizing error of an isotonic regression in both dimensions. The first model always leads to a superior (or equally good) fit, and the difference between the two is used to gauge the respective evidence for the two competing hypotheses.

CMR requires the use of aggregated data because it uses bootstrapping to compare single-process and multiple-process models within the dataset, and uses that resampling procedure to supply a traditional *p*-value corresponding to a test of the null hypothesis of a single latent order. Of course, those data may be aggregated over a group, or over trials within individual subjects.

2. PIRST: A Permutation test for isotonic regression on state-trace functions

The top panel of Fig. 4 shows a state-trace function *with each condition* separately fit by isotonic regression. The regressions are indicated by the small colored points connected by lines. Because the actual data points in each condition cannot be directly connected and still preserve monotonicity, some of the points of the fitted function deviate from the actual data. Like linear regression, isotonic regression minimizes squared error (unweighted, in this case, and throughout this paper). Unlike linear regression, it is constrained only to weak monotonicity, not to linearity. The combined squared deviations of the fit of these two functions yields one component of our test statistic—the original SSE.

Rather than permuting the values of the data points to generate a sampling distribution, our technique permutes the

[2] Strictly speaking, multiple latent variables could yield this same outcome, though only under restricted conditions (Dunn & Kalish, 2018). Parsimony compels a single-process interpretation in these circumstances.

Please cite this article as: A.S. Benjamin, M.L. Griffin and J.A. Douglas, A nonparametric technique for analysis of state-trace functions. Journal of Mathematical Psychology (2019), https://doi.org/10.1016/j.jmp.2019.03.006.

assignment of these points to the condition variable. In doing so, it avoids giving precedence to one variable over another (which would be necessary if we were to choose to permute along either the x- or the y-axis of the graph).

There are two steps in the permutation process. The first is to select the points eligible for permutation. A point is eligible if it is flanked, on both sides of both the x- and the y-axis, by one or more points from the opposite condition. It is ineligible if it is either smaller or larger than every point from the opposite condition. That is, if there are no data points from the other condition with smaller (or larger) x-values *or* y-values then the data point is eligible for permutation. Pre-selection for eligibility helps the permutation algorithm by disallowing permutations that do not affect the outcome. However, this step can be skipped under conditions in which implementation is unwieldy, or if the permutation technique is generalized in the future to experiments in which there are more than two state conditions.

The fit of an isotonic regression line does not change if the condition label of these ineligible points changes, so their status is irrelevant to our test. This is a corollary to the problem reviewed earlier about how a state-trace plot can be totally undiagnostic if the data from the two conditions do not overlap with each other on both dimensions. Likewise, this permutation technique only induces meaningful variability across permutations when some of the points in the state-trace plot lie within a region of overlap. We will explicitly consider the cost of nonoverlapping data in constructing simulated state-trace plots in the next section of this paper.

The second step randomly permutes the condition label across all eligible points, preserving the number of data points from each condition. What this means is that points stay in the same positions on the state-trace plot across all permutations, but that some of them will be randomly reassigned from one condition to another within each permutation. The way in which this is done is by assigning all of the condition labels (i.e., 'Condition 1' and 'Condition 2') from the eligible subset of data points back to that set of points randomly. The permutation is conducted again if the outcome of the permutation is the same assignment as the original data. After a permutation, the isotonic regressions are refit and the error is measured again. An example permutation and recomputation of the regressions is shown in the bottom panel of Fig. 4. The sum of squared error (SSE) for each of many permutations is computed. In the example case developed here, Permutations are done at the level of individual subjects, and generated randomly for each permutation iteration and for each subject.

The final step is to assess a (one-sided) test statistic by computing the proportion of permuted data sets for which SSE is greater than the original SSE. When this value is high, it suggests that the true (original) arrangement of the data provides a more convincing fit to two monotone functions than does a random arrangement of the data, all while preserving the actual values of the data points themselves. When the true state of affairs is a single function, the permutation of the data will not affect the quality of the fit in the long-run. This is because the condition variable does not affect the proximity of the data points to the latent single function—only measurement noise does. Consequently, the original error will lie near the mode of the distribution of SSE from the permuted data and yield a mean of 0.5.[3] We do not develop in this paper a specific decision rule, in part because that decision rule requires an explicit consideration of the relative costs of false positives (concluding the action of multiple processes when only a single latent one is truly at work) versus misses (failing to appropriately conclude the action of multiple processes). We assess the validity of our technique and CMR by using ROC functions, which reveal each technique's ability to discern single- from multiple-process data over all possible decision rules.

The starting point of fitting two functions—inherently the more complex and more flexible model—might seem an indirect approach to estimating whether one or two functions provides a better fit to a set of data. We originally attempted a technique that started with a single function fit to the entirety of the data and evaluated the quality of that fit. However, that starting point prevents the application of a sensible permutation algorithm. Scrambling the condition labels by definition does not change the fit of a single function to the data set. It was possible to preserve the condition labels but to permute the values on either the x- or the y-axis, but this would force a choice as to which variable underlay the error term. Given the equivalent status of both Variable A and Variable B in state-trace analysis, this choice was undesirable and we abandoned the approach.[4]

Indeed, one weakness of PIRST is the fact that it forces the user to make a decision about how to assign variables in the original regression. There is no conceptual difference between the x- and y-variables in a state-trace design; both contain measurement error and, often, neither is thought to be logically a predictor of the other. Future developments might marry regression that minimizes bivariate error, like CMR, with type of nonparametric test we develop here. For present purposes, it is important for users of PIRST to identify ahead of time how the conditions should be assigned to the x- and y- dimensions of the state-trace plot, in order to reduce opportunities for capitalizing upon inflated Type I error that would result from trying both possible assignments. Alternatively, both assignments could be tested and reported.

Other experimental designs and models of sampling error. The idea of permuting condition labels can be applied to any set of data. In fact, the technique as developed here can be directly applied to group data rather than to individuals. However, in mixed- or between-subject designs, attention must be paid to preserving the correlational structure across the points in the state-trace plot imposed by the design. We do not pursue these ideas further in this article.

Similarly, the logic of this test may be applicable to designs with more than two conditions. We develop it here exclusively for the 2 (condition) $\times$ 2 (outcome) $\times$ k (trace factor) case but include some pointers along the way that may be beneficial for future development.

A resampling approach in combination with cross-validation could also be used to evaluate the robustness of a particular conclusion to deviations due to sampling error. We leave this development for future work as well.

3. Simulation analyses

The process of generating simulated state-trace plots involved multiple steps, each of which will be described in detail below.

[3] The sampling distribution is most accurate when ties among SSE scores are few; in fact, when there are relatively few data points (as in some of the examples in the figures here), ties can be frequent. Ties do not count towards evidence for multiple processes, as can be seen by the decision rule (in which only permuted data sets in which SSE is *higher* are counted as evidence towards multiple processes).

[4] In a third approach, we counted the number of points within the entire set of data that forced a violation of monotonicity. We compared that value to a similar score, corrected for different numbers of opportunities, computed for the two individual conditions. If we assume that the true function *for each individual condition* is monotone under both latent-variable theories, then deviations from monotonicity in those functions serve as a measure of noise in our measurement, and can be used to assess whether the rate evident in the single function was greater than expected conditional upon our estimate of noise. However, the technique reported in this paper outperformed that approach consistently, and so we do not report the details of that approach here.

Please cite this article as: A.S. Benjamin, M.L. Griffin and J.A. Douglas, A nonparametric technique for analysis of state-trace functions. Journal of Mathematical Psychology (2019), https://doi.org/10.1016/j.jmp.2019.03.006.

Boldfaced text within this description indicates variables that are included in the analysis using the algorithm introduced previously. Code for running these simulations is publicly available at https://osf.io/y6c5r/, as are the data yielded by the specific simulations conducted here and the outcome of the inferential algorithms applied to the simulated data.

(1) We start each state-trace plot with a single latent linear function from which a **variable number of points**, corresponding to the *trace* variable, are sampled. This corresponds to the single-variable model discussed earlier and shown in Fig. 1. These data are spread evenly throughout the range (with some small margins) of the state-trace space.

(2) Data for the individual conditions of the *state* variable are generated by adjusting the points generated in step (1). The conditions are varied to include a variable amount of **overlap**: the points from the two conditions either overlap considerably (∼90%) or overlap much less (∼60%).

(3) The performance of the algorithm is evaluated as a function of the **linearity of the underlying function**. In the nonlinear case, the data from step (2) are projected onto a concave-up or concave-down function.

(4) Critically, our algorithm must discriminate between data that come from a single latent function versus two latent functions. This is manipulated in our simulations through a **latent interaction** variable that adds a constant value to the y-values from one condition. When this interaction is set to 0, the data come from a single function. The higher the value of the interaction term, the more prominent the deviation from a single function.

(5) In the final step, random Gaussian **noise** is added to all of the data points. The noise is manipulated to evaluate the performance of our algorithm across different levels of measurement noise.

Generation of data points. The starting point for generating data is a single linear latent variable *S*, from which we generate *N*2* points, corresponding to the number of measures across the *trace* variable for each of two *state* conditions. In our simulations, *N* is either 4, 10, or 50. The distribution of *N* latent points is governed by the following equation, which spreads the *i* values across the scale without placing them close to the margins (for reasons that will soon become evident):

$$S_i = 0.7\frac{i}{N} + 0.15$$

The *N* data points from each state condition are then generated by adding or subtracting a constant to these latent strength values, which spreads the *N*2* points evenly: $A_{i,C} = S_i \pm 0.35\left(\frac{p-0.5}{N}\right)$, where *C* indicates the state condition label. The parameter *p* determines the overlap between conditions, and is discussed below.

Overlap. The overlap between the data points from the two state conditions is critical to assessing the performance of any analytic technique for state-trace functions. As reviewed previously and demonstrated in Fig. 3, some state-trace plots are totally nondiagnostic because of zero overlap.

In one condition, designed to simulate *high overlap* of data points across conditions, *p* was set to 1 when there were 4 points per condition, to 1 when there were 10 points per condition, and to 4 when there were 50 points per condition. The parameter *p* determines the number of points from each condition that will lie outside the overlap region when there is no noise.

In a second condition, designed to simulate *low overlap* of data points, *p* was set to 2 when there were 4 items per condition, to 4 when there were 10 items per condition, and to 18 when there were 50 points per condition. This set of circumstances provides a more challenging test for our algorithm. The actual values of the data are determined by subtracting a constant (0.5) from *p*, as shown in the equation for $A_{i,C}$. This correction ensures that the points from the two conditions are offset from one another.

Linear and nonlinear functions. For some simulations, a latent linear function provides the basis for the state-trace plot. In that case, the x and y values of the data points are set directly to the values of *A* at this point.

For other simulations, the data were transformed into a nonlinear function by projecting them onto an arc from a circle. To ensure that *A* lay within the range of the circle corresponding to the appropriate concavity, *A* is remapped to the portion of the circle from 270° to 360°:

$$A*_{i,C} = 90A + 270$$

and corresponding functions are generated that are either concave-up:

$$y = \sin\left(A*_{i,C}\right) + 1$$
$$x = \cos(A*_{i,C})$$

or concave-down:

$$y = \cos\left(A*_{i,C}\right)$$
$$x = \sin\left(A*_{i,C}\right) + 1$$

both of which are shown in Fig. 5.

Interaction constant. To this point, the generated data are all derived from a single-process model. To simulate the effects of a true two-process model, an **interaction** constant was added to the y values for one condition. The value of this constant was either 0 (single-process model), 0.1, 0.15, or 0.2, corresponding to different effect sizes.

Noise. Finally, all of the data points within the plot were perturbed by Gaussian noise with a mean of 0. The standard deviation of the imposed noise was either 0.1, 0.2, or 0.4, corresponding to different levels of measurement noise a data set might contain.

Simulation parameters. Within each simulation, there were 4, 10, or 50 data points generated per condition. For each simulated experiment, the size of the effect and number of data points was held constant, and 100 simulated subjects were generated. Fig. 5 shows example simulated subjects for the 4 data-point condition, across all of the other relevant dimensions. There are four quadrants to this figure, corresponding to the four combinations of latent function shape (linear or nonlinear) and separation between the conditions (large or small). Within each quadrant, the columns indicate increasing effect size of the latent interaction, with the left graphs showing the action of a single latent process, the middle graphs showing dual processes with a small effect size, and the right graphs showing dual processes with a large effect size. The rows within each quadrant in and the rows indicate the contribution of noise, with the top being no noise (indicated for illustration only; this condition is not included in our simulations), and the bottom including a small amount of noise ($\sigma = 0.1$).

Analysis using PIRST. The data from each simulated subject was analyzed by the permutation-regression technique detailed previously. The outcome of this analysis is shown in Fig. 6, which shows a heatmap of the proportion of cases in which the original SSE exceeded the SSE from the permuted data sets, for each effect size, overlap condition, and number of points in the state-trace plot. Recall that this value will be high when the multiprocess explanation provides a superior account of the data.

Please cite this article as: A.S. Benjamin, M.L. Griffin and J.A. Douglas, A nonparametric technique for analysis of state-trace functions. Journal of Mathematical Psychology (2019), https://doi.org/10.1016/j.jmp.2019.03.006.

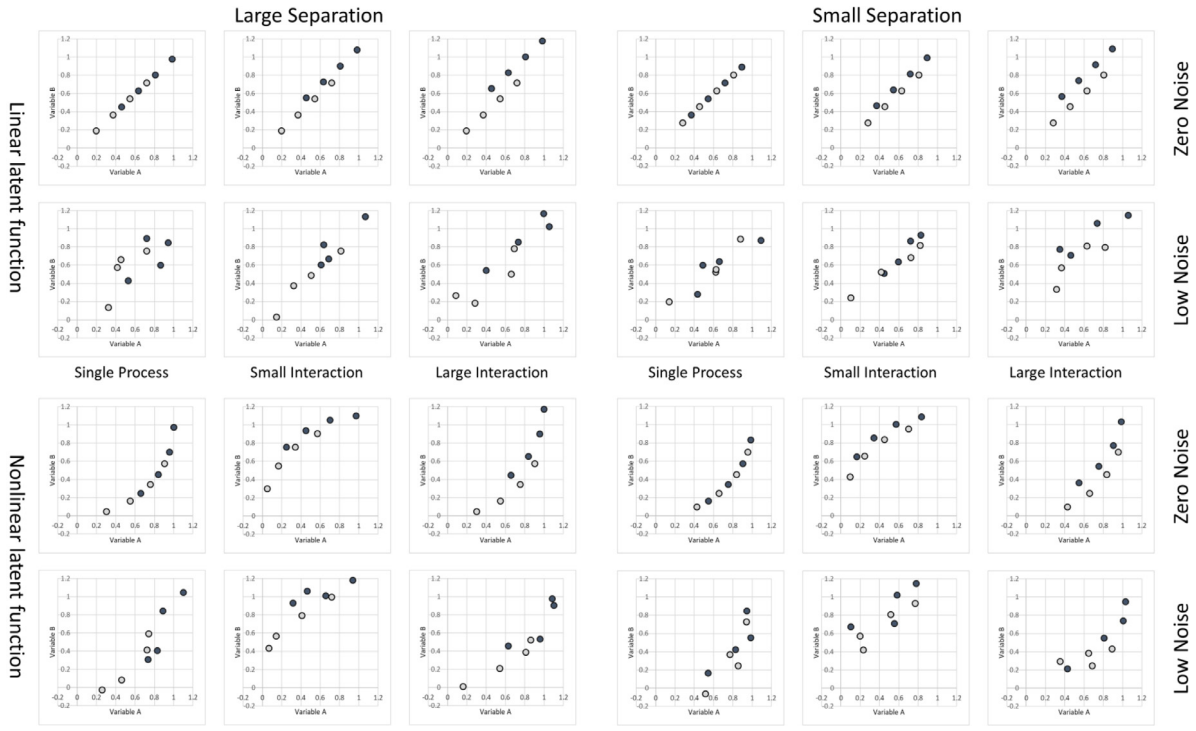

Fig. 5. Examples of simulated individual-subject state-trace plots varying in shape of the underlying latent function, noise, separation between conditions, and size of the latent interaction. See text for a more detailed description of the conditions.

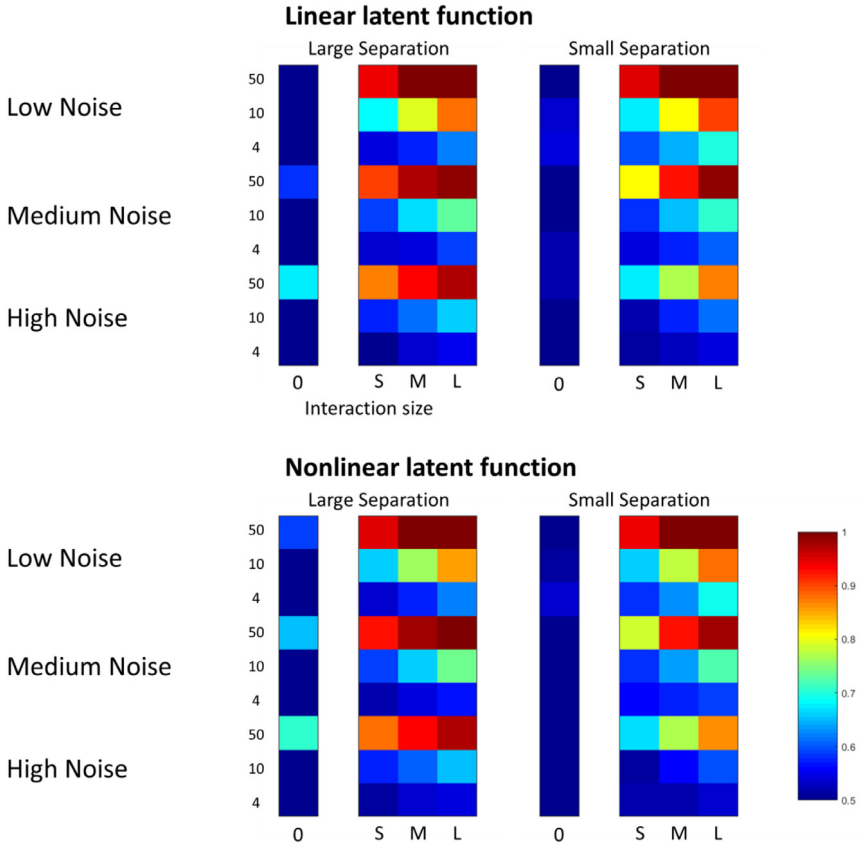

Fig. 6. Heatmaps of the rate at which PIRST yields greater evidence against the single process model. Hotter colors indicate more evidence against the single process model. The results of the simulations are described in detail in the text. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Please cite this article as: A.S. Benjamin, M.L. Griffin and J.A. Douglas, A nonparametric technique for analysis of state-trace functions. Journal of Mathematical Psychology (2019), https://doi.org/10.1016/j.jmp.2019.03.006.

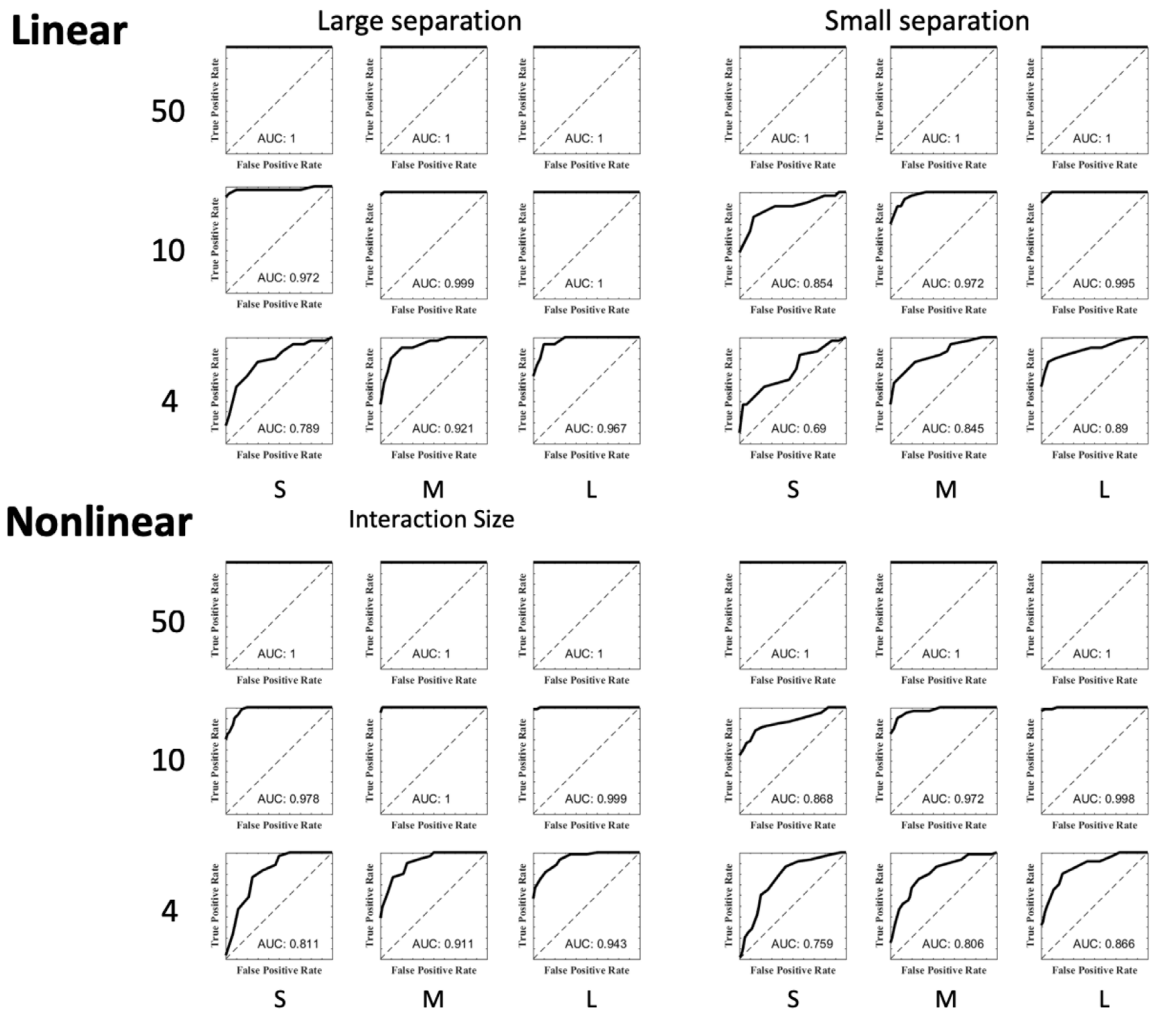

Fig. 7. ROC functions for PIRST, collapsed across levels of noise.

The PIRST procedure also was assessed with receiver-operating characteristic (ROC), analysis, in which the criterion for adjudicating between single- and multiple-process accounts was varied from 0 to 1 in steps of 0.01. An ROC plots the true positive detection rate (concluding multiple processes when there are multiple latent variables) against the false positive rate (concluding multiple processes when there is only a single latent variable). The overall diagnosticity of the procedure is assessed by the total area under that curve (AUC), which ranges from 0.5 when the technique can only diagnose latent structure at a chance rate to 1.0 when it yields perfect performance.

For the simulations presented here, we took the proportion of permutations in each condition in which the permuted SSE exceeded the original SSE, and compared that value to a decision criterion that varied from 0 to 1. When that criterion is 1, the proportion of cases in which the permuted SSE is larger will be at the lowest possible value and will intersect the y-axis at that value. When the criterion is 0, the hit and false-alarm rates will be 1. Criterion values between these two extremes determine the shape of the ROC and the consequent AUC.

To roughly simulate the conditions under which state-trace functions are generated from actual experiments, we computed ROCs and associated AUC values in batches of 100 simulated subjects. The top portion of Table 1 shows the AUC values for every condition of the simulation and Fig. 7 shows the ROCs collapsed across the noise variable.

False alarms. Note that in all cases, values of this statistic rarely exceed 0.5 by much when the latent interaction is absent, indicating that this test is not prone to indicating evidence for multiple processes when multiple processes are not evident. The rare exceptions that do occur do so when the state-trace plot includes a high number of points per condition. This can be seen in the leftmost points of the heatmap in Fig. 6.

Noise. Noise imposed upon the individual data points exhibits a prominent effect on the ability of the model to recover evidence for multiple processes when they actually do contribute to the data. At high levels of noise, high values of detection are only evident under ideal conditions: when there are many data points and when separation between the conditions is small. Overall discriminability, collapsed over the other variables, is revealed by the AUC, which drops from 0.99 to 0.89 over the levels of noise.

Size of the latent interaction. More evidence for multiple processes is evident when the latent interaction is larger. This can be seen in the increasingly hot colors that appear to the right of each quartet of boxes in Fig. 6, and in the lower AUC values for smaller interactions in Table 1. Mean AUC values rise from 0.91 to 0.98 as the interaction size increases. This result indicates that the technique is more apt to reveal effects of multiple processes when the contribution of the latent interaction of those processes is larger.

Overlap of conditions. Recall that state-trace plots have greater potential to be diagnostic when the data points overlap to a

Please cite this article as: A.S. Benjamin, M.L. Griffin and J.A. Douglas, A nonparametric technique for analysis of state-trace functions. Journal of Mathematical Psychology (2019), https://doi.org/10.1016/j.jmp.2019.03.006.

Table 1
Area under the curve (AUC) values for PIRST (top rows), CMR with no partial order constraints (middle rows table), and CMR with partial order constraints (bottom rows). Higher values indicate greater recovery of the true latent state. Extent of deviations from 1 (perfect diagnosticity) are indicated by increasing shades of red.

			Low Noise						Medium Noise						High Noise					
			Low Overlap			High Overlap			Low Overlap			High Overlap			Low Overlap			High Overlap		
	Interaction	Size	S	M	L	S	M	L	S	M	L	S	M	L	S	M	L	S	M	L
PIRST	Linear	50	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		10	1	1	1	1	1	1	1	1	1	1	1	1	0.89	1.00	1	0.64	0.98	1
		4	0.93	1	1	0.94	1	1	0.89	0.93	0.99	0.78	0.96	1	0.65	0.87	0.91	0.65	0.84	0.87
	Concave	50	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		10	1	1	1	1	1	1	1	1	1	1	1	1	0.90	1	0.99	0.70	0.95	1
		4	0.93	0.99	1.00	0.77	0.92	0.99	0.85	0.93	0.98	0.90	0.93	0.99	0.65	0.77	0.80	0.64	0.64	0.74
CMR	Linear	50	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
no constraint		10	1	1	1	1	1	1	1	1	1	1	1	1	0.94	0.95	1	0.90	1.00	1
		4	0.78	0.58	0.99	0.65	0.74	1	0.70	0.52	0.83	0.55	0.57	0.805	0.60	0.78	0.92	0.64	0.76	0.96
	Concave	50	0.5	0.5	0.5	1	1	1	0.83	0.83	0.83	1	1	1	0.90	0.99	1	0.99	1	1
		10	0.73	0.73	0.73	1	1	1	0.90	0.96	0.98	1.00	1	1	0.74	0.78	0.99	0.83	0.98	1
		4	0.98	0.98	0.99	1	1	1	0.86	0.95	0.91	0.75	0.98	0.96	0.70	0.74	0.84	0.79	0.81	0.86
CMR	Linear	50	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
with constraint		10	1	1	1	1	1	1	0.995	1	1	1	1	1	0.94	0.96	1	0.97	1.00	1
		4	0.78	0.59	0.99	0.67	0.74	1	0.70	0.54	0.83	0.55	0.57	0.805	0.61	0.78	0.91	0.61	0.76	0.95
	Concave	50	0.5	0.5	0.5	1	1	1	0.85	0.85	0.85	1	1	1	0.88	0.991	1	0.99	1	1
		10	0.68	0.70	0.70	1	1	1	0.90	0.97	0.98	1.00	1	1	0.75	0.80	1.00	0.85	0.99	1
		4	0.98	0.98	0.99	1	1	1	0.83	0.95	0.92	0.75	0.98	0.96	0.66	0.74	0.81	0.72	0.80	0.85

greater extent. That intuition notwithstanding, PIRST did not reveal overall poorer discriminability under low- than high-overlap conditions. However, overall performance was high across the simulations and may not prove to be a strict test of the degree to which data overlap contributes to test performance.

Data points per condition. PIRST is more able to detect the presence of multiple processes when each individual's state-trace function includes a larger number of data points. Mean AUC values drop from 1.0 (with 50 points) to 0.86 (with only 4 points). However, as can be seen in Fig. 6, it also appears that a large number of data points increases the rate at which the technique yields a false alarm in the absence of a latent interaction. When these two factors are assessed together, it can be seen in Fig. 7 and Table 1 that overall diagnosticity of PIRST decreases with the number of points in the state-trace plot. It is worth pointing out, however, the PIRST performs well with only 10 points per condition, which, though larger than many state-trace experiments in the literature today, is probably not beyond the reach for many experimental protocols.

Latent function shape. The technique seems to be highly robust to deviations from linearity in the underlying function. The bottom heatmaps appear very similar to the ones generated from a linear function, and the AUC values are affected only slightly.

As can be seen, PIRST acts appropriately. It is more likely to indicate support for multiple processes when the effect size is larger, and when there are more data within the state-trace plot. In all cases, the test statistic is higher when the underlying model is multiprocess than when it is not. Across the simulations conducted here, the average AUC for PIRST is 0.95.

Analysis using CMR. The coupled monotonic regression technique was applied to the same simulated data set. This application was mostly automated, using the open-source software provided by Kalish et al. (2016). There are two main parameters to set for the technique: the number of bootstrapped samples generated, and the (optional) addition of any partial order constraints. A partial order constraint can be imposed when there are a priori expectations that the trace manipulation should have a specific directional effect (for example, increasing study time can be expected to improve performance on both dependent variables). We conducted multiple simulations using CMR. In one, we follow the lead of Kalish et al. (2016) and impose no partial order constraints on the analysis. In the second, we impose a partial order constraint. Because trace variables in state-trace experiments (like study time) often provide straightforward order constraints, users of CMR are often likely to be able to take advantage of them, and we wanted to evaluate performance under conditions more relevant to the end user. It turned out that differences in outcomes between the simulations with and without the constraint were not large. Because the outcomes were so similar between the two simulation conditions for CMR, we focus our discussion on the patterns that are evident across simulation conditions. When we cite averaged statistics from CMR in the forthcoming discussion, we are referring to the version with no partial order constraints unless otherwise noted.

We also followed the recommendation of Kalish et al. to use 10,000 bootstrap samples, with the exception of the 50 point conditions. In those conditions, we used only 1000 bootstrap samples.

In order to ensure that the reduction in bootstrap samples did not yield uncomfortably variable estimates of fit, we re-ran a subset of the conditions with 10,000 samples and compared the p-values across those replications. In all cases, the observed difference in p-values were less than 0.01, and the majority were less than 0.001. Based on this outcome, we are comfortable that the reduction in samples did not materially affect the outcome of the fits.

Because CMR computes a traditional null-hypothesis *p*-value, the outcome of the procedure itself is not directly comparable to PIRST, which does not. To directly compare the techniques, we generated ROC functions and computed AUC for CMR. As before, ROCs were generating by varying the criterion from 0 to 1 in steps of 0.01. ROCs collapsed across noise levels are shown in Figs. 8 and 9, and AUC values for every condition are shown in Table 1.

Noise. Across the values of noise implemented here, CMR is impressively resistant to its effects. AUC values do not drop between small (0.90) and large (0.90) levels of noise. Of course, under other compromising conditions, such as small latent interaction size, the effects of noise are evident. These effects can be seen in Table 1.

Size of the latent interaction. CMR responds appropriately, with greater discriminability of underlying structure when the latent interaction is large (AUC = 0.95) than when it is small (AUC = 0.87).

Overlap of conditions. Unlike PIRST, CMR responds as expected to variations in the overlap of data points across conditions. AUC values are higher when overlap is high (0.94) than when it is low (0.87).

Please cite this article as: A.S. Benjamin, M.L. Griffin and J.A. Douglas, A nonparametric technique for analysis of state-trace functions. Journal of Mathematical Psychology (2019), https://doi.org/10.1016/j.jmp.2019.03.006.

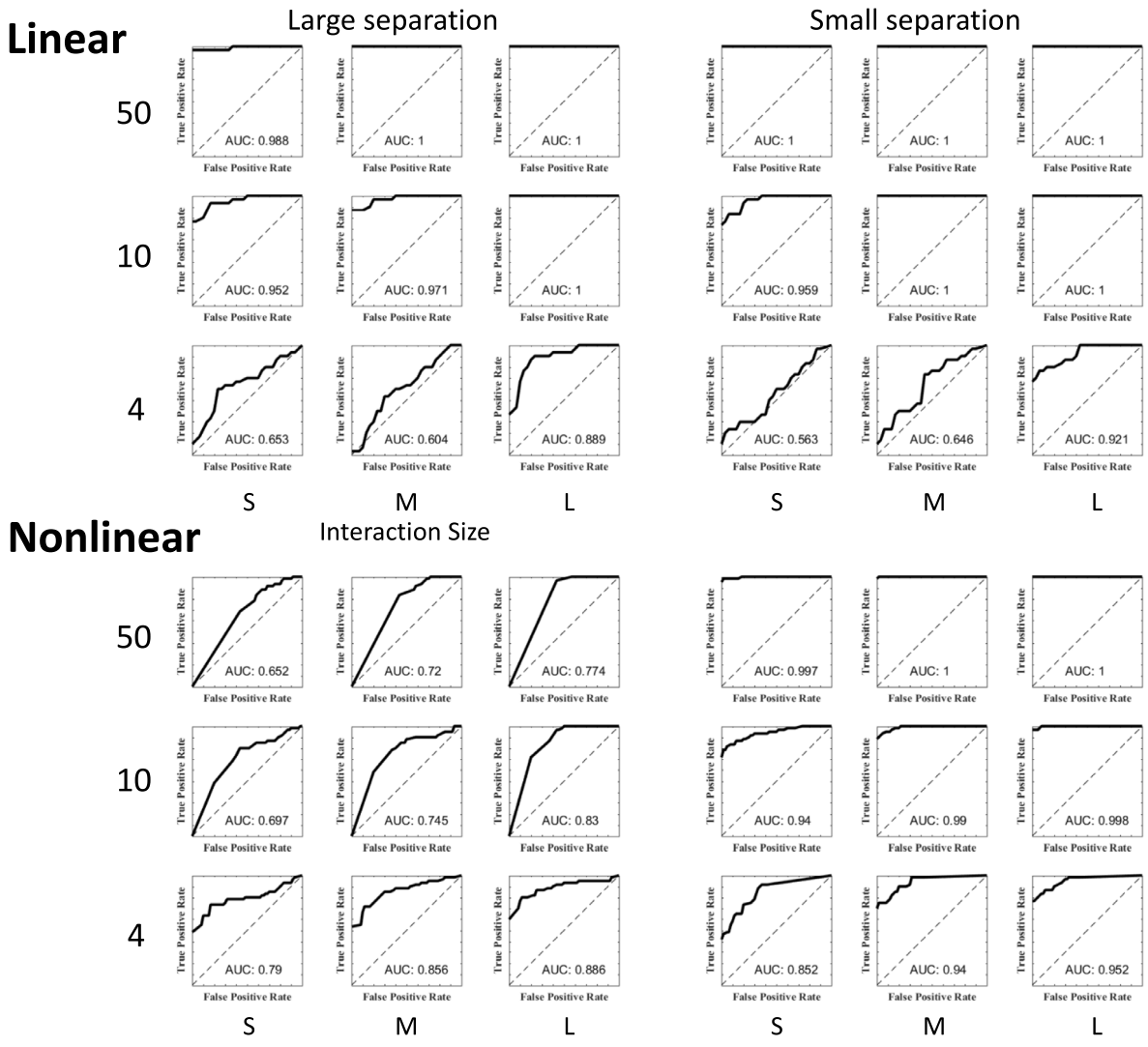

Fig. 8. ROC functions for CMR *with* partial order constraints, collapsed across levels of noise.

Data points per condition. CMR appears to be highly resilient to reductions in the number of data points in the state-trace plot. AUC values range from 0.88 in the 50-point condition to 0.89 in the 4-point condition. However, upon inspection, it can be seen that the effect of data points depends on the contribution of noise. When noise is high, CMR responds as expected, with lower discriminability in the 4-point condition (AUC = 0.79) than in the 50-point condition (AUC = 0.98). However, it exhibits unexpected behavior in the low-noise condition, where it performs more poorly in the 50-point condition (0.75) than in the 4-point condition (0.99).

The puzzling nature of this result led us to replicate a subset of these conditions and confirm the result. Upon examining the actual distribution of p-values for these conditions (available in the online supplement), it can be seen that the procedure is false-alarming at ceiling rates in the low-overlap, low-noise condition. This leads to the chance level diagnosticity seen for these cells in Table 7. This result is obviously of some concern for users of this procedure if it turns out to be general.

Comparison of PIRST and CMR. Overall, both procedures perform quite capably, with mean AUC values of 0.95 for PIRST and 0.90 for CMR across the simulations we conducted here. That small difference should not be taken to mean much, since these values depend on the particular simulations chosen and may well differ given other variables or other implementations of the variables explored here.

Table 2 shows the differences between PIRST and CMR across the same conditions shown in Table 1. There are two general patterns evident in this table. First, PIRST performs more capably under conditions of low and medium noise but is overtaken by CMR under conditions of high noise. Second, these effects are most prominent in the 4-item condition, especially under conditions of high overlap.

One additional result is of note. CMR, as noted earlier, exhibits a surprising tendency to diagnose latent structure more poorly under certain conditions when there are more points in the state-trace function. Those conditions include a concave latent function with low overlap between conditions. In fact, when there are 50 points, CMR has no diagnostic capacity. Whether these findings are a quirk of our simulations or a general problem inherent to CMR remains a target for future work.

There are, however, two important caveats in interpreting the results presented here. First, as mentioned earlier, a wider range of conditions and deeper exploration of relevant variables will be necessary to determine with any generalizability the relative strengths and weaknesses of each approach. Second, the area under the ROC is an important tool for understanding the potential diagnosticity of a procedure, but it does not speak directly to the

Please cite this article as: A.S. Benjamin, M.L. Griffin and J.A. Douglas, A nonparametric technique for analysis of state-trace functions. Journal of Mathematical Psychology (2019), https://doi.org/10.1016/j.jmp.2019.03.006.

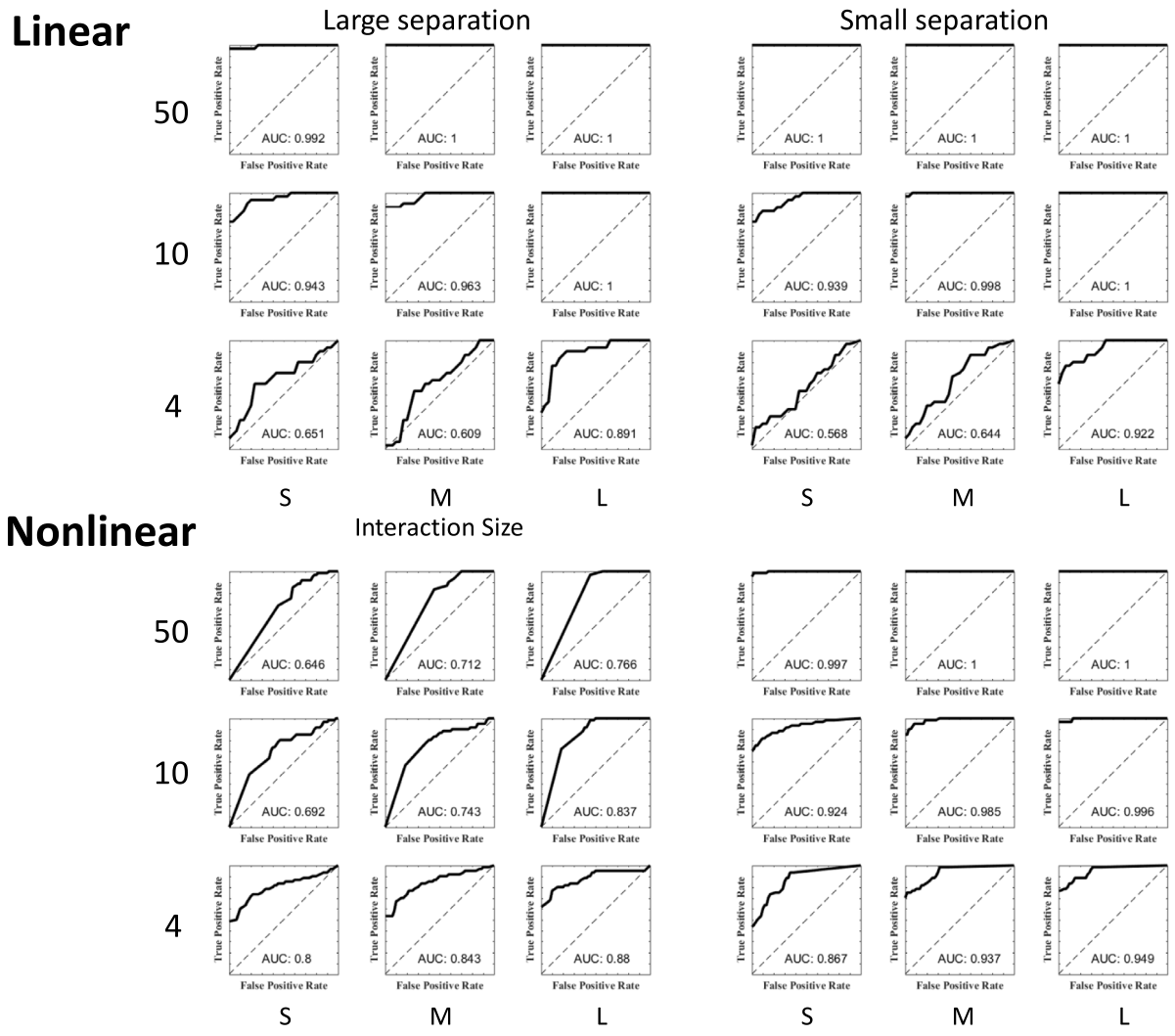

Fig. 9. ROC functions for CMR with *no* partial order constraints, collapsed across levels of noise.

Table 2
Heatmap of the differences in AUC between PIRST and CMR. Top table uses CMR with no partial order constraints, and the bottom table uses CMR with partial order constraints. Yellow boxes indicate conditions that favor PIRST; blue boxes indicate conditions that favor CMR.

PIRST vs. CMR with NO constraint

	Low Noise						Medium Noise						High Noise					
	Low Overlap			High Overlap			Low Overlap			High Overlap			Low Overlap			High Overlap		
Interaction Size	S	M	L	S	M	L	S	M	L	S	M	L	S	M	L	S	M	L
Linear																		
50	0	0	0	0	0	0	0	0	0	0	0	0	0.04	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	-0.05	0.05	0	-0.26	-0.02	0
4	0.15	0.43	0.01	0.29	0.27	0	0.19	0.41	0.16	0.24	0.39	0.20	0.06	0.10	-0.01	0.01	0.08	-0.10
Concave																		
50	0.50	0.50	0.50	0	0	0	0.18	0.18	0.18	0	0	0	0.10	0.01	0	0.01	0	0
10	0.28	0.28	0.28	0	0	0	0.10	0.04	0.02	0.00	0	0	0.16	0.22	0.00	-0.13	-0.03	0.00
4	-0.06	0.01	0.01	-0.22	-0.08	-0.01	0.00	-0.02	0.07	0.16	-0.05	0.03	-0.05	0.04	-0.04	-0.15	-0.17	-0.12

PIRST vs. CMR WITH constraint

	Low Noise						Medium Noise						High Noise					
	Low Overlap			High Overlap			Low Overlap			High Overlap			Low Overlap			High Overlap		
Interaction Size	S	M	L	S	M	L	S	M	L	S	M	L	S	M	L	S	M	L
Linear																		
50	0	0	0	0	0	0	0	0	0	0	0	0	0.05	0	0	0	0	0
10	0	0	0	0	0	0	0.01	0	0	0	0	0	-0.05	0.03	0	-0.33	-0.02	0
4	0.15	0.42	0.01	0.27	0.27	0	0.19	0.39	0.16	0.23	0.39	0.20	0.05	0.10	-0.01	0.04	0.09	-0.08
Concave																		
50	0.50	0.50	0.50	0	0	0	0.15	0.15	0.15	0	0	0	0.12	0.01	0	0.01	0	0
10	0.32	0.30	0.30	0	0	0	0.10	0.03	0.02	0.00	0	0	0.14	0.20	0.00	-0.14	-0.04	0.00
4	-0.05	0.02	0.01	-0.22	-0.08	-0.01	0.02	-0.02	0.07	0.15	-0.05	0.03	-0.01	0.04	-0.01	-0.08	-0.16	-0.12

implementation of a decision rule and the adequacy of a single decision rule across subjects, conditions, or experiments.

4. Summary

State-trace experiments provide a convenient and direct way of contrasting two common hypotheses about the bases for performance: that they derive from a single psychological influence, or from multiple influences. Theories about the joint contribution of dual processes like this appear throughout the history of psychology. At the center of the best-selling popular psychology book *Thinking, Fast and Slow* (Kahneman, 2011) is the presumption that two "systems" contribute to reasoning, one of which is slow and deliberative, and the other of which is fast and heuristic. This distinction goes back (at least) to William James (1890), who postulated a similar distinction between fast associative processes and (slower) "true" reasoning.

Yet the multitude of experiments that arise from such distinctions rarely take seriously the question of how to assess the presence of two distinct processes within a set of data. Dual processes are either taken as a given, and the data are interpreted conditional upon that assumption, or they are tested by one of the many imperfect means that traditional statistical tools have to offer. When the rigor of state-trace analysis is applied to a question, an outcome supporting the operation of multiple processes is not guaranteed. In fact, the application of the state-trace procedure to studies of reasoning has revealed – in contrast to the intuitions of luminaries like James and Kahneman – stronger support for a single than for multiple systems of reasoning (Hayes, Stephens, Ngo, & Dunn, 2018; Stephens, Dunn, & Hayes, 2018).

The underlying problem is a mismatch between the measurement qualities of psychological data and the requirements of popular analytic tools. Linear models require linear assumptions. When data can be reparameterized in such a way so as to have defensively interval qualities (e.g., Matzen & Benjamin, 2009; Wagenmakers et al., 2012), then linear interactions can meaningfully reveal dissociations (Benjamin, 2010). But the typical data gleaned from experiments in psychology – accuracy, response times, rating scale measurements – do not naturally have this capacity and must be treated with caution. State-trace analysis does so by making the minimal assumption of ordinality in measurement.

PIRST asks only the question: do the original data look more convincingly like two separate functions than permuted versions of those data? If they do, then the conclusion is warranted that multiple processes are at work. If they do not, we should be disposed to concluding the action of only a single process.

Acknowledgments

The authors wish to thank members of the Human Memory and Cognition Laboratory at the University of Illinois for feedback, as well as Philip Huebner and Jon Willits for assistance with implementing the computationally intensive simulations.

References

Bamber, D. (1979). State-trace analysis: A method of testing simple theories of causation. *Journal of Mathematical Psychology*, *19*(2), 137–181.
Barlow, R., Bartholomew, D., Bremner, J., & Brunk, H. (1972). *Statistical inference under order restrictions*. New York, NY: Wiley.
Benjamin, A. S. (2010). Representational explanations of process dissociations in recognition: The DRYAD theory of aging and memory judgments. *Psychological Review*, *117*, 1055–1079.
Dunn, J. C., & James, R. N. (2003). Signed difference analysis: Theory and application. *Journal of Mathematical Psychology*, *47*(4), 389–416.
Dunn, J. C., & Kalish, M. L. (2018). *State-trace analysis*. New York: Springer, [ISSN: 2510-1889].
Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin*, *53*(2), 134–140.
Hayes, B. K., Stephens, R. G., Ngo, J., & Dunn, J. C. (2018). The dimensionality of reasoning: Inductive and deductive inference can be explained by a single process. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(9), 1333–1351.
Iverson, G. J. (2006). An essay on inequalities and order-restricted inference. *Journal of Mathematical Psychology*, *50*, 215–219.
Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*(5), 513–541.
James, W. (1890). *The principles of psychology (Vol. I)*. Henry Holt.
Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.
Kalish, M. L., Dunn, J. C., Burdakov, O. P., & Sysoev, O. (2016). A statistical test of the equality of latent orders. *Journal of Mathematical Psychology*, *70*, 1–11.
Loftus, G. R. (1978). On interpretation of interactions. *Memory & Cognition*, *6*(3), 312–319.
Loftus, G. R., Oberg, M. A., & Dillon, A. M. (2004). Linear theory, dimensional theory, and the face-inversion effect. *Psychological Review*, *111*(4), 835–863.
Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*(3), 252–271.
Matzen, L. E., & Benjamin, A. S. (2009). Remembering words not presented in sentences: How study context changes patterns of false memories. *Memory & Cognition*, *37*, 52–64.
Mccarley, J. S., & Grant, C. (2008). State-trace analysis of the effects of a visual illusion on saccade amplitudes and perceptual judgments. *Psychonomic Bulletin & Review*, *15*(5), 1008–1014.
Michell, J. (1986). Measurement scales and statistics: A clash of paradigms. *Psychological Bulletin*, *100*(3), 398–407.
Newell, B. R., & Dunn, J. C. (2008). Dimensions in data: Testing psychological models using state-trace analysis. *Trends in Cognitive Sciences*, *12*(8), 285–290.
Newell, B. R., Dunn, J. C., & Kalish, M. (2010). The dimensionality of perceptual category learning: A state-trace analysis. *Memory & Cognition*, *38*(5), 563–581.
Prince, M., Brown, S., & Heathcote, A. (2012). The design and analysis of state-trace experiments. *Psychological Methods*, *17*(1), 78–99.
Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108.
Regenwetter, M., & Cavagnaro, D. R. (2018). Tutorial on removing the shackles of regression analysis: How to stay true to your theory of binary response probabilities. *Psychological Methods*, *24*(2), 135–152.
Sidman, M. (1952). A note on functional relations obtained from group data. *Psychological Bulletin*, *49*(3), 263–269.
Stephens, R. G., Dunn, J. C., & Hayes, B. K. (2018). Are there two processes in reasoning? The dimensionality of inductive and deductive inferences. *Psychological Review*, *125*(2), 218–244.
Verhaeghen, P., & Cerella, J. (2002). Aging, executive control, and attention: A review of meta-analyses. *Neuroscience & Biobehavioral Reviews*, *26*(7), 849–857.
Wagenmakers, E. J., Krypotos, A. M., Criss, A. H., & Iverson, G. (2012). On the interpretation of removable interactions: A survey of the field 33 years after Loftus. *Memory & Cognition*, *40*(2), 145–160.
Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, *114*(1), 152–176.

Please cite this article as: A.S. Benjamin, M.L. Griffin and J.A. Douglas, A nonparametric technique for analysis of state-trace functions. Journal of Mathematical Psychology (2019), https://doi.org/10.1016/j.jmp.2019.03.006.