Evolutionary Ecology Research, 2014, **16**: 179–194

The limitations of inferring decision rule use from individuals' sampling behaviour: a computational test of old and new algorithms

Bart J. Kensinger and Barney Luttbeg

Department of Zoology, Oklahoma State University, Stillwater, Oklahoma, USA

ABSTRACT

Background: The foraging and mate choices of individuals depend on how individuals gather information about available options and use that information to make decisions. Several alternative decision rules have been proposed, but little progress has been made in using observations of behaviour to discover what decision rule is being used. An important approach has been to combine the rules into an algorithm that chooses between them ('the prevailing algorithm').

Question: How accurately does the prevailing algorithm classify which decision rule is most likely to produce observed sequences of individuals choosing one item from a pool of many? Can we create better algorithms?

Methods: Using computer simulations, we construct sampling sequences that follow the threshold, best-of-*n*, and comparative Bayes rules. We vary sampling cost, pool size, signal variance, and population variance. We mimic sequences seen in nature and avoid parameter values in which individuals sample infinitely or not at all. We apply the prevailing algorithm to determine which rule created each sequence, and measure its probability of success. We use a recursive partitioning exercise to produce two new algorithms: one based on individuals' sampling sequences, and one based on the population's averaged metrics. We compare the performances of these two new algorithms and the prevailing algorithm. Finally, we apply all three algorithms to previously published empirical data sets.

Conclusions: The prevailing algorithm did no better than a random classification, but the individual algorithm had 64% success and the population algorithm had 97% success. The latter also succeeded in identifying the predominant rule even if different individuals of the group had different rules. The population algorithm can reliably infer what rule the majority of a population is using. However, using observed sequences to infer the decision rule of an individual is unreliable.

Keywords: decision rules, mate choice, foraging.

Correspondence: B.J. Kensinger, Department of Zoology, Oklahoma State University, Stillwater, OK 74078, USA. e-mail: bart.kensinger@okstate.edu

INTRODUCTION

Individuals often have to choose an item from a set of available items. These items may be mates, social groups, territories, or foraging patches that differ in risk and reward. In each case, the individual can improve its expected fitness by choosing a good item, but this is limited by uncertainty about the quality of items and the costs of gathering information about the quality of the items. The processes (decision rules) that individuals use to make these decisions can shape what items they choose, thus shaping the dynamics of larger systems, such as mating systems (Gibson and Langen, 1996) and predator–prey systems (Luttbeg and Schmitz, 2000; Luttbeg and Trussell, 2013). In addition, the processes that individuals use to make decisions may be constrained by their memory and cognitive abilities (Ryan *et al.*, 2009). It is because of these higher-order impacts on larger systems that behavioural ecologists have been interested in understanding how individuals make decisions.

Several alternative decision rules have been proposed (Parker, 1978, 1979, 1983; Wittenberger, 1983; Dombrovsky and Perrin, 1994; Collins *et al.*, 2006; Castellano *et al.*, 2012). They differ in assumptions about whether individuals track the estimated quality of individual objects or the population distribution of qualities, and whether individuals randomly or selectively choose what items they sample for information. We have chosen to focus on the best-of-*n* (Janetos, 1980), flexible thresholds (Real, 1990), and comparative Bayes (Luttbeg, 1996) rules, in line with previous investigators (Uy *et al.*, 2001; Mabry and Stamps, 2008; Selonen and Hanski, 2010). The best-of-*n* rule states that individuals should randomly sample *n* items and then return to and choose the item that appeared to have the highest quality. The threshold rule states that individuals should randomly sample items and choose an item when it appears to exceed a predetermined threshold of quality. We use a threshold rule in which the threshold declines with time. The comparative Bayes rule states that individuals should use their estimates of the quality of each potential item to choose whether to sample an item and if so, which item, or else to choose the item that they currently estimate to be the best.

Efforts have been made to determine which proposed decision rule best depicts the rule that individuals are using. One proposed approach has been to manipulate the variance or mean of the quality of items from which an individual is choosing, and observe if individuals respond in ways that the alternative rules predict (Beckers and Wagner, 2011; Izzo and Gray, 2011; Wiegmann *et al.*, 2013). This could be a powerful approach if individuals used the rules as exactly formulated. However, the approach is highly sensitive to the assumptions inherent in the rules, and slight changes in assumptions can cause the predicted behaviours of individuals using different decision rules to converge. In addition, these experiments usually require laboratory conditions to control carefully the manipulations; thus they are limited to the few species that can be studied in the lab at semi-natural spatial scales. Other complications with this approach are that animals may behave differently in the lab, and that mate-searching tactics are plastic and can change with previous experience (Tuckerman *et al.*, 1993; Fiske and Kalas, 1995), possibly affecting experiments with repeated measures.

A second approach has been to infer what decision rule an individual is using based upon its observed sampling sequence before choosing an item. Several studies pioneered this approach (Arak, 1988; Trail and Adams, 1989), but the first attempt to distinguish between the three decision rules that we are studying was done with the female mate choice behaviour of satin bowerbirds, *Ptilonorhynchus violaceus* (Uy *et al.*, 2001). Uy and colleagues used observed sampling behaviour to infer what decision rule best explained an observed sequence of potential mates visited by a female.

Mabry and Stamps (2008) and Selonen and Hanski (2010) combined the rules into the prevailing algorithm as follows. If the proportion of observed returns is >0.05, the algorithm classifies the sequence as being produced by the comparative Bayes rule (Fig. 1A). If the proportion of observed returns is ≤0.05, and the chosen item is either the last sample and sampled only once, or not sampled at all, then the algorithm classifies the sequence as being produced by the threshold rule. The remaining sequences are classified as being produced by the best-of-*n* rule. This prevailing algorithm, as far as we know, has not yet been tested to see how effectively it identifies the decision rule an individual is using. Thus, in the first part of this paper we expose the prevailing algorithm to sequences produced by known decision rules to examine how often the prevailing algorithm correctly identifies the rule that made the sequence.

We find that the prevailing algorithm does not have a high success rate. Thus in the second part of the study we create two new classification algorithms. The first algorithm uses individual sequence data, and the second algorithm uses the average of sampling metrics from a population of individuals. We compare the performances of these two new algorithms and the prevailing algorithm, and highlight the superior performance of the population-based algorithm. Finally, we apply all three algorithms to previously published empirical data sets. This exercise illustrates the importance of testing algorithms before their application since the conclusions about populations' rule use change.

METHODS

We assumed that individuals attempt to maximize the net benefit they receive from their choice of an item, which is the benefit received by the chosen item minus accrued sampling costs. Individuals are choosing an item from a pool (p) of available items. These items could be mates, foraging patches, territories, or anything that is chosen once and affects the fitness of the chooser. We do not include the chance of an item being taken by another individual and precluding our focal individual from choosing that item. The quality of each item within this pool is drawn from a normal distribution with an arbitrary mean of 30 and a population variance (r^2). Each time the individual samples an item it pays a sampling cost (c) and receives a cue indicating the item's quality. These costs accrue linearly as would energy or time spent visiting items. The cue is drawn from a normal distribution with the mean being the item's actual quality, and a signal variance (s^2) – used only by the comparative Bayes model.

We specified that individuals could conduct only 99 samples after which an item must be chosen. We fix this sample limit, because with finite items and infinite time the threshold rule can get trapped into searching for a non-existent item above the threshold (Luttbeg, 2002). The time limit has a negligible effect on our conclusions because it is reached in less than 1% of our simulated sequences.

The *best-of-n rule* [sometimes referred to as the pooled-comparison rule (Janetos, 1980; Real, 1990)] states that an individual should sample a fixed number (n) of randomly selected items and then return to and choose the item with the highest sampled quality (Janetos, 1980). Following the common formulation of the best-of-*n* rule (Janetos, 1980; Real, 1990), we assumed that items are randomly encountered, that individuals behave as if they know the distribution of item quality, and that they receive perfect information about the quality of items.

The optimal level of n depends on sampling costs and the variance in the distribution of item quality, but is independent of the mean of the distribution of item quality. Optimal n's were found using normal order statistics (Luttbeg, 2002).

The *threshold rule* states that an individual should set a threshold, sample randomly, and choose the first item that appears to have a quality exceeding the threshold. The optimal threshold depends on the mean and variance of the distribution of item quality, sampling costs, and time constraints. Again, like the best-of-n rule, we follow the common formulation of the rule (Real, 1990) and assume that items are encountered randomly and that individuals gain perfect information about the quality of encountered items. We found optimal thresholds using a dynamic state variable model (Luttbeg, 2002; Luttbeg and Langen, 2004).

The *comparative Bayes rule* states that an individual should sample the item that maximizes their expected future fitness and, when the cost of sampling exceeds the expected fitness benefit of sampling, the individual chooses the item with the highest associated expected fitness (Luttbeg, 1996). Key differences between this rule and the other two rules are that individuals gain only imperfect information about the quality of items, and that they can choose which items to sample based upon their estimates of the quality of those items and their level of certainty about that estimate. The rule assumes that individuals have prior estimates of the quality of each available item and that these prior estimates are normally distributed. We determined optimal sampling and choice decisions given the comparative Bayes decision rule (Luttbeg, 1996, 2002).

Metrics of sequences

For our computer-model-generated sequences, we use the terminology established in previous work (Gibson and Langen, 1996; Luttbeg and Langen, 2004). In a sequence ABAACA (where an individual has encountered item A, item B, item A again, and so on), an individual has sampled 3 *items* and conducted 5 *samples*. The last *item* in a sequence is the identity of the *chosen item* and it is not considered one of the *samples*. A *repeat* is when an individual samples the same item consecutively, such as AA. A *switch* is when an individual moves from one item to another, such as AB. A *return* is a switch in which the individual samples an item that has been sampled earlier in the sequence, such as ABCAD. ABCBD is a special case of *return*, called *return to previous*, in which the individual returns to the previous item. For several of the metrics we also included an adjusted value. For instance, the adjusted proportion of returns was the number of returns divided by (samples − 1). This is because the first sample cannot be a return. We included the normal and adjusted versions in the metrics for algorithm development. Table 1 summarizes the sequence metrics.

Testing the algorithm

To test the algorithm, we programmed the three decision rules. We used the C programming language and the Xcode compiler (see evolutionary-ecology.com/data/2880Appendix A). The sampling and choice behaviour produced by each of the three decision rules depend on the size of the pool of available items; the population variance (r^2) in the quality of those items; the costs of sampling items; and the variance in the signal (s^2) an individual receives when they sample an item. For each decision rule we created sequences of individuals sampling and choosing an item while varying these parameter

Table 1. Metrics of sequence data

Metric	Definition	Example	Value
Sampled	Number of distinct items encountered	ABCA	3
Chosen	The item that is chosen	ABC**A**	A
Samples	Number of encounters with items	**ABAAC**A	5
Samples/items		ABCABC	5/3
Repeats	Individual samples the same item consecutively	A**BB**C**DD**B	2
Adjusted proportion of repeats	Repeats/(samples – 1)	A**BB**C**DD**B	2/5
Switch	An individual moves from one item to another	**AB**	1
Returns	A switch in which the individual samples an item that has been sampled earlier in the sequence	**A**BC**A**D**A**E	2
Proportion returns	Returns/samples	AB**C**D**C**A	1/5
Adjusted proportion returns	Returns/(samples – 1)	AB**C**D**C**A	1/4
Return to previous	A switch in which the individual returns to the item that it sampled immediately before	A**B**C**B**D	1
Adjusted proportion returns to previous	Returns to previous/(samples – 2)	A**B**C**B**D	1/2
Proportion returns	Returns/samples	AB**C**D**C**A	1/5
Last sample repeat	Last sample = (sample – 1) ≠ chosen	ABC**DD**E	Yes
Last sample return to previous	Last sample = return to previous ≠ chosen	AB**C**D**C**E	Yes
Last sample chosen repeat	Last sample = (sample – 1) = chosen	ABC**DDD**	Yes
Last sample chosen return to previous	Last sample = return to previous = chosen	AB**C**D**CC**	Yes
Last item sampled for the first time chosen		ABC**D**ADB**D**	Yes
Most sampled chosen	Chosen item was sampled ≥ second most sampled item	A**B**A**BB**B	Yes
Non-random last sample	Last sample was sampled ≥ second most sampled item & last sample ≠ chosen	A**B**CD**B**A	Yes
Non-random last sample chosen	Last sample was sampled ≥ second most sampled item & last sample = chosen	A**B**CD**BB**	Yes

Note: **Bold** letters emphasize important sequence portions for each metric. With the exception of *switch* and *chosen*, we used all of these metrics in the recursive partitioning exercise to build the new algorithms.

values. We systematically varied each of the four parameters one at a time while holding the other three variables fixed at the following standard conditions (cost = 0.05; $r^2 = 1$; $s^2 = 1$; pool size = 50). We varied the values for pool size (5, 10, 20, 50, 80), cost (0.001, 0.05, 0.10, 0.15, 0.20), and r^2 (0.01, 0.05, 1, 3, 5). Signal variance and sampling costs jointly affect how much an individual samples prior to choosing an item, so we varied signal variance (0.01, 0.05, 1, 3, 5) as long as sampling still occurred. For the comparative Bayes rule, when signal variance is too high, individuals stop sampling items and choose the first item they encounter. For each decision rule we produced 100 replicates of sampling sequences for each parameter combination. For each resulting sampling sequence, we applied the prevailing algorithm and determined if it correctly classified which rule produced the sequence.

Potentially better algorithms

Next, we attempted to identify new algorithms that have better rates of correct classification. We call one the ‘new individual algorithm’, the other the ‘new population algorithm’.

The first branching point of our new individual algorithm (Fig. 1B) asks whether the chosen item was the last item sampled for the first time. If not, the algorithm classifies the sequence as being produced by the best-of-*n* rule. Otherwise, the algorithm classifies the sequence as being produced either by the threshold rule or the comparative Bayes rule. The algorithm classifies the sequence as being produced by the comparative Bayes rule if the most sampled item in the sequence was chosen. Otherwise, the algorithm classifies the sequence as being produced by the threshold rule.

The first branching point of the new population algorithm (Fig. 1C) asks whether the last item sampled for the first time was chosen in at least 38% of sequences. If not, the algorithm classifies all of the sequences from that population as being produced by the best-of-*n* rule. The second branching point asks whether the chosen item was sampled greater than or equal to average in at least 88% of sequences. If so, the algorithm classifies all of the sequences from that population as being produced by the comparative Bayes rule. Otherwise, the algorithm classifies all of the sequences from that population as being produced by the threshold rule.

To test our new algorithms, we created two more data sets composed of 187,500 sequences of every combination of the following parameters: cost (0.05, 0.1, 0.15, 0.2, 0.25); signal variance (0.05, 1.3, 2.1, 2.9, 3.7); population variance (2, 2.4, 2.8, 3.2, 3.6); and pool size (5, 15, 25, 35, 45). We chose parameter values such that less than 1% of the time individuals either sampled only once or ran out of time and randomly chose an item. For each parameter combination, each rule was used to create 100 replicate sequences, mimicking what a population of 100 individuals using a specific rule would do. We calculated the sequence metrics described in Table 1.

Scoring the algorithms

Potentially all three rules can produce very similar sampling sequences, especially if two different organisms are receiving a different quality of information, or if the cost of sampling is unequal, or if they are choosing from populations with different values of pool size or variance. This problem (different rules producing the same sequence) is the largest hindrance to all the algorithms, and constitutes the primary reason we use a recursive partitioning analysis to draw out the key differences in sequences produced by different rules.

We used a decision-tree (algorithm) building approach to find the best algorithm for classifying what decision rule produced the various sequences (Breiman *et al.*, 1984). This was done using the *rpart: Recursive Partitioning* package in R. This package finds a point-estimate algorithm that classifies the highest number of sequences correctly based on each sequence’s parameter values. The penalty for adding branching points can be adjusted manually. We included all the metrics described in Table 1, except *chosen* and

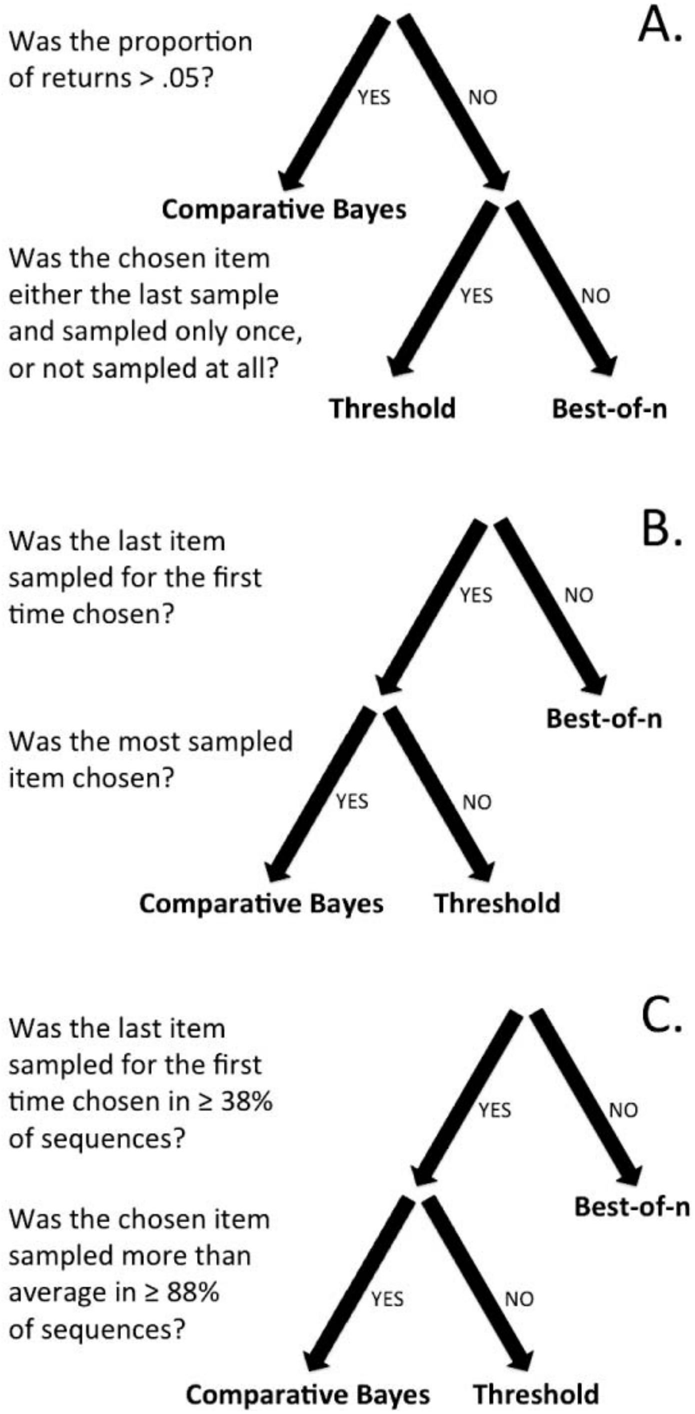

Fig. 1. (A) The prevailing algorithm. (B) The new individual algorithm. (C) The new population algorithm.

switch, as potential branching point criteria. Default parameters and the ‘class’ method were used for *rpart* settings. The value of ‘cost’ for reducing branching points was systematically decreased producing best-fit algorithms for 2, 3, and 4 branching points. We considered algorithms of more than 4 branching points to be of limited practicality and potentially overfit to our particular data set.

First, we performed the algorithm analysis using individual sampling sequences to classify what decision rule each individual used. Next, we averaged the metric values for each population of 100 sequences (for each parameter combination), and then performed a second classification algorithm analysis using population-level data. For both algorithms we chose the two-branch model because there was little improvement in their performance when a third or fourth branch was added.

To measure how well the two new algorithms performed compared with the prevailing algorithm, we had them classify sequences from a final data set (test data set) created in the identical manner of the previously described data sets. This way the identification and performance of algorithms would be independent and we would avoid identifying new algorithms that by chance performed well on only one data set.

Robustness of the population algorithm

By classifying what decision rule a population is using, one is assuming that all of the individuals in a population are using the same decision rule. We tested how robust the population algorithm was at identifying the predominant decision rule when not all individuals were using that rule. For each decision rule we used the test data set but replaced 10%, 20%, or 30% of the sequences with sequences from individuals using the two different decision rules (half of each decision rule). We randomly selected which sequences were being replaced. For example, a ‘threshold’ 10% alternative rule population was composed of 90 sequences produced by individuals using the threshold rule, five sequences produced by individuals using the comparative Bayes rule, and five using the best-of-*n* rule, all of them using the same parameter values.

RESULTS

The prevailing algorithm

The prevailing algorithm (Fig. 1A) classified only 32% of the test sequences to the correct decision rule (Fig. 3). Sequences tended to be classified as resulting from the threshold or comparative Bayes rules more frequently than from the best-of-*n* rule.

Examining each parameter individually, we see that the prevailing algorithm is least sensitive to signal variance and most sensitive to cost. Signal variance affects comparative Bayes only at low values, and cost has an opposite effect on the correct classification of comparative Bayes and threshold/best-of-*n* (Fig. 2). Both pool size and population variance affect the correct classification of all three rules in unique ways.

The new algorithms

Both new algorithms outperformed the prevailing algorithm (Fig. 3). And the new population algorithm outperformed the new individual algorithm. The new individual

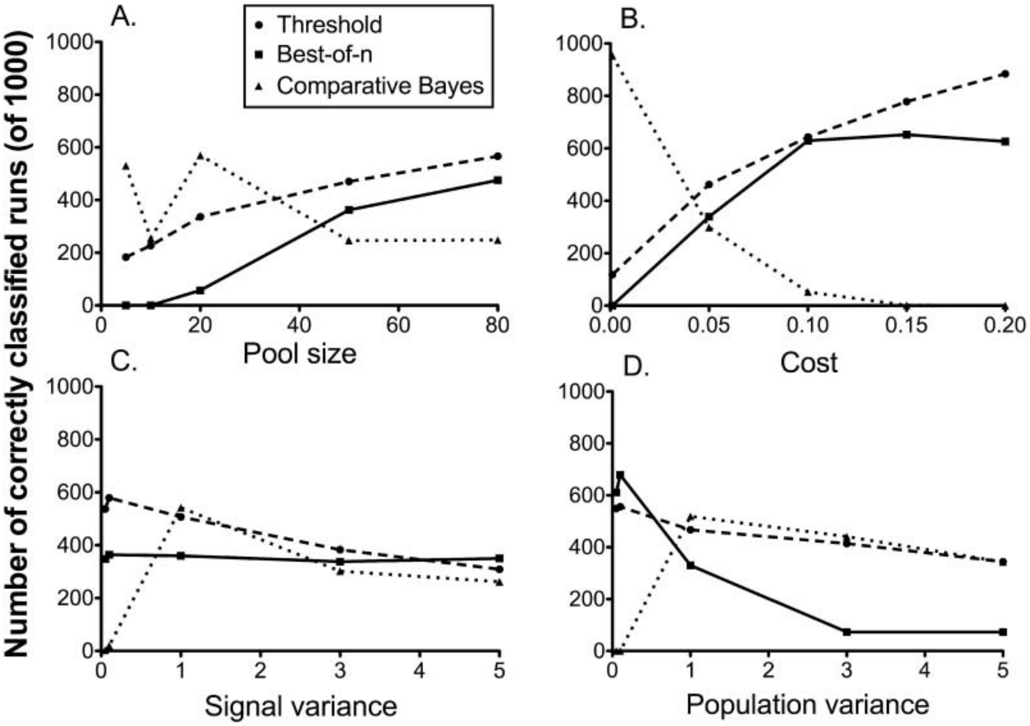

Fig. 2. Classification success of the prevailing algorithm when we sequentially varied: (A) pool size; (B) cost; (C) signal variance; and (D) population variance. Unspecified default values are as follows: pool size (50); cost (0.05); signal variance (1); and population variance (1).

algorithm classified 64% of sequences to the correct decision rule. But the new population algorithm classified 97% of sequences to the correct decision rule.

We also tested sensitivity of the performance of the population algorithm to changes in the values used at the two branching points. When we varied the values used in the algorithm, we found that the population algorithm still performed fairly well no matter which metric was varied (Fig. 4).

We tested the performance of the population algorithm under mixed-rule conditions and found it to be robust to heterogeneity in the decision rules of populations. When populations had 10%, 20%, and 30% of individuals not using the primary rule for the population, the population algorithm correctly classified the primary decision rule for 93%, 81%, and 64% of populations respectively.

Finally, we examined what biases the three algorithms have by applying them to the same test data set. The prevailing algorithm performed most poorly when diagnosing best-of-*n*, but was poor at diagnosing all three rules (Fig. 3). Our new individual algorithm diagnosed the threshold rule even more poorly than the old algorithm, but was relatively good at identifying the other two rules. The population algorithm outperformed the other two algorithms at identifying all three rules, and was over 95% correct classifying each rule.

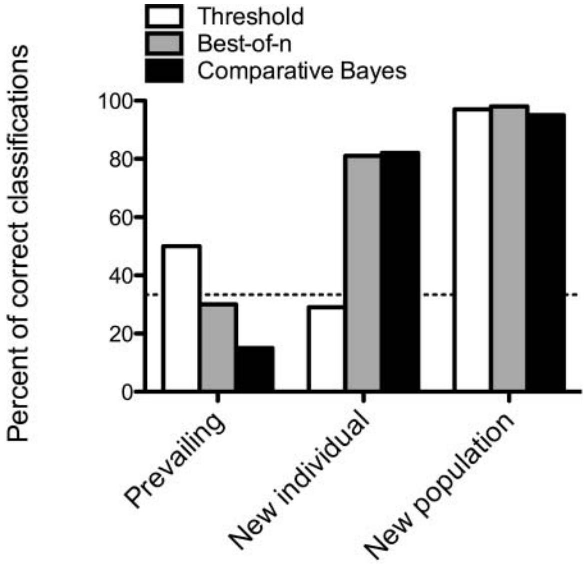

Fig. 3. Classification success of the three algorithms applied to our test data set. The horizontal dashed line is the expected 33% of correct classifications due to chance.

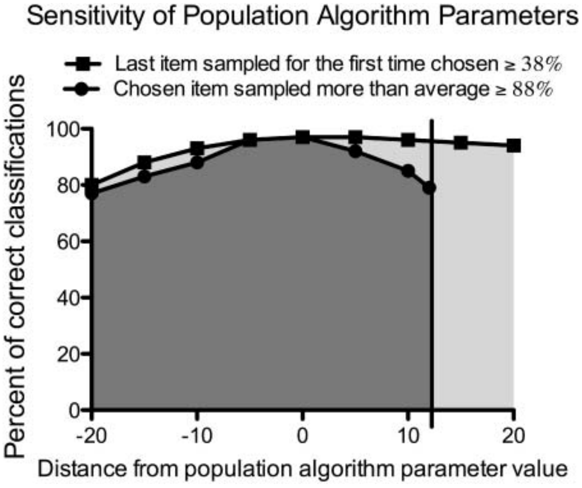

Fig. 4. Sensitivity analysis of the population algorithm. Here we varied the population algorithm parameters above and below the algorithm's optimal values (zero on the x-axis), and evaluated its performance.

Results with empirical data sets

We applied all three algorithms to eleven empirical data sets in which the authors provided the raw data on which items were sampled, in what order, and which item was ultimately chosen (Table 2). We had seven mate choice data sets, one for territory choice, and one for foraging.

Table 2. Comparison of classification results, based on the algorithm used, for eleven empirical data sets

Organism	Context	Data set	The majority of the individuals used	Sample size	Prevailing			New individual			New population
					TH (%)	BN (%)	CB (%)	TH (%)	BN (%)	CB (%)	
Great Reed Warbler (*Acrocephalus arundinaceus*)	Mate choice	1	BN	11	9	0	91	0	73	27	BN
Pied flycatcher (*Ficedula hypoleuca*)	Mate choice	2	Inconsistent with any rule	12	25	0	75	8	58	33	BN
Sage grouse (*Centrocercus urophasianus*)	Mate choice	3	TH	16	38	6	56	25	31	44	TH
Sand goby (*Pomatoschistus minutus*)	Mate choice	4	TH	27	100	0	0	0	0	100	CB
Two-spotted goby (*Gobiusculus flavescens*)	Mate choice	5	TH	28	89	0	11	11	4	86	CB
Brush mouse (*Peromyscus boylii*)	Territory	6	See previously used algorithm	47	15	19	66	54	31	14	BN
Pied flycatcher	Mate choice	7	Pooled-comparison	78	45	0	55	5	47	47	CB
Ortolan bunting (*Emberiza hortulana*)	Mate choice	8	64% TH, 18% pooled-comparison	135	78	0	22	7	11	81	CB
Scrub-jay (*Aphelocoma californica*)	Foraging	9	CB	315	29	0	71	33	23	44	TH
Human (*Homo sapiens*) – no time limit	Coin-filled eggs	10	–	270	62	0	37	0	100	0	BN
Human – with a time limit	Coin-filled eggs	11	–	1295	50	1	49	5	42	53	CB

Note: We interpreted, as best we could, the authors' conclusions for the rule used by the majority of each population. In some cases, the authors do not rule out the possibility of other rules, or use multiple lines of reasoning to infer which rule individuals are using. Dashes indicate that the author has not made any conclusions about which rule was used.

Data sets: 1, Bensch and Hasselquist (1992); 2, Dale *et al*. (1992); 3, Gibson (1996); 4, Forsgren (1997); 5, Myhre *et al*. (2012); 6, Mabry and Stamps (2008); 7, Dale and Slagsvold (1996); 8, Dale *et al*. (2006); 9, Luttbeg and Langen (2004); 10, T.A. Langen (unpublished data); 11, T.A. Langen (unpublished data).

Rules: TH, threshold; BN, best-of-*n*; CB, comparative Bayes.

In six of the eleven sets, the prevailing algorithm classified 70% or more of the individuals as using either the threshold or comparative Bayes rule. In the remaining five, it classified a fairly even number of sequences as being produced by the threshold and comparative Bayes rules. Few sequences were diagnosed as best-of-*n* in ten of the data sets and, in the remaining set, only 19% of sequences were classified as being produced by the best-of-*n* rule.

The new individual algorithm classified individual sequences substantially differently from the prevailing algorithm. In fact, they almost never agreed. Perhaps owing to the previously described bias of the new individual algorithm, threshold was classified as the predominant rule in only one of the eleven data sets (Fig. 3).

The new population algorithm classified five of the data sets as using the comparative Bayes rule, four as using the best-of-*n* rule, and two as using the threshold rule. The population algorithm agreed with the new individual algorithm for eight data sets. But it agreed with the prevailing algorithm for only a single data set.

Four of the seven mate choice data sets indicated that the population was using the comparative Bayes rule. Our single data sets for territory choice and foraging were classified as best-of-*n* and threshold, respectively.

DISCUSSION

Many ecological decisions regarding mate choice, foraging, and dispersal affect an organism's fitness, so the decision rules that individuals use to make these decisions should have fitness consequences (Jennions and Petrie, 1997). Furthermore, the decision rules that individuals use to make choices between alternative items may actually shape larger systems such as mating systems and the distribution of individuals across a landscape. For example, if a female is using a threshold or best-of-*n* rule, a male may have to perform only one good set of signals to be chosen by her. However, if she is using a comparative Bayes rule, he may need to court her by performing the signal multiple times. In order to understand how these rules shape larger systems, we must be able to classify correctly the rules that best describe patterns in nature.

The low number of available data sets limits the generalities that can be safely made about contextual or taxonomic rule-use bias. In fact, we discovered no obvious difference in the inferred rule used by birds, mammals, and fish. However, we do note that the comparative Bayes rule was found more often in mate-choice data than foraging and territory data. Perhaps mate choice decisions have a higher fitness consequence than foraging and territory decisions, thus leading to selection for the more memory-intensive comparative Bayes rule. In contrast, foraging behaviour may be performed so regularly that the constraints of memory select for a simple rule like threshold.

A popular approach for inferring what decision rules individuals are using is to compare observed patterns of individuals sampling and choosing items against the expected patterns of alternative decision rules. Although the prevailing algorithm for distinguishing between threshold, best-of-*n*, and comparative Bayes decision rules has been used in a few empirical studies, our study is the first to test the algorithm's effectiveness. When we simulated sampling sequences and imposed decision rules on them, we found that the prevailing algorithm performed very poorly at identifying what decision rule was in use. The prevailing algorithm had a success rate of only 32%, essentially the same as the 33% that would characterize random choice of a rule.

A number of weaknesses prevent the prevailing algorithm from correctly identifying the decision rule being used. First, individuals using the best-of-*n* and threshold rules sometimes randomly return to items that they have previously sampled. The prevailing algorithm classifies sequences in which the proportion of returns exceeds 0.05 as being produced by the comparative Bayes rule, thus misclassifying sequences when individuals using the best-of-*n* and threshold rules randomly return to an item. Small pool size, low costs, and high population variance are the most problematic conditions for identifying best-of-*n* and threshold (Fig. 2).

Second, the prevailing algorithm assumes that individuals using comparative Bayes will frequently return to items (Fig. 1A). But this assumption will be false if it is costly to return, or the population is low, or if signal variances are low. Also, when pool size is high, individuals using the comparative Bayes rule should rarely return to items. In all such cases, the large number of one-time visits to the various items lowers the rate of returning (Fig. 2) and reduces the correct identification of sequences produced by the comparative Bayes rule.

Faced with the inadequacy of the prevailing algorithm, we produced two alternative algorithms. One, the new individual algorithm, identified the threshold rule more poorly than the prevailing algorithm, but was nearly twice as successful at detecting both best-of-*n* and comparative Bayes. Overall, the new individual algorithm was nearly twice as likely as the prevailing algorithm to correctly classify which rule produced a sequence (64% vs. 32%).

We were able to improve upon the new individual algorithm by recognizing an inherent problem with it. Even when individuals are using the same decision rule with identical parameter values, they can produce very different sampling sequences because of random variation in the quality of items that an individual encounters and variation in the information they receive about those items. Our results show that within a group of individuals using the same decision rule, the individual algorithm falsely concludes that there is variation in the decision rule being used. Thus, we produced the new population algorithm.

The new population algorithm seeks to classify what rule is in use by a group of individuals. It had a 97% success rate. Even if the population had individuals using different rules, this algorithm had a high success rate at inferring the predominant decision rule being used by the population.

Several investigators have suggested that individuals may adjust their searching rules (either by changing rules or changing the parameters in a rule) as environmental conditions change. Suggested changes include predation risk (Hubbell and Johnson, 1987), mate density (Crowley *et al.*, 1991), costs of memory (Hutchinson *et al.*, 1993; Roitberg *et al.*, 1994), time constraints (Real, 1990), and age (Fiske and Kalas, 1995). Questions about variation in decision rule use between individuals and within individuals are very interesting. However, as yet we have no empirical evidence to support their existence (Castellano *et al.*, 2012). [For an in-depth review of mate choice rules, see Jennions and Petrie (1997).] Previously, some have concluded that there is variation within or between individuals in what decision rules are being used (Uy *et al.*, 2001; Dale *et al.*, 2006; Mabry and Stamps, 2008; Selonen and Hanski, 2010). But our results show that even when individuals are using identical rules, variation in sampling sequences will interfere with the reliable identification of the decision rule used for a single observed sequence. Variation of the decision rule may exist but the conclusions of the individual algorithms are not good evidence for it. Classification algorithms can reliably infer only the decision rules being used by the majority of the population, and we should avoid using them to infer the decision rule being used by a single individual.

Examining our new algorithms can give insights into what best distinguishes behaviour created by the alternative rules (Fig. 1B, 1C). It is noteworthy that for both the individual and the population algorithms, the metric '*last item sampled for the first time is chosen*' determines whether a sequence is classified as being produced from the best-of-*n* rule. This metric is similar to the second branching point of the prevailing algorithm: '*last sample chosen and sampled only once*'. The importance of this metric is intuitive because with the best-of-*n* rule, an individual should choose only the last sampled item approximately 1/*n* of the time, so best-of-*n* should generally not choose the last sample unless there is very little sampling.

The best metrics for the second branching point of both the new individual and population algorithm consider whether the chosen item is sampled more often than other items, and, when this is true, the sequence gets classified as produced by the comparative Bayes rule rather than the threshold rule. Thus we learn that non-random returning to the chosen item, not returning in general, provides the improved classification performance.

We based all our models on the assumption that encounter rates for all items in the pools were equal for every sampling event. However, in a spatially explicit world the probability of encountering neighbours is higher than encountering more distant individuals. Furthermore, over time, the probability of encountering a particular item will change because it grows nearer or farther away. Adding spatial complexity to the models might change both our individual and population algorithms dramatically. We find this a particularly interesting direction for future research, since non-random returning is important for our decision rule algorithms, and the probability of returning to an item should often decrease with distance to that item.

The population algorithm provides new insight into the characteristics of sequences important to classification of decision rule models. Most notably, comparative Bayes and threshold rules are often confounded. Previously, a high proportion of returns was considered characteristic of comparative Bayes, and no returns was considered characteristic of threshold. We have found that this is not always true. Individuals using the threshold rule often return to items, with the exception of the chosen item. In contrast, individuals using comparative Bayes do not necessarily randomly return to items, but selectively return to the chosen item. These are the primary reasons why our conclusions differ from those in the original papers (Table 2). The classification success of the population algorithm argues that researchers should consider it when concluding which sequence characteristics are produced by different rules.

ACKNOWLEDGEMENTS

We thank the National Science Foundation for its generous support of this work (OCE-0929691 to B. Luttbeg). We are also grateful to all of the researchers who graciously provided us with their sequence data to reanalyse.

REFERENCES

Arak, A. 1988. Mate selection in the Natterjack toad: active choice or passive attraction. *Behav. Ecol. Sociobiol.*, **22**: 317–327.

Beckers, O.M. and Wagner, W.E., Jr. 2011. Mate sampling strategy in a field cricket: evidence for a fixed threshold strategy with last chance option. *Anim. Behav.*, **81**: 519–527.

Bensch, S. and Hasselquist, D. 1992. Evidence for active female choice in a polygynous warbler. *Anim. Behav.*, **44**: 301–311.

Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. 1984. *Classification and Regression Trees*. Belmont, CA: Wadsworth International.

Castellano, S., Cadeddu, G. and Cermelli, P. 2012. Computational mate choice: theory and empirical evidence. *Behav. Process.*, **90**: 261–277.

Collins, E.J., McNamara, J.M. and Ramsey, D.M. 2006. Learning rules for optimal selection in a varying environment: mate choice revisited. *Behav. Ecol.*, **17**: 799–809.

Crowley, P.H., Travers, S.E., Linton, M.C., Cohn, S.L., Sih, A. and Sargent, R.C. 1991. Mate density, predation risk, and the seasonal sequence of mate choices – a dynamic game. *Am. Nat.*, **137**: 567–596.

Dale, S. and Slagsvold, T. 1996. Mate choice on multiple cues, decision rules and sampling strategies in female pied flycatchers. *Behaviour*, **133**: 904–944.

Dale, S., Rinden, H. and Slagsvold, R. 1992. Competition for a mate restricts mate search of female pied flycatchers. *Behav. Ecol. Sociobiol.*, **30**: 165–176.

Dale, S., Steifetten, O., Osiejuk, T.S., Losak, K. and Cygan, J.P. 2006. How do birds search for breeding areas at the landscape level? Interpatch movements of male Ortolan Buntings. *Ecography*, **29**: 886–898.

Dombrovsky, Y. and Perrin, N. 1994. On adaptive search and optimal stopping in sequential mate choice. *Am. Nat.*, **144**: 355–361.

Fiske, P. and Kalas, J.A. 1995. Mate sampling and copulation behavior of Great Snipe females. *Anim. Behav.*, **49**: 209–219.

Forsgren, E. 1997. Mate sampling in a population of sand gobies. *Anim. Behav.*, **53**: 267–276.

Gibson, R. 1996. Female choice in sage grouse: the roles of attraction and active comparison. *Behav. Ecol. Sociobiol.*, **39**: 55–59.

Gibson, R.M. and Langen, T.A. 1996. How do animals choose their mates? *Trends Ecol. Evol.*, **11**: 468–470.

Hubbell, S.P. and Johnson, L.K. 1987. Environmental variance in lifetime mating success, mate choice, and sexual selection. *Am. Nat.*, **130**: 91–112.

Hutchinson, J.M.C., McNamara, J.M. and Cuthill, I.C. 1993. Song, sexual selection, starvation and strategic handicaps. *Anim. Behav.*, **45**: 1153–1177.

Izzo, A.S. and Gray, D.A. 2011. Heterospecific courtship and sequential mate choice in sister species of field crickets. *Anim. Behav.*, **81**: 259–264.

Janetos, A.C. 1980. Strategies of female mate choice – a theoretical analysis. *Behav. Ecol. Sociobiol.*, **7**: 107–112.

Jennions, M.D. and Petrie, M. 1997. Variation in mate choice and mating preferences: a review of causes and consequences. *Biol. Rev. Camb. Phil. Soc.*, **72**: 283–327.

Luttbeg, B. 1996. A comparative Bayes tactic for mate sampling and choice. *Behav. Ecol.*, **7**: 451–460.

Luttbeg, B. 2002. Sampling the robustness and optimality of alternative decision rules with varying assumptions. *Anim. Behav.*, **63**: 805–814.

Luttbeg, B. and Langen, T.A. 2004. Comparing alternative models to empirical data: cognitive models of Western Scrub-Jay foraging behavior. *Am. Nat.*, **163**: 263–276.

Luttbeg, B. and Schmitz, O.J. 2000. Predator and prey models with flexible individual behavior and imperfect information. *Am. Nat.*, **155**: 669–683.

Luttbeg, B. and Trussell, G.C. 2013. How the informational environment shapes how prey estimate predation risk and the resulting indirect effects of predators. *Am. Nat.*, **181**: 182–194.

Mabry, K.E. and Stamps, J.A. 2008. Searching for a new home: decision making by dispersing Brush Mice. *Am. Nat.*, **172**: 625–634.

Myhre, L.C., de Jong, K., Forsgren, E. and Amundsen, T. 2012. Sex roles and mutual mate choice matter during mate sampling. *Am. Nat.*, **179**: 741–755.

Parker, G.A. 1978. Evolution of competitive mate searching. *Annu. Rev. Entomol.*, **23**: 173–196.

Parker, G.A. 1979. Sexual selection and sexual conflict. In *Sexual Selection and Reproductive Competition in Insects* (M.S. Blum and N.A. Blum, eds.), pp. 123–166. New York: Academic Press.

Parker, G.A. 1983. Mate quality and mating decisions. In *Mate Choice* (P. Bateson, ed.), pp. 141–166. Cambridge: Cambridge University Press.

Real, L. 1990. Search theory and mate choice. I. Models of single-sex discrimination. *Am. Nat.*, **136**: 376–405.

Roitberg, B.D., Reid, M.L. and Li, C. 1994. Choosing hosts and mates: the value of learning. In *Insect Learning: Ecological and Evolutionary Perspectives* (D.R. Papaj and A.C. Lewis, eds.), pp. 174–194. New York: Chapman & Hall.

Ryan, M.J., Akre, K.L. and Kirkpatrick, M. 2009. Cognitive mate choice. In *Cognitive Ecology II* (R. Dukas and J. Ratcliffe, eds.), pp. 137–155. Chicago, IL: University of Chicago Press.

Selonen, V. and Hanski, I.K. 2010. Decision making in dispersing Siberian flying squirrels. *Behav. Ecol.*, **21**: 219–225.

Trail, P.W. and Adams, E.S. 1989. Active mate choice at cock-of-the-rock leks – tactics of sampling and comparison. *Behav. Ecol. Sociobiol.*, **25**: 283–292.

Tuckerman, J.F., Gwynne, D.T. and Morris, G.K. 1993. Reliable acoustic cues for female mate preference in a katydid (*Scudderia curvicauda*, Orthoptera: Tettigoniidae). *Behav. Ecol.*, **4**: 106–113.

Uy, J.A.C., Patricelli, G.L. and Borgia, G. 2001. Complex mate searching in the satin bowerbird *Ptilonorhynchus violaceus. Am. Nat.*, **158**: 530–542.

Wiegmann, D.D., Angeloni, L.M., Seubert, S.M. and Wade, J.G. 2013. Mate choice decisions by searchers. *Curr. Zool.*, **59**: 184–199.

Wittenberger, J.F. 1983. Tactics of mate choice. In *Mate Choice* (P. Bateson, ed.), pp. 435–447. Cambridge: Cambridge University Press, Cambridge.