

# Assessing bimodality to detect the presence of a dual cognitive process

Jonathan B. Freeman · Rick Dale

© Psychonomic Society, Inc. 2012

**Abstract** Researchers have long sought to distinguish between single-process and dual-process cognitive phenomena, using responses such as reaction times and, more recently, hand movements. Analysis of a response distribution's modality has been crucial in detecting the presence of dual processes, because they tend to introduce bimodal features. Rarely, however, have bimodality measures been systematically evaluated. We carried out tests of readily available bimodality measures that any researcher may easily employ: the bimodality coefficient (BC), Hartigan's dip statistic (HDS), and the difference in Akaike's information criterion between one-component and two-component distribution models ( $AIC_{diff}$ ). We simulated distributions containing two response populations and examined the influences of (1) the distances between populations, (2) proportions of responses, (3) the amount of positive skew present, and (4) sample size. Distance always had a stronger effect than did proportion, and the effects of proportion greatly differed across the measures. Skew biased the measures by increasing bimodality detection, in some cases leading to anomalous interactive effects. BC and HDS were generally convergent, but a number of important discrepancies were found.  $AIC_{diff}$  was extremely sensitive to bimodality and identified nearly all distributions as bimodal. However, all measures served to detect the presence of bimodality in comparison to unimodal simulations. We provide a validation with experimental data, discuss

methodological and theoretical implications, and make recommendations regarding the choice of analysis.

**Keywords** Bimodality · Response distribution · Dual-process models · Single-process models · Mouse-tracking

## Introduction

The idea that the human mind makes use of two distinct kinds of processing enjoys a long tradition, one that can be traced back at least two millennia. In 350 B.C.E., Aristotle claimed that the mind is composed of two separate systems, one supporting intuition and another supporting reasoning. Over the last few decades, experimental psychology has seen a wide proliferation of theories and models that mirror this age-old view. These dual-process accounts have been extensively reviewed elsewhere (Evans, 2008; Keren & Schul, 2010). Often, these accounts propose the tandem operation of one process that is relatively fast, nonconscious, automatic, and coarse and a separate process that is relatively slow, conscious, deliberate, and fine-grained (Evans, 2008).

In the domain of social cognition, for example, it is theorized that people initially evaluate others in a nonconscious, automatic fashion. This rapid evaluation may be subsequently modified by a more conscious and deliberate assessment, which takes more time. For instance, an individual's prejudice may lead to a rapid negative reaction to another person, but this may be controlled by a more deliberate motivation to be nonprejudicial (Devine, 1989). In memory research, dual-process accounts argue that an item's recognition is the result of two separate processes, familiarity and recollection. Familiarity is relatively fast, involving a coarse assessment of whether an item has been previously

---

J. B. Freeman (✉)  
Department of Psychological & Brain Sciences,  
Dartmouth College,  
6207 Moore Hall,  
Hanover, NH 03755, USA  
e-mail: jon.freeman@dartmouth.edu

R. Dale  
University of California, Merced,  
Merced, CA, USA

encountered and lacking access to specific details. Recollection, on the other hand, is relatively slow, involving a more fine-grained assessment and explicit retrieval of an item's details (Atkinson & Juola, 1973; Jacoby, 1991). In language research, some accounts, such as the unrestricted-race account, propose that an initial syntactic structure is selected on the basis of evidence accumulated during sentence processing. When syntax is substantially ambiguous, this account posits that a separate, slower reanalysis process may then intervene on the initial, quick commitment to a syntactic structure if it turns out to be inappropriate (van Gompel, Pickering, Pearson, & Liversedge, 2005; van Gompel, Pickering, & Traxler, 2001).

Common to all of these accounts is the presence of two processes that work on different temporal scales. A consequence of this is the prediction that a participant's set of responses, in certain conditions, is being drawn from two separate populations. This is because, during some trials within an experimental condition, the second, slow process may be in agreement with the first, quick process, but on other trials, the two processes may be in disagreement. For example, according to an unrestricted-race account, on some trials in a sentence processing experiment an initially selected syntactic structure will turn out to be inappropriate and therefore need to be corrected by subsequent reanalysis. On other trials in the same condition, the initially selected structure will in fact be correct and need no intervention (van Gompel et al., 2005). As such, the unrestricted-race account predicts that a participant's response distribution derives from two separate trial populations, one involving zero intervention and a second involving intervention. Thus, this account, like many others hypothesizing two independent processes, predicts that the distributions of behavioral measures based on these responses will exhibit bimodality.

Whereas dual-process accounts predict the presence of bimodality in certain response distributions, other single-process accounts seek to disconfirm that any bimodality is present, because these predict unimodal distributions. For example, constraint-based accounts of sentence processing argue that the selection of a syntactic structure is accomplished by a single process involving dynamic competition, rather than two independent processes involving reanalysis. Thus, constraint-based accounts predict that certain response distributions will be unimodal. In particular, when syntax is ambiguous, these accounts predict that trials will reveal a single, continuous range of competition between the possible interpretations of the ambiguity, thereby giving way to a unimodal distribution over behavioral responses (Farmer, Anderson, & Spivey, 2007). A similar difference in predictions occurs in social categorization research. When categorizing the sex of a sex-atypical male face (e.g., one containing slight feminine features), for example, discrete stage-based approaches to social categorization predict a

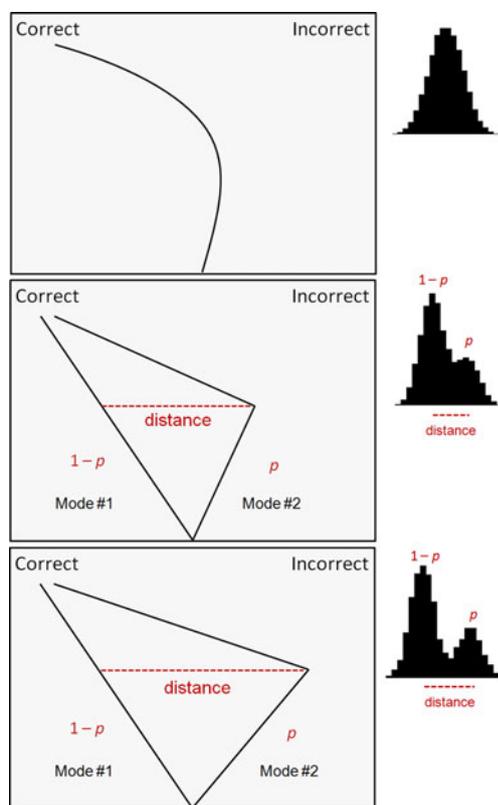
bimodal response distribution. This is because these approaches assume that an initial categorization is made on the basis of coarse perceptual features (e.g., female), which sometimes may need to be intervened on by a more fine-grained reanalysis (e.g., male) if the initial categorization turns out to be incorrect. At other times, the initial categorization would be correct and need no intervention. Thus, on some trials there is intervention, and on other trials there is none, forming a bimodal distribution. A dynamic interactive approach, on the other hand, predicts a unimodal distribution. This is because it assumes that such sex-atypical faces will always trigger the same single process involving dynamic competition between sex categories (Freeman & Ambady, 2011a), but that the competition among possible interpretations this system can make will give way to a normal distribution over the strength of the competition (Freeman, Ambady, Rule, & Johnson, 2008).

Examining a response distribution's characteristics in order to distinguish between competing theoretical accounts has a long history with reaction time measurements (Ratcliff, 1979). Recently, distributional analyses have become increasingly important with the advent of more continuous, temporally fine-grained measures that index participants' tentative commitments to various response alternatives during online processing. For example, in studies recording hand movement trajectories (via a computer mouse, wireless remote, or electromagnetic position tracker), analysis of a response distribution's modality has been crucial in distinguishing between accounts of categorization (Dale, Kehoe, & Spivey, 2007; Freeman & Ambady, 2009, 2011b; Freeman et al., 2008; Freeman, Pauker, Apfelbaum, & Ambady, 2010), language processing (Dale & Duran, 2011; Farmer et al., 2007; Spivey, Grosjean, & Knoblich, 2005), decision making (McKinstry, Dale, & Spivey, 2008), learning (Dale, Roche, Snyder, & McCall, 2008), visual search (Song & Nakayama, 2008), and attentional control (Song & Nakayama, 2006). We take this particular methodology as one especially ripe for investigating measures of bimodality and for distinguishing between single-process and dual-process phenomena.

For example, in one series of studies, participants categorized faces' sex by moving the computer mouse from the bottom center of the screen to either the top-left or top-right corner, which were marked "male" and "female" (Freeman et al., 2008). When categorizing sex-atypical faces, participants' mean mouse trajectories showed a continuous attraction to the opposite sex-category response (on the opposite side of the screen), relative to sex-typical faces. This mean continuous-attraction effect could reflect either a single-process phenomenon involving dynamic competition or, alternatively, a dual-process phenomenon involving an initial analysis and subsequent reanalysis. If the effect has a unimodal distribution, in which some trials involve strong

attraction, some medium attraction, and some weak attraction, it would suggest that sex-atypical faces triggered dynamic competition between parallel, partially active sex categories. However, if the effect has a bimodal distribution, in which some trials involve zero attraction and others involve extremely strong attraction, it would suggest that categorization of sex-atypical faces involved dual processes that sometimes agreed and sometimes conflicted. On some trials, an initial perceptual analysis and subsequent fine-grained reanalysis would agree, resulting in zero attraction (a discrete movement straight to the correct category). On other trials, the initial analysis (e.g., female) would turn out to be incorrect and require intervention from later reanalysis (e.g., male), resulting in extremely strong attraction (an initial discrete movement to the incorrect category, which would have to be redirected midflight by a corrective movement straight to the correct category; Freeman et al., 2008).

Thus, in accounts hypothesizing two independent processes, sometimes the two processes can agree with each other. For the sake of presentation here, we will call this kind of response “Mode 1.” At other times the two processes can be in conflict, and we will call this “Mode 2.” A schematic illustration appears in Fig. 1, which shows how



**Fig. 1** A schematic demonstration of how a dual process introduces bimodal features, using the distribution of responses. In this paradigm, typically a stimulus is presented and participants move the mouse from the bottom center of the screen to the top-left or top-right corner (Freeman & Ambady, 2010; Spivey et al., 2005)

a dual process introduces bimodal features using the two-choice mouse-tracking paradigm (although it applies to any behavioral measure—e.g., reaction times). In this paradigm, typically a stimulus is presented and participants move the mouse from the bottom center of the screen to the top-left or top-right corner (Freeman & Ambady, 2010; Spivey et al., 2005). In this figure, the top-left corner represents the correct response and the top-right corner represents the incorrect response. Each panel is a depiction of one experimental condition. The top panel shows one unimodal population of trajectories that all show an attraction toward the incorrect response (sometimes strong, sometimes medium, sometimes weak), which is often interpreted by single-process accounts as dynamic competition. The lower panels show bimodal populations of trajectories, in which dynamic competition is not present. Instead, some proportion ( $1 - p$ ; Mode 1) of trials involve a discrete movement toward the correct category, and the rest of the trials ( $p$ ; Mode 2) involve an initial discrete movement toward the incorrect category, which is then redirected in midflight by a discrete movement to the correct category. The middle panel depicts a population of trajectories with a recognizable amount of separation between Mode 1 and Mode 2 responding, whereas the bottom panel depicts a population with more extreme separation. Both panels depict a pattern of results consistent with dual-process accounts, where Mode 1 responses occur when an earlier and a later process agree and Mode 2 responses occur when the two processes conflict. Importantly, when all trajectories in each panel are averaged together into a mean trajectory for the experimental condition, the three mean trajectories would look quite similar, resembling something like the top panel’s trajectory (see Freeman et al., 2008, Study 3). This highlights the importance of examining distributional characteristics, as the underlying pattern of responses may be quite different, although the mean effects look virtually the same.

As is shown in Fig. 1, a response distribution’s shape would be strongly affected by single versus dual modes of responding, with dual modes introducing bimodal features. Specifically, two parameters are likely to affect the distributional shape. One of these parameters is the *distance* between mean responses in Mode 1 and Mode 2. In a mouse-tracking paradigm, this could be the difference in the trajectories’ deviations toward the incorrect response between Mode 1 and Mode 2 responses. As Mode 2 responses become more extreme, the distance increases between the two peaks of the bimodal distribution, as illustrated in Fig. 1. This is not limited to a mouse-tracking paradigm; for example, this distance could refer to a difference in reaction times (e.g., Atkinson & Juola, 1973; Ratcliff, 1979). The other parameter is the *proportion* of responses in Mode 2. If the likelihood of a Mode 2 response is 25 % (as in Fig. 1), it is easy to superficially observe bimodality in

the response distribution. However, if the likelihood of a Mode 2 response is only 5 %, for example, observing bimodality is likely to be substantially more difficult, because the Mode 2 population could be obscured by the considerably larger Mode 1 population, thereby feigning unimodality.

### Distinguishing between unimodality and bimodality

Researchers have used several measures to distinguish between unimodality and bimodality, including the bimodality coefficient (BC; SAS Institute, 1989), Hartigan's dip statistic (HDS; Hartigan & Hartigan, 1985), and the difference in Akaike's information criterion (AIC; Akaike, 1974) between one-component and two-component Gaussian mixture distribution models (McLachlan & Peel, 2000). An extensive discussion of these measures is beyond the scope of this article, but we provide a brief description of each. In the present work, we focus on utilizing these measures "out of the box"—that is, on how a researcher's estimation of bimodality may be done with readily available scripts and other sources. The measures that we employ have this property of accessibility and ease of application (see the [Appendix](#) for our code).

The BC is based on an empirical relationship between bimodality and the third and fourth statistical moments of a distribution (skewness and kurtosis). It is proportional to the division of squared skewness with uncorrected kurtosis,  $BC \propto (s^2 + 1)/k$ , with the underlying logic that a bimodal distribution will have very low kurtosis, an asymmetric character, or both; all of these conditions increase BC. The values range from 0 and 1, with those exceeding .555 (the value representing a uniform distribution) suggesting bimodality (SAS Institute, 1989).

The HDS is a statistic calculated by taking the maximum difference between the observed distribution of data and a uniform distribution that is chosen to minimize this maximum difference. The idea is that repeated sampling from the uniform (with the sample size of the original data) produces a sampling distribution over these differences; a bimodal (or  $n$ -modal) distribution is one in which the HDS is at or greater than the 95th percentile among all sampled values. In other words, as compared to the uniform distribution (which Hartigan & Hartigan, 1985, argued to be the best choice for testing unimodality), a multimodal distribution has statistically significant disparities in its distribution function. Thus, the HDS is given to null-hypothesis logic and is inferential; if  $p < .05$ , the distribution is considered to be bimodal or multimodal (Hartigan & Hartigan, 1985).

Finally, the AIC is a well-known information-theoretic goodness-of-fit measure for an estimated statistical model, with lower AIC values indicating better fit. To assess modality, one can fit the observed data using one-component (i.e., unimodal) and two-component (i.e., bimodal) Gaussian

mixture distribution models to determine which of the two models minimizes AIC (McLachlan & Peel, 2000). If the one-component model minimizes AIC, the distribution is better described as unimodal; if the two-component model minimizes AIC, the distribution is better described as bimodal. Importantly, the AIC weighs the likelihood score of a model against the number of parameters used to construct the model. If the AIC for a bimodal mixture model is smaller than that of a unimodal model, it suggests that the goodness of fit exceeds the cost of having an additional component in the model.<sup>1</sup>

As described earlier and exemplified by Fig. 1, the presence of a dual process can affect a response distribution by introducing bimodal features. The degree of bimodality would be influenced by two important factors, the distance in mean responses between Modes 1 and 2, and the proportion of responses in Mode 2 (vs. Mode 1). To examine how these factors affect the distributional shape and the detection of bimodality using BC, HDS, and AIC measures, we systematically manipulated them in a number of simulations. We also manipulated the degree of positive skewness in the response distribution. It is quite common for reaction time distributions to exhibit positive skewness, and indices of spatial attraction, curvature, or deviation in hand-trajectory data (see Freeman & Ambady, 2010) commonly exhibit positive skewness as well. Given that such distributions often feature positive skewness, it is important to understand how skew might influence the detection of bimodality.

The HDS measure initially proposed in Hartigan and Hartigan (1985) was meant to test the null hypothesis of unimodality against the alternative hypothesis of multimodality, with the null of an asymptotic uniform distribution. Though it is widely utilized in the bimodal context, the test was intended to explore departures from unimodality of a kind that may have more than two modes. Essentially, it tests the departure of an observed density function from a unimodal one (assumed to have a single inflection point between convex and concave segments). This means that HDS is relatively more robust to skew: Regardless of the location of the center of the observed function, HDS tests the observed function against the presence of a single inflection point. In both BC and AIC, high skew may significantly impact the test for the presence of more than one mode, so that both may sometimes find a spurious second mode in long-tailed, skewed distributions (a point originally made by Hartigan & Hartigan, 1985). Nevertheless, researchers have often used measures such as BC or AIC with distributions containing high skew.

<sup>1</sup> We found that an improvement in  $AIC_{diff}$  could be achieved by choosing a threshold proportion, often within the range of .1 to .25. However, there appears to be no a priori basis upon which to choose this optimal threshold, making it difficult to apply in the context of a single sample.

Another important, yet often underappreciated, issue in judging the modality of distributions in psychological experiments is sample size. For example, in the original mouse-tracking study (Spivey et al., 2005), the authors noted that having too few trials within an individual subject posed problems for assessing bimodality on a per-subject basis; instead, they opted for assessing it at the group level. This has now become the norm in mouse-tracking research (see, e.g., Freeman & Ambady, 2010). How many trials, then, is too few? In general, the sampling error of skewness and kurtosis are high at smaller sample sizes (10 or fewer), suggesting that the BC, which is computed from these parameters, may be unstable at smaller sample sizes. HDS is based on an empirical resampling from a uniform distribution, and thus may naturally correct for smaller sample sizes, as the simulated distribution will accommodate this. Examining the performance of all three measures—BC, AIC, and HDS—with unimodal and bimodal distributions of varying skew and sample size, among other factors that are known to drive modality, has therefore long been needed.

#### The present work

Given that the BC and AIC measures (and not the HDS measure) may be substantially biased by skew and sample size, we hypothesized that HDS may be an overall more robust measure for judging the modality of a distribution. As we have discussed, adjudicating between unimodality and bimodality has become an increasingly important issue in psychological experiments, with unimodal and bimodal outcomes often leading to opposite interpretations of cognitive dynamics (e.g., Dale & Duran, 2011; Farmer et al., 2007; Freeman et al., 2008). Here, we provide a comprehensive analysis of three modality measures with the aim of making a recommendation for measure selection in future research. To this end, we have taken a two-pronged approach. First, we systematically examined the measures' performance in simulations in which the factors described above were tightly controlled and varied. Then, to validate the results and increase their generality, we examined the measures' performance with previously published experimental data, which contained distributions theoretically known to be either unimodal or bimodal.

#### Simulations

We independently varied the four parameters mentioned above—*distance*, *proportion*, *skew*, and *size*—across 1,760 different simulations (8 levels of distance  $\times$  11 levels of proportion  $\times$  5 levels of skew  $\times$  4 levels of size). In each simulation, 250, 500, 1,000, or 2,000 simulated observations were randomly sampled from a standard Gaussian

distribution (depending on the size parameter). These might correspond to trials aggregated across subjects that could reflect, for example, reaction times or an hand trajectory's spatial attraction, curvature, or deviation. They could also correspond to trials within a single subject. In general, however, these numbers reflected a standardized distribution that may be observed in a wide range of behavioral measures, such as reaction times (likely the most common behavioral measure subjected to distribution analysis; Van Zandt, 2000).

#### Method

To simulate a separate population of Mode 2 responses, a varying proportion of the total observations was shifted by a varying distance (i.e., a varying amount of *SDs*) by adding a constant (the distance parameter). The proportion was varied as 0 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, and 50 %. Thus, when proportion = 5 % and size = 2,000, 100 observations belong to Mode 2, and when proportion = 50 % and size = 2,000, 1,000 observations belong to Mode 2. We assumed that when proportion = 0 %, a distribution is unimodal, and when proportion > 0 %, it is bimodal. The distance was varied over 1, 2, 3, 4, 5, 6, 7, and 8 *SDs*. When distance = 1 *SD*, the population of Mode 2 responses is quite close to the population of Mode 1 responses (1 *SD* away), and when distance = 8 *SD*, the population of Mode 2 responses is quite extreme, considerably shifted away from the population of Mode 1 responses (8 *SDs* away). To introduce varying levels of positive skew, each observation was then exponentiated to the power of 1, 2, 3, 4, or 5 (the skew parameter). Thus, when skew = 1, the distribution is normal with zero skew, and when skew = 5, the distribution is highly positively skewed. This was done after shifting responses. All observations were then restandardized ( $M = 0$ ,  $SD = 1$ ).

We assessed bimodality in the final resulting distribution by computing BC and HDS in MATLAB (Mechler, 2002). The resulting distribution was also fit to a one-component and a two-component Gaussian mixture distribution model (using an expectation-maximization algorithm with `gmdistribution.fit` in MATLAB; McLachlan & Peel, 2000), and AIC was computed for the two models (one-component,  $AIC_1$ ; two-component,  $AIC_2$ ). We then computed a proportional difference score ( $AIC_{diff}$ ) that represents the relative better fit of the bimodal mixture model over the unimodal one:  $(AIC_1 - AIC_2) / \max(AIC_1, AIC_2)$  (see Wagenmakers & Farrell, 2004). See Appendix for all code.

Each of the 1,760 simulations was run 10 times, wherein a new set of responses were sampled from a Gaussian distribution and processed again through the above procedures.

## Results

We present the results of the simulations in the following way. First, we examine the extent to which the bimodality measures would predict that scores derived from a unimodal distribution (proportion = 0 %) or a bimodal distribution (proportion > 0 %) were indeed bimodal. Second, we assess how the various parameters influence the behavior of the measures. To do so, we evaluate the relative contributions of the distance, proportion, skew, and size parameters on the bimodality measures. We regressed BC, HDS, and  $AIC_{diff}$  (in separate analyses) onto mean-centered distance, proportion, skew, and size values. Because we suspected the possibility of interactive influences on the bimodality measures, we modeled interaction effects as well. Most effects were highly significant ( $p < .001$ ), as can be seen below in Table 1. Depicted in Figs. 2 and 3 are contour maps that plot BC, HDS, and  $AIC_{diff}$  values as a function of the distance and proportion parameters, separately for zero-skew distributions (skew exponent = 1) and highly positively skewed distributions (skew exponent = 5), and separately for large distributions (size = 2,000; Fig. 2) and small distributions (size = 250; Fig. 3). We expected that as proportion and distance increased, bimodality should be more easily detected. More important was how skew and size might influence the bimodality measures, perhaps interactively with the proportion and distance parameters. Our focus was on the relative sensitivity of the bimodality measures in response to change in these parameters. First, we explored how the bimodality measures helped to detect proportions of 0 % (unimodal) versus > 0 % (bimodal).

**Table 1** Standardized  $\beta$  values for distance, proportion, skew, and size effects on bimodality measures

Parameter	BC	HDS	$AIC_{diff}$
Skew	0.167***	-0.206***	0.354***
Distance	0.921***	-0.589***	0.795***
Proportion	-0.093***	-0.44***	-0.195***
Size	-0.014***	-0.067***	0.012***
Skew $\times$ Distance	-0.090***	0.170***	-0.126***
Skew $\times$ Proportion	-0.157***	-0.026***	-0.120***
Distance $\times$ Proportion	0.043***	0.050***	-0.036***
Skew $\times$ Size	-0.015***	-0.018***	-0.002
Distance $\times$ Size	0.022***	-0.032***	0.002
Proportion $\times$ Size	0.006**	0.057***	0.000
Skew $\times$ Distance $\times$ Proportion	0.026***	0.091***	0.014***
Skew $\times$ Distance $\times$ Size	0.008***	0.027***	-0.001
Skew $\times$ Proportion $\times$ Size	-0.001	0.009	-0.004
Distance $\times$ Proportion $\times$ Size	0.000	0.067***	0.003
Skew $\times$ Distance $\times$ Proportion $\times$ Size	-0.001	0.002	-0.002

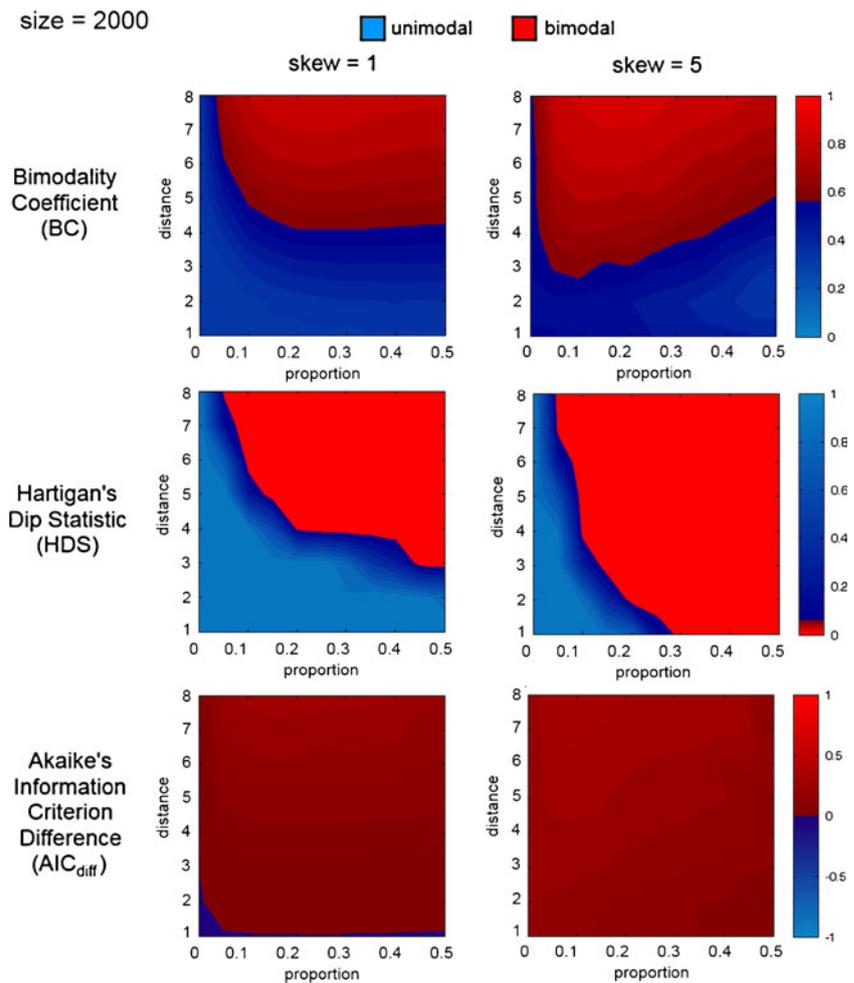
\*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$

## Bimodality detection

All three measures significantly covaried with the bimodality of the simulated data sets. To show this, we created a dichotomous code for each of the 17,600 simulations, scoring unimodal (proportion = 0 %; coded as 0) and bimodal (proportion > 0 %, coded as 1) simulations. Because bimodality is theoretically known to be the case for any proportion that is > 0 %, we could use this dichotomous score in a logistic regression to test whether the three measures, BC, HDS, and  $AIC_{diff}$ , at minimum covaried with the presence of bimodality. This was the case, and the three measures covaried to similar extents. The coefficient estimates in the logistic regression were 8.7, -5.8, and 8.5 for BC, HDS, and  $AIC_{diff}$ , respectively ( $ps < .0001$ ). This result merely indicates that the measures behaved as expected: They covaried systematically around the presence of bimodality, on average, across all parameters. However, it is a separate question whether a researcher, inspecting one of these measures, would infer from it that there was significant presence of bimodality. For this to occur, the researcher would judge BC to be greater than .555, HDS to be  $p < .05$ , and  $AIC_{diff}$  to be greater than 0.

When looking at the scores in this way, the measures behaved considerably differently. In a signal detection framework, we first considered false positives or “misses” in the simulations, when bimodality was inferred though none was actually present in the population (given our parameters). Among the simulations with unimodality (proportion = 0 %), BC was beyond the .555 threshold 21 % of the time; HDS had a significant  $p$  value 0 % of the time; and  $AIC_{diff}$  judged the two-component Gaussian model to be a more economical fit 81 % of the time.  $AIC_{diff}$  thus generated a large number of false positives, judging 81 % of the unimodal simulations to be best fit by a two-component model. BC was second in its number of false positives, judging 21 % of the unimodal simulations to be past threshold for bimodality. HDS produced 0. We carried out a similar test, this time looking at the inference that one would make about bimodality when the proportion was > 0 %: that is, for true positives or “hits.” In this case, BC hit threshold in 65 % of the cases; HDS in 58 % of the cases; and  $AIC_{diff}$  in 94 % of the cases. The pattern of false negatives or “false alarms” therefore had a complementary pattern: HDS was more conservative, judging only just over half of the simulated bimodal distributions as being bimodal.  $AIC_{diff}$  judged most of the bimodal distributions to be bimodal. This may suggest that the Gaussian mixture model assesses distributions to be more economically fit by two-component models overall. Together, these results led HDS to have the highest sensitivity to bimodality ( $d' = 3.43$ ), BC to have the second highest ( $d' = 1.46$ ), and  $AIC_{diff}$  the lowest ( $d' = 0.74$ ). A summary of the signal detection characteristics of the three measures appears in Table 2.

**Fig. 2** Contour maps depicting the BC, HDS, and  $AIC_{diff}$  measures as a function of the distance and proportion parameters, separately for zero-skew distributions (skew exponent=1) and highly positively skewed distributions (skew exponent=5), for simulations in which size=2,000. Blue colors indicate unimodality, and red colors indicate bimodality



The previous analysis tested the detection capability of the bimodality measures when the population was known. We also asked how each of our measures would be influenced by the four parameters. To this end, we constructed a regression model predicting each of the measures using the parameters that we manipulated in the simulations. Because we found that the size parameter had the smallest and most negligible influence on the three measures (discussed below), for ease of presentation we will first describe the influences of the distance, proportion, and skew parameters on each measure, and focus our discussion on distributions of size = 2,000. Afterward, we will describe the influence of size on the three measures.

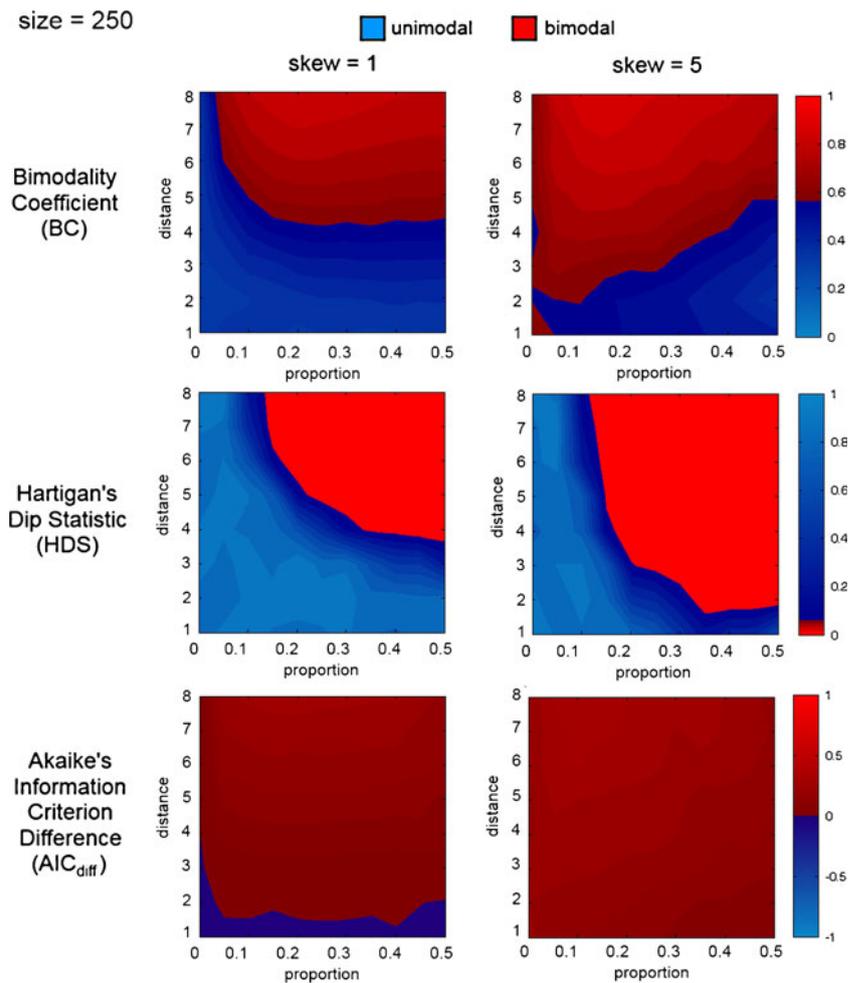
#### *Bimodality coefficient (BC)*

As described above, the BC ranges from 0 to 1, with values greater than .555 suggesting bimodality. The distance between Mode 1 and Mode 2 responses had a considerably stronger influence on the BC ( $\beta = .92$ ) than did the

proportion of Mode 2 responses ( $\beta = -.09$ ), the degree of skew present ( $\beta = .17$ ), or the size of the sample ( $\beta = -.01$ ). Increases in distance led to increases in BC, indicating more bimodality. Distributions with a Mode 2 population farther than 6 SDs away from the Mode 1 population were all recognized as bimodal, and distributions with a Mode 2 population within 3 SDs from the Mode 1 population were all recognized as unimodal, regardless of proportion or skew.

Somewhat unexpectedly, proportion had a very small negative overall influence on BC. This main effect was qualified by strong interactive influences. In zero-skew distributions (skew exponent = 1), a Mode 2 proportion of 5% required approximately a distance of 6 SDs to detect bimodality ( $BC > .555$ ), and the required distance declined to approximately 4 SDs by a proportion of 20%, remaining at that level until a proportion of 50%. In highly positively skewed distributions (skew exponent = 5), however, a proportion of 5% required an approximate distance of 3 SDs to detect bimodality, and this remained roughly the same until a proportion of 20%, which rose thereafter to approximately 5 SDs by a proportion of 50% (Fig. 2). Thus, skew had

**Fig. 3** Contour maps depicting the BC, HDS, and  $AIC_{diff}$  measures as a function of the distance and proportion parameters, separately for zero-skew distributions (skew exponent=1) and highly positively skewed distributions (skew exponent=5), for simulations in which size=250. Blue colors indicate unimodality, and red colors indicate bimodality



several interesting influences on bimodality detection with the BC measure. First, it made bimodality detection easier, with an overall greater propensity for  $BC > .555$ . Second, whereas BC increased as proportion increased for zero-skew distributions, BC decreased as proportion increased for highly positively skewed distributions. Thus, the presence of positive skew made distributions with higher proportions of Mode 2 responses require a larger, rather than a smaller, distance for bimodality to be detected. This might have been the case because the elongated positive tail of the Mode 1 distribution, due to positive skew, led the Mode 1

distribution to “blend” into the Mode 2 distribution, thereby requiring a greater distance between the two distributions for bimodality to be detected.

#### *Hartigan's dip statistic (HDS)*

For the HDS measure, we used  $p$  values resulting from Hartigan's dip test. Thus, HDS values range from 0 to 1, with values less than .05 indicating significant bimodality, and values greater than .05 but less than .10 suggesting bimodality with marginal significance. As with the BC measure, distance had a considerably stronger influence ( $\beta = -.59$ ) than did proportion ( $\beta = -.44$ ) or skew ( $\beta = -.21$ ). The size parameter appeared to have a negligible influence ( $\beta = -.07$ ). As distance increased, HDS decreased. More heavily than in the BC results, the presence of positive skew biased HDS values, leading to a greater propensity overall for distributions to be recognized as bimodal. Differing from the BC results, however, proportion had a strong influence on HDS ( $\beta = -.44$ , as compared with  $\beta = -.09$  for the BC measure). Thus, as proportion increased, HDS

**Table 2** Signal detection results with bimodality measures using simulated data

Measure	$d'$	Hit	Miss	False Alarm	Correct Rejection
BC	1.46	64 %	36 %	14 %	86 %
HDS	3.43	65 %	35 %	0 %	100 %
$AIC_{diff} > 0$	0.74	95 %	5 %	81 %	19 %

reliably decreased, and the required distance to detect bimodality reliably decreased as well. This finding makes sense given the reliance of HDS upon distribution differences, because a higher density of shifted observations would induce higher maximum differences in this range of the distribution.

As with the BC measure, the distance threshold for detecting bimodality changed as a function of proportion. In zero-skew distributions, a low proportion of 5 % required a distance of roughly 7 *SDs* to detect bimodality, and this declined to roughly 5 *SDs* with the higher proportion of 10 %. The distance threshold for detecting bimodality decreased gradually thereafter, with a proportion of 25 % requiring a distance of 4 *SDs*, and a proportion of 50 % requiring a distance of 3 *SDs*. In highly positively skewed distributions, this same pattern obtained as well, except that the presence of positive skew overall lowered the distance threshold for detecting bimodality across all levels of proportion. Thus, for example, whereas a proportion of 10 % required a distance of 5 *SDs* in zero-skew distributions, it required only a distance of 3 *SDs* in highly positively skewed distributions.

#### *Akaike's information criterion difference ( $AIC_{diff}$ )*

$AIC_{diff}$  values range from  $-1$  to  $1$ , with positive values suggesting bimodality, and negative values suggesting unimodality. As with the BC and HDS measures, distance had a considerably larger influence ( $\beta = .80$ ) than either proportion ( $\beta = -.20$ ) or skew ( $\beta = .35$ ). As with the other measures, size again appeared to have a negligible influence ( $\beta = .01$ ). The influence of proportion was relatively strong, but in the unexpected direction, with increases in proportion overall leading to decreases in  $AIC_{diff}$  and leading to the recognition of distributions as unimodal. Most striking was that the  $AIC_{diff}$  measure recognized nearly all distributions as bimodal, except for zero-skew distributions with a distance of 1 *SD*. Thus, the measure is extremely liberal, recognizing bimodality with even the lowest proportions and smallest distances of Mode 2 responses. At least one reason for this, it appeared, is that the two-component model does not converge after 100 iterations in cases with minimal bimodality (i.e., low distance values). This may have rendered the AIC estimation in the two-component fit unreliable, and in many cases it resulted in an  $AIC_{diff}$  that was positive (i.e., lower AIC for the bimodal model).

#### *Influence of sample size*

As described above, the size parameter had the weakest influence across the BC ( $\beta = -.01$ ), HDS ( $\beta = -.07$ ), and  $AIC_{diff}$  ( $\beta = .01$ ) measures. That said, it did have some minor but interesting interactions with the other parameters on bimodality detection (see Table 1). For the BC measure,

size had a negligible influence for distributions with low skew (e.g., skew exponent = 1). For distributions with high skew (e.g., skew exponent = 5), however, a smaller sample size made distributions with a large distance but small proportion more likely to be detected as bimodal. For the HDS measure, the lowering of sample size made bimodality detection overall more conservative. No substantive differences by size were found for distributions with low skew (skew exponent = 1). However, when distributions contained high skew (skew exponent = 5), those with a large proportion but small distance were less readily detected as bimodal with small samples (size = 250). For the  $AIC_{diff}$  measure, there were no major differences across sample sizes. Overall, bimodality detection became more conservative with a smaller sample size, but, given how liberal the measure is, nearly all distributions were recognized as bimodal regardless. One notable result was that in distributions with low skew (skew exponent = 1) and small distance, the lowering of sample size decreased bimodality detection, whereas this was not the case in distributions with high skew (skew exponent = 5). Taken together, across the three measures, sample size had the weakest influence on bimodality detection and introduced relatively negligible effects.

#### Discussion

The analyses of our simulated data suggest that the HDS measure may be the most robust measure for detecting bimodality. It had the highest sensitivity ( $d'$ ) in distinguishing bimodality from unimodality, and it also was immune to some of the interactive effects with skew that plagued the BC measure. Thus far, HDS appears to be an optimal choice. Whereas analyses of simulated data afforded us control and comprehensiveness, we also sought to validate their applicability. Thus, we examined the performance of the measures in experimental data for which the modality of particular distributions was theoretically known.

#### Experimental data

To this end, we made use of the previously published studies in Freeman et al. (2008). For Studies 1 and 2 in that work, the researchers asked participants to categorize the sex of 20 sex-typical and 20 sex-atypical faces using a mouse-tracking paradigm. Thus, on every trial the participants moved the mouse from the bottom center of the screen to a “male” or “female” response in the top-left or the top-right corner. Study 1 was based on real faces, and Study 2 on computer-generated faces. We would expect trajectories for the sex-typical faces to be relatively direct, with little curvature (e.g., as is seen with the Mode 1 responses in the

bottom panels of Fig. 1). In Study 3, participants were presented with the combined 40 sex-typical faces of Studies 1 and 2. This time, however, as soon as the participants initiated a movement, on half of the trials the “male” and “female” response buttons suddenly changed color and switched sides. On the other half of the trials, the responses remained constant. For the “switch” trials, therefore, we would expect the trajectories to have extreme curvature, in the form of discrete-like errors, when the participants initially pursued one alternative and then redirected in mid-flight toward the other alternative (e.g., as is seen with the Mode 2 responses in the bottom panels of Fig. 1). For the normal trials, we would expect trajectories to be direct and to exhibit low curvature.

Thus, when averaging the degrees of curvature across all trajectories for sex-typical faces in Studies 1 and 2, we would expect a unimodal distribution in which all trajectories exhibited low curvature. When averaging the degrees of curvature across all trajectories for sex-typical faces in Study 3, however, we would expect a bimodal distribution. This would occur because on half of the trials participants’ trajectories would exhibit extreme curvature (“switch” trials), and on the other half their trajectories would exhibit low curvature. As such, the trajectory data of Freeman et al.’s (2008) studies afford us two distributions, one theoretically known to be unimodal and the other theoretically known to be bimodal. Here, we determined how the three bimodality measures would perform in distinguishing the modality of these distributions.

## Method

Study 1 involved 23 participants, each categorizing 20 sex-typical faces, with eight excluded trials. Study 2 involved 25 participants each categorizing 20 sex-typical faces, with 12 excluded trials. This resulted in a total of 940 trials for Studies 1 and 2, forming the theoretical unimodal distribution. Study 3 involved 21 participants each categorizing 40 sex-typical faces, with 27 excluded trials. Half of these trials were “switch” trials, and half were control trials. This resulted in a total of 813 trials for Study 3, forming the theoretical bimodal distribution. Note that the stimuli from which the two distributions are derived were identical.

## Results

All trajectories were fit to 100 normalized time bins (101 time steps) using linear interpolation and were rescaled to a standard coordinate space. A measure of trajectory curvature, area under the curve (AUC), was computed for every trial; this was the geometric area between the observed trajectory and an idealized response trajectory (a straight

line between the trajectory’s start- and endpoints). Any curvature heading away from the opposite-category response was computed as negative area (and summed with any positive area; see Freeman & Ambady, 2010).

Using the  $z$  distributions of the trajectories’ AUC values, the theoretical unimodal distribution ( $n = 940$ , skewness = 0.20, kurtosis =  $-0.79$ ) was recognized as unimodal by BC ( $b = .469$ ) and HDS ( $p = .12$ ), but not by  $AIC_{diff}$  (.04), which, as we noted, tends to overestimate bimodality. Conversely, the theoretical bimodal distribution ( $n = 813$ , skewness = 0.13, kurtosis =  $-1.35$ ) was recognized as bimodal by BC ( $b = .614$ ), HDS ( $p < .0001$ ), and  $AIC_{diff}$  (.16). Thus, two of the three measures successfully distinguished the experimental unimodal and bimodal distributions.

## Discussion

These results demonstrate that two of the bimodality measures were able to successfully distinguish unimodal and bimodal distributions in a relevant experimental context, such as the mouse-tracking paradigm. This increases the generality of the measures and indicates that they are able to be applied to experimental data.

## General discussion

The results of our simulations revealed a number of important divergences between the BC, HDS, and  $AIC_{diff}$  measures in their sensitivity to bimodality, as well as a number of convergences. Across all measures, the distance between Mode 1 and Mode 2 populations had a considerably stronger influence on bimodality detection than did proportion or skew, with increases in distance leading to increases in bimodality. The proportion of Mode 2 responses differed in its influence, with a main effect of proportion quite weak in the BC simulations and quite strong in the HDS simulations. A main effect of proportion was also relatively strong in the  $AIC_{diff}$  simulations, but it did not bear a consequential effect on distinguishing between unimodality versus bimodality because virtually all distributions were recognized as bimodal. The introduction of positive skew had considerable influences on bimodality detection, but in interestingly different ways between the measures, which are discussed later. Finally, sample size had a relatively weak influence on all three measures.

Overall, in part on the basis of these influences, HDS had the highest sensitivity ( $d'$ ) in distinguishing unimodality from bimodality, with BC coming in second, and  $AIC_{diff}$  last (see Table 2). All three measures, however, did successfully identify an experimental distribution theoretically known to be bimodal as being bimodal. The HDS and BC

measures (but not  $AIC_{diff}$ ) also successfully identified an experimental distribution theoretically known to be unimodal as being unimodal. This is encouraging, and suggests that both the HDS and BC measures may, in some cases, be successfully applied to experimental data. However, bear in mind that the two experimental distributions examined here were limited cases. They provide promise for the applicability and generality of the measures, but represent two points in a quite large parameter space explored more systematically with the simulations. Future research should consider examining the applicability of the measures to a wider range of experimental distributions. For the time being, we find that the measures are able to be successfully applied in some cases, but the more comprehensive results of the simulations suggest that the HDS measure may be most successful with the largest possible set of distributions. Thus, both HDS and BC generally appear to be applicable to a relevant experimental context such as the mouse-tracking paradigm. But, more importantly, the simulation results suggest that HDS would be a more sensitive and accurate measure, when considering the full parameter space that exists across researchers' experiments.

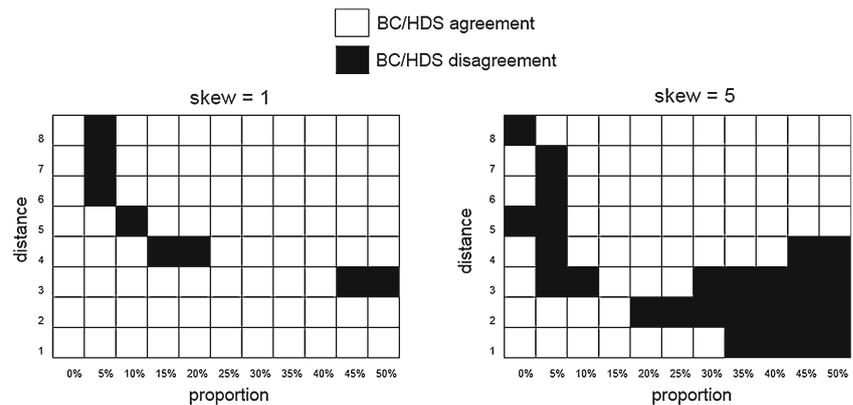
The BC and HDS measures were generally convergent, recognizing bimodality and unimodality in similar distributions, whereas the  $AIC_{diff}$  measure behaved quite differently. The  $AIC_{diff}$  measure was extremely liberal at even the lowest proportions and smallest distances of Mode 2 responses, recognizing bimodality in virtually all distributions. Because the BC and HDS measures operated similarly, we directly compared these two measures in further detail. We specifically focused on distributions with large sample size ( $n = 2,000$ ). For both measures, distance had a strong main effect, with larger distances leading to more bimodality. The main effects of proportion, however, differed considerably. As expected, the HDS measure was strongly influenced by proportion, with increases in proportion resulting in decreases in HDS and the lowering of the required distance to detect bimodality. The BC measure, however, was negligibly influenced by a main effect of proportion. Instead, in BC simulations (but not HDS simulations), skew interacted with proportion in such a way that, in highly positively skewed distributions, increases in proportion resulted in decreases in BC and the raising, rather than the lowering, of the distance required to detect bimodality. As discussed earlier, this may have been due to a "blending" of the elongated positive tail of the Mode 1 population into the Mode 2 population. The HDS measure was immune to this anomalous interactive effect.

In fact, when we examine the BC and HDS measures more closely, we can see why this was the case. The BC relies on the intuition that bimodality will involve a potential

increase in asymmetry (skew, along with a drop in kurtosis). Thus, increasing skew even in a completely unimodal context increases BC consistently, and with extreme skew it can induce spurious inferences of two modes. While HDS did appear to have a main effect of skew, when we consider the way in which the measure is calculated, an interesting pattern falls out (for a full description of HDS, see Henderson, Parmeter, & Russell, 2008). The HDS operates by judging the degree to which the theoretical uniform distribution (the "asymptotic unimodal" case, with a single in-the-limit inflection point) strays from the observed distribution. The measure does so by looking to the convex and concave portions of the observed distribution, and it finds the uniform distribution that minimizes the maximum distance between the uniform and observed distributions. By obtaining a sampling distribution of a dip statistic (based on those distances), it determines the extent to which the observed case strays. This means that any distribution with a single inflection point is very unlikely to be assessed as bimodal. This is the case even with extreme skew: As BC increases with skew in a unimodal case (0 % proportion), HDS does not approach  $p < .05$  at all; in fact, increasing skew *increases* the  $p$  value, reducing the probability of inferring bimodality. The relationship between HDS and skew derives from proportions  $> 0$  %. As noted earlier, as distance and proportion increase, skew only increases the probability of assessing bimodality (put simply, it helps HDS). The reason for this seems to derive from this brief description of HDS: The increase in skew may lead to a sharp increase in the distance along the concave/convex portions of the observed distribution, thus increasing the dip statistic and drawing it away from the sampling distribution based on the uniform.

These basic theoretical differences help frame other results from the simulations. To further compare the BC and HDS measures, Fig. 4 depicts the distributions for which the BC and HDS measures were in agreement or disagreement. The measures had 91 % agreement for zero-skew distributions, and 71 % agreement for highly positively skewed distributions. Regardless of skew, disagreements were found in two areas: low proportions of medium-to-high distance and high proportions of low distance. Positive skew exacerbated the amount of disagreement. In disagreements among low proportions of medium-to-high distances, HDS recognized the distributions as unimodal, whereas BC recognized them as bimodal. In disagreements among high proportions of low distances, HDS recognized the distributions as bimodal, whereas BC recognized them as unimodal. These disagreements are mainly attributable to the differential influences of the proportion parameter on BC and HDS measures, such that the HDS measure was considerably more calibrated to proportion. For example, a highly positively skewed distribution of only 5 % proportion was

**Fig. 4** Grids showing the simulations in which the BC and HDS measures agreed in bimodality detection (white cells) and simulations in which the measures disagreed (black cells), for simulations in which size=2,000



recognized as bimodal by the BC measure with a surprisingly low minimum distance of 4 *SDs*, whereas the required distance for the HDS measure to detect bimodality for a 5 % proportion was considerably higher: 7 *SDs*. This accounts for the disagreements among low proportions of medium-to-high distances. As for the other type of disagreements, involving high proportions of low distances, these are accounted for by the anomalous interactive effect of skew and high levels of proportion with the BC measure, discussed earlier. As proportion increased, BC decreased (rather than increased), especially in highly positively skewed distributions. Because HDS was immune to these effects, the two measures disagreed with high proportions of low distances.

On the basis of our simulation results, we are inclined to recommend the use of HDS. This is mainly because of its simultaneous and appropriate calibration to both the distance and proportion of a dual response population (Mode 2). Although the HDS measure is biased by positive skew, with greater positive skew overall increasing bimodality detection (as occurred with the other measures, as well), skew had minimal interactive influences with other parameters, and therefore did not “warp” the results as it did with other measures. For example, in the BC simulations, skew interacted with proportion such that higher proportions led to a raising, rather than a lowering, of the threshold for detecting bimodality, whereas the HDS measure was immune to this anomalous effect. Furthermore, because the BC measure was not especially calibrated to proportion, even the lowest proportions with relatively minimal distances were recognized as bimodal. Thus, overall, the HDS was a more robust measure.

As for the  $AIC_{diff}$  measure, it proved to be extremely sensitive to even the most minimal introduction of bimodal features (the lowest levels of proportion and distance), but was biased (as it tended not to recognize unimodality), and thus its utility for distinguishing between unimodality versus bimodality suffered. However, if a researcher were interested in detecting even the faintest of bimodal features, the  $AIC_{diff}$  measure might be advisable. Previous

discussions of the AIC measure and mixture models had argued against the measure being used for assessing bimodality due to computational expense (Hartigan & Hartigan, 1985), although nowadays the ready availability of libraries permits their application easily. Unfortunately, it is not clear how one could infer confidence on the basis of AIC differences. Most often, a discrete selection heuristic is used, although other strategies may be available to characterize these AIC differences (Wagenmakers & Farrell, 2004).

One concern in conducting distributional analyses is outlier screening. Typically, researchers use a standard criterion somewhere between 2 and 4 *SDs* to eliminate outliers (Van Selst & Jolicœur, 1994). An initial issue in outlier screening is researchers’ distributional assumptions. If researchers do not anticipate a bimodal or non-Gaussian distribution and do not explore distributional characteristics, they may eliminate a dual-response population (Mode 2) altogether, and wrongly consider a meaningful, separate population of responses to be spurious. A more complicated issue is the fine line between what could be considered a few outliers versus what could be considered an actual dual-response population (Mode 2). For example, consider the theoretical bimodal distribution of mouse trajectories’ attraction toward the incorrect response from Freeman et al. (2008), composed of approximately 50 % Mode 1 trajectories and 50 % Mode 2 trajectories, as can be seen in Fig. 1. This bimodal distribution therefore had an approximate proportion of 50 % and a distance typical of the two-choice mouse-tracking paradigm. In that study, the Mode 1 and Mode 2 populations had a distance of approximately 3 *SDs*. Thus, if a researcher were to eliminate outliers exceeding a standard criterion somewhere between 2 and 4 *SDs*, a hefty portion of the total trials would have been eliminated, and the Mode 2 population would have been considerably truncated. The large portion of eliminated trials might alarm the researcher, leading to a closer inspection of distributional characteristics. However, a Mode 2 proportion of only 10 % or 15 % that was shifted approximately 3 to 4 *SDs* away from the Mode 1 population could easily pass as standard outliers without alarm and be eliminated. In such a case, such shifted responses could

genuinely reflect a dual-response population (Mode 2) or, alternatively, outliers.

In such cases, the decision between a dual-response population versus outliers may be difficult and would require careful consideration of theoretical assumptions. If shifted responses were considered outliers but in fact reflected a genuine dual-response population, researchers' interpretation of the data would greatly suffer by altogether overlooking the presence of a dual cognitive process. On the other hand, if the shifted responses were considered genuine data (a Mode 2 population), but in reality were mere outliers, the BC and HDS measures might spuriously recognize the distribution as bimodal, so long as it was also positively skewed. Even with a Mode 2 proportion of only 5 %, the BC measure would recognize bimodality with the same distance (approximately 3–4 *SDs*), so long as the distribution was positively skewed. The HDS measure would not, however, as it was more appropriately calibrated to the proportion parameter. In short, researchers would benefit from carefully considering whether a collection of shifted responses reflects a genuine dual-response population (Mode 2) or spurious outliers, especially in low proportions.

### Limitations

It is important to note the limitations of the present work. First, we have focused here on bimodality estimations in an “out-of-the-box” style by using readily available methods. Researchers interested in detailed density estimation techniques may go beyond bimodality and into the domain of fitting complex mixture models (for some exploration of the mathematical properties of these measures, see Hartigan & Hartigan, 1985; Minnotte, 1997; Silverman, 1981; and in other scientific domains, Hellwig et al., 2010; Milligan & Cooper, 1985). We hope that the estimations described here are helpful for researchers and are able to be used as general-purpose techniques to distinguish unimodal and bimodal distributions in commonly observed experimental data. However, some specific cases may warrant more sophisticated estimation techniques that go beyond the scope of this work.

More generally, bimodality is a complex phenomenon, and the way that two tandem processes might affect distributional characteristics in psychological experiments is not always easy to discern. Bimodality may thus potentially manifest in ways that are not captured by the estimations described in the present work. Moreover, we have focused on the limited case of one, albeit highly popular, theoretical view on multiple cognitive processes: dual-process accounts. Of course, other views argue for the existence of three or more cognitive processes, which in some cases might lead to theoretical predictions of multimodality involving more than two modes. For instance, a popular

model of implicit social cognition is the quadruple-process model, which posits four distinct processes that contribute to overt responses in certain implicit measures: association activation, discriminability, overcoming bias, and guessing (Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005). In some cases, distributions of four modes may therefore be predicted. Fortunately, the HDS and  $AIC_{diff}$  measures are designed to detect multimodality of more than two modes, but the BC measure is not. And, of course, the very interpretations of these measures may be called into question. Some researchers may proffer the interpretation that any such hypothetical “processes” are in fact different modes of operation of one general-purpose cognitive system. Future research will need to examine how the measures used here fare with multimodal distributions involving more than two modes, and how to interpret those processes identified or proposed.

Another point is that our simulations involved a range of sample sizes that was relatively large, closer to the number of observations that would originate from an aggregated group of participants rather than from individual-participant data. Indeed, researchers have often estimated bimodality in such aggregate data sets, and both benefits and costs are involved when examining distributional characteristics in group rather than individual participant data (see, e.g., Vincentizing methods: Ratcliff, 1979). Nevertheless, the present results should hold true for any distribution, whether from aggregated data or not. One limitation, however, is that our smallest distributions contained only 250 observations, and many experimental tasks involve fewer than 250 trials per participant. That said, we found that sample size (the “size” parameter) had quite minimal influences on bimodality detection. For this reason, it is likely that the same pattern of results would obtain for smaller samples as well (e.g., 50 observations) that would be closer to individual-participant data sets. Future work would benefit from examining bimodality detection at a wider range of sample sizes and from directly comparing the results when using aggregate versus individual-participant data.

### Conclusion

In summary, the analyses presented here offer a comparison among oft-used bimodality measures that may be added to the methodological and statistical toolbox of the psychological sciences. By utilizing basic tests for bimodality, researchers may be able to use straightforward methods to readily test hypotheses about a sample. In fact, some experiments may very well employ these distributional characteristics as empirical patterns to be predicted and tested in behavioral experiments (e.g., Dale & Duran, 2011; Freeman et al., 2008). In these cases, the inference of bimodality may be a central feature of developing theory. A given debate,

such as single- versus dual-process models, could accumulate data across multiple studies. The significance of even weak bimodality, if significant at all, should be reproducibly detected by the techniques that we have described here.

Distinguishing between single-process and dual-process phenomena has only intensified with the recent advent of more continuous, temporally fine-grained measures that track cognition in real time (Freeman & Ambady, 2010; Spivey & Dale, 2006; Spivey et al., 2005). Whether using such measures, reaction times, or some other behavioral measure, analysis of a response distribution's modality has been crucial in detecting the presence of a dual process, as dual processes tend to introduce bimodal features. Although these facts have been known for at least three decades (e.g., Atkinson & Juola, 1973; Ratcliff, 1979), rarely have

measures of bimodality been systematically evaluated and compared. In our simulations of the BC, HDS, and  $AIC_{diff}$  measures of bimodality, we have found that the HDS may be overall a more appropriate measure, although this conclusion is qualified by several points discussed above. We hope that this work proves useful for future research seeking to distinguish between single-process and dual-process cognitive phenomena, and more generally for calling greater attention to the response distribution in understanding cognition.

**Author note** This work was supported in part by a National Institutes of Health NRSA fellowship (F31-MH092000) to J.B.F. and by a National Science Foundation research grant (BCS-0826825) to R.D. We thank Nalini Ambady for helpful support.

---

## Appendix: MATLAB code for BC, $AIC_{diff}$ , and HDS

In all of the code below, **x** is a vector containing a single sample.

### *BC*

---

```
function [b] = bmtest(x)
    %m3 = skew
    %m4 = kurt
    %n = data size
    m3 = skew(x);
    m4 = kurtosis(x);
    n = length(x);
    b=(m3^2+1) / (m4 + 3 * ( (n-1)^2 / ((n-2)*(n-3)) ));
```

### *AIC<sub>diff</sub>*

---

```
f1 = gmdistribution.fit(x,1); % single component, unimodal
f2 = gmdistribution.fit(x,2); % two components, bimodal
AICd = (f1.AIC-f2.AIC)/max([f1.AIC,f2.AIC])
```

### *HDS*

---

```
% Utilizing function by F. Mechler, 2002 translated into MATLAB, found at:
% http://www.nicprice.net/diptest/
```

```
[dip,p_value,xlow,xup] = HartigansDipSignifTest(x,500);
```

---

## References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, AC-19*, 716–723. doi:10.1109/TAC.1974.1100705
- Atkinson, R. C., & Juola, J. F. (1973). Factors influencing speed and accuracy of word recognition. In S. Kornblum (Ed.), *Attention and performance IV* (pp. 583–612). New York, NY: Academic Press.
- Conroy, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. J. (2005). Separating multiple processes in implicit

- social cognition: The quad model of implicit task performance. *Journal of Personality and Social Psychology*, *89*, 469–487.
- Dale, R., & Duran, N. D. (2011). The cognitive dynamics of negated sentence verification. *Cognitive Science*, *35*, 983–996.
- Dale, R., Kehoe, C., & Spivey, M. J. (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Memory & Cognition*, *35*, 15–28.
- Dale, R., Roche, J., Snyder, K., & McCall, R. (2008). Exploring action dynamics as an index of paired-associate learning. *PLoS One*, *3*, e1728. doi:10.1371/journal.pone.0001728
- Devine, P. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*, 5–18.
- Evans, J. St. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*, 255–278.
- Farmer, T. A., Anderson, S. E., & Spivey, M. J. (2007). Gradiency and visual context in syntactic garden-paths. *Journal of Memory and Language*, *57*, 570–595.
- Freeman, J. B., & Ambady, N. (2009). Motions of the hand expose the partial and parallel activation of stereotypes. *Psychological Science*, *20*, 1183–1188.
- Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, *42*, 226–241. doi:10.3758/BRM.42.1.226
- Freeman, J. B., & Ambady, N. (2011a). A dynamic interactive theory of person construal. *Psychological Review*, *118*, 247–279. doi:10.1037/a0022327
- Freeman, J. B., & Ambady, N. (2011b). When two become one: Temporally dynamic integration of the face and voice. *Journal of Experimental Social Psychology*, *47*, 259–263.
- Freeman, J. B., Ambady, N., Rule, N. O., & Johnson, K. L. (2008). Will a category cue attract you? Motor output reveals dynamic competition across person construal. *Journal of Experimental Psychology: General*, *137*, 673–690. doi:10.1037/a0013875
- Freeman, J. B., Pauker, K., Apfelbaum, E. P., & Ambady, N. (2010). Continuous dynamics in the real-time perception of race. *Journal of Experimental Social Psychology*, *46*, 179–185.
- Hartigan, J. A., & Hartigan, P. M. (1985). The dip test of unimodality. *The Annals of Statistics*, *13*, 70–84.
- Hellwig, B., Hengstler, J. G., Schmidt, M., Gehrman, M. C., Schormann, W., & Rahnenfuhrer, J. (2010). Comparison of scores for bimodality of gene expression distributions and genome-wide evaluation of the prognostic relevance of high-scoring genes. *BMC Bioinformatics*, *11*, 276.
- Henderson, D. J., Parmeter, C. F., & Russell, R. R. (2008). Modes, weighted modes, and calibrated modes: Evidence of clustering using modality tests. *Journal of Applied Econometrics*, *23*, 607–638.
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513–541. doi:10.1016/0749-596X(91)90025-F
- Keren, G., & Schul, Y. (2010). Two is not always better than one: A critical evaluation of two-system theories. *Perspectives on Psychological Science*, *4*, 533–550.
- McKinstry, C., Dale, R., & Spivey, M. J. (2008). Action dynamics reveal parallel competition in decision making. *Psychological Science*, *19*, 22–24.
- McLachlan, G., & Peel, D. (2000). *Finite mixture models*. Hoboken, NJ: Wiley.
- Mechler, F. (2002). A direct translation into MATLAB from the original FORTRAN code of Hartigan's Subroutine DIPTST algorithm. Retrieved from [www.nicprice.net/diptest](http://www.nicprice.net/diptest)
- Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, *50*, 159–179.
- Minnotte, M. C. (1997). Nonparametric testing of the existence of modes. *The Annals of Statistics*, *25*, 1646–1660.
- Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, *86*, 446–461. doi:10.1037/0033-2909.86.3.446
- SAS Institute Inc. (1989). *SAS/STAT user's guide*. Cary, NC: Author.
- Silverman, B. W. (1981). Using kernel density estimates to investigate multimodality. *Journal of the Royal Statistical Society: Series B*, *43*, 97–99.
- Song, J. H., & Nakayama, K. (2006). Role of focal attention on latencies and trajectories of visually guided manual pointing. *Journal of Vision*, *6*(9), 11:982–995. doi:10.1167/6.9.11
- Song, J. H., & Nakayama, K. (2008). Target selection in visual search as revealed by movement trajectories. *Vision Research*, *48*, 853–861.
- Spivey, M. J., & Dale, R. (2006). Continuous dynamics in real-time cognition. *Current Directions in Psychological Science*, *15*, 207–211.
- Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences*, *102*, 10393–10398.
- van Gompel, R. P. G., Pickering, M. J., Pearson, J., & Liversedge, S. P. (2005). Evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language*, *52*, 284–307.
- van Gompel, R. P. G., Pickering, M. J., & Traxler, M. J. (2001). Reanalysis in sentence processing: Evidence against current constraint-based and two-stage models. *Journal of Memory and Language*, *45*, 225–258.
- Van Selst, M., & Jolicœur, P. (1994). A solution to the effect of sample size on outlier elimination. *Quarterly Journal of Experimental Psychology*, *47A*, 631–650. doi:10.1080/14640749408401131
- Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review*, *7*, 424–465. doi:10.3758/BF03214357
- Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, *11*, 192–196. doi:10.3758/BF03206482