Testing take-the-best in new and changing environments

Michael D. Lee¹ · Gabrielle Blanco¹ · Nikole Bo¹

Published online: 7 September 2016
© Psychonomic Society, Inc. 2016

Abstract Take-the-best is a decision-making strategy that chooses between alternatives, by searching the cues representing the alternatives in order of cue validity, and choosing the alternative with the first discriminating cue. Theoretical support for take-the-best comes from the “fast and frugal” approach to modeling cognition, which assumes decision-making strategies need to be fast to cope with a competitive world, and be simple to be robust to uncertainty and environmental change. We contribute to the empirical evaluation of take-the-best in two ways. First, we generate four new environments—involving bridge lengths, hamburger prices, theme park attendances, and US university rankings—supplementing the relatively limited number of naturally cue-based environments previously considered. We find that take-the-best is as accurate as rival decision strategies that use all of the available cues. Secondly, we develop 19 new data sets characterizing the change in cities and their populations in four countries. We find that take-the-best maintains its accuracy and limited search as the environments change, even if cue validities learned in one environment are used to make decisions in another. Once again, we find that take-the-best is as accurate as rival strategies that use all of the cues. We conclude that these new evaluations support the theoretical claims of the accuracy, frugality, and robustness for take-the-best, and that the new data sets provide a valuable resource for the more general study of the relationship between effective decision-making strategies and the environments in which they operate.

Keywords Take-the-best · Fast and frugal decision making · Non-compensatory decision making · Heuristic decision making

Introduction

The “fast and frugal heuristics” approach to understanding human decision-making (Gigerenzer & Goldstein, 1996; Gigerenzer, Todd, & the ABC Group, 1999) is based on two major theoretical assumptions. One is that the world is competitive, which means that the ability to make fast decisions is an advantage. The other is that the world is uncertain and changeable, which means that making robust decisions—that is, decisions that are not sensitive to fine-grained details of the environment—is an advantage.

Taken together, these theoretical assumptions provide a basis for people not using all of the information available in an environment to make decisions. Instead, following the fast and frugal theory, people may make decisions based on a limited search for the most important information. The claim is that this use of non-compensatory decision strategies, sensitive to information structures in the environment, allows people to make fast and robustly accurate decisions.

Take-the-best

Perhaps the best-studied fast-and-frugal heuristic is take-the-best (Gigerenzer & Goldstein, 1996). This is a model of forced choice between two alternatives, each of which is represented in terms of the presence or absence of a set of cues. Associated with each cue is a measure of its validity, which is defined as the probability the cue belongs to the correct alternative, for those situations where one alternative has the cue but the other does not. Take-the-best assumes

✉ Michael D. Lee
mdlee@uci.edu

¹ Department of Cognitive Sciences, University of California, Irvine, Irvine, CA 92697-5100, USA

that cues are searched according to their validities, starting with the most valid cue, and search is terminated as soon as a discriminating cue is found. At that point, the alternative with the discriminating cue is chosen, or a random choice is made if search exhausts all the cues without discriminating between the alternatives.

Take-the-best clearly follows the fast and frugal theoretical assumptions. Because search terminates as soon as a discriminating cue is found, decisions are typically made quickly, based on cues with high validity. In environments with the right types of structures, take-the-best is capable of making accurate decisions (Gigerenzer & Brighton, 2009; Lee & Zhang, 2012). For example, in environments with diminishing returns, in which the most valid cues are much more important than the other cues, it makes sense to terminate search once a high-validity discriminating cue has been found, since the remaining lower-validity cues could not change the decision. Alternatively, in environments with a correlated structure, in which most discriminating cues favor the same alternative, it makes sense to terminate search at the first discriminating cue, since further search is likely to find further evidence in favor of the same decision.

Take-the-best is technically a non-compensatory decision-making strategy, because it usually does not use all of the available cue information. If the first discriminating cue found favors one alternative, the remaining cues—which will not be searched—cannot change (compensate for) this initial choice. It is natural to compare take-the-best to compensatory decision-making strategies that search all of the cues exhaustively, and use all of the information in some way. Two compensatory strategies commonly considered as contrasts are the tally and weighted additive (WADD) strategies. In the tally strategy, the number of cues favoring one alternative is compared to the number of cues favoring the other, and the alternative with the greatest number is chosen, or a random choice is made in the case of a tie. In the WADD strategy, the cue validities favoring each alternative are combined and compared. Sometimes the validities themselves, or a “chance corrected” form of the validities, are summed (Gigerenzer & Goldstein, 1996; Hilbig & Moshagen, 2014). If the intention is to use WADD as a normative compensatory strategy, then it is incorrect to sum the validities, and, instead, the log odds of the validities must be summed (Lee & Cummins, 2004; Lee & Zhang, 2012; Lee, 2016).

Evaluating take-the-best

The extent to which environments have the regularities that support fast and frugal decision making is an open research question, and makes the evaluation of heuristics like take-the-best in real-world environments an important topic of study. Czerlinski, Gigerenzer, and Goldstein

(1999) evaluated take-the-best using 20 environments, many of which were also used in a later evaluation reported by Brighton (2006). Recently, Şimşek (2013) greatly expanded the number of available environments, in a study of linear decision rules for 51 environments. In both the Czerlinski et al. (1999) and Şimşek (2013) studies—which overlap in the environments they consider—the environments span a wide range of areas, and vary significantly in the number of stimuli and cues they use. Unfortunately, however, many of cues are inherently continuous properties of the stimuli—such as the average temperature in a city—and so strong assumptions are made to transform them to a binary representation. This is typically done by a median split procedure. There are exceptions throughout the environments, including inherently binary cues such as whether a famous actor or actress is American, or whether a city has rent control, but these constitute a relatively small number of the cues. None of the environments, other than the original German cities environment developed by Gigerenzer and Goldstein (1996), appears to be based solely on naturally binary properties of the stimuli.

One obvious problem with using cues defined by median splits is that they do not represent the environment well. Stimuli with very similar values on the underlying continuous variable can have different cue values (if they are close to, but either side of, the median), while stimuli with very different values can have the same cue value (if one is close to the median, but the other is far away). A more specific problem, in the context of decision-making models, relates to cue discriminability. This is a complementary measure to validity, and is defined as the proportion of pairs of stimuli that differ with respect to a cue. The use of the median split guarantees that every cue has a discriminability of one-half, since half the stimuli have the cue but half do not.¹ While take-the-best only relies on cue validity, the fast and frugal framework naturally lends itself to extended or alternative models that do incorporate discriminability. For example, Newell, Rakow, Weston, and Shanks (2004) consider the search criterion they call “success”, which combines cue validity and discriminability to assess the probability a cue by itself makes correct decisions. Lee and Newell (2011) and Lee and Zhang (2012) consider a family of search criteria found by the weighted linear combinations of validity and discriminability for each cue. Clearly, environments in which every cue has the same discriminability are not useful for evaluating these models of the search process in decision making.

¹Or approximately so, in the case where the median split is applied to an odd number of stimuli.

Other evaluations of take-the-best, or closely related strategies, have focused on a few environments from specific applied domains such as legal and medical decision making (Dhami, 2003; Dhami & Ayton, 2001; Dhami & Harries, 2001). These data sets are typically based on genuinely binary cues, and are impressively large and grounded in real-world observation. Lee and Zhang (2012) developed city environments for Italy, the US, and the UK, mimicking the approach taken to develop the original German cities environment. These environments, together with the legal and medical ones, are well-suited to evaluate take-the-best, but relatively few of them exist. The final possibility for large scale evaluation of take-the-best in real-world environments lies in the repositories developed in machine learning research communities to test classification algorithms (e.g., The University of Toronto, 1996; Lichman, 2013). In general, these data sets rely on features or dimensions of objects that are not naturally binary cues, although, as for the environments considered by Czerlinski et al. (1999) and Şimşek (2013), there are exceptions.

Current aims

Overall, there is a need for the development of new environments suited to the evaluation of take-the-best. These environments need to represent the objects in terms of naturally binary cues, and have a criterion that is a basis for forced-choice decisions. The first aim of this paper is to develop a number of such environments, and use them to evaluate the performance of take-the-best and alternative compensatory decision-making strategies.

The second aim of this paper stems from the emphasis of fast and frugal heuristics coping with the uncertainty that stems from environmental change. As Gigerenzer et al. (1999, p. 18) put it: “Fast and frugal heuristics avoid this trap by their very simplicity, which allows them to be robust in the face of environmental change and enables them to generalize well to new situations”. A key evaluation implied by this theoretical claim is how take-the-best performs as an environment changes. For example, for the original German cities environment, the question is how take-the-best performs as the populations of the cities change over time, and as cues like whether or not a city has a team in the Bundesliga change.

To the best of our knowledge, no evaluations of this sort have been with take-the-best, although the importance of dynamic environments is emphasized by Serwe and Frings (2006) in studying the closely related recognition heuristic. Accordingly, we develop environments for the German, Italian, US, and UK city domains, based on their structure at different points in time. We then evaluate the performance of the take-the-best, tally, and WADD strategies across these changing environments.

New environments

We developed four new environments, involving the lengths of 65 bridges based on 12 cues, the prices of hamburgers at 30 fast food chain restaurants based on 6 cues, the number of people visiting 20 theme parks based on 7 cues, and the Times Higher Education world university rankings score of 65 US Universities based on 9 cues. Details of the environments, including criterion and cue values for all of the stimuli, are available on the Open Science Framework at <https://osf.io/yd7mw/>.

The cues used in each environment are listed in Table 1. These cues were chosen because of their intuitive reasonableness, and without any specific consideration of their likely usefulness as input to specific decision strategies. In particular, the cues were not chosen because they were expected to favor take-the-best, and no evaluation of take-the-best or any other decision-making strategy using the environments was conducted until the environments were finalized. The cues were simply chosen because they were naturally binary, the relevant information was available, and they seemed likely to be relevant to the criterion in some way.

Table 1 provides useful summary measures of the environments and the cues, especially as they relate to non-compensatory strategies for decision making like take-the-best. First, following Baucells, Carrasco, and Hogarth, (2008), measures of dominance and cumulative dominance are listed for each environment. A stimulus dominates another if it has as much or more evidence that it should be chosen for every cue. One loose interpretation is that single-cue strategy will make the same decision, regardless of which cue is used, when dominance exists. A stimulus cumulatively dominates another if the combined evidence over the cues—searched in some order—is always as large or greater as each additional cue is included in the total. One loose interpretation is that a strategy that combines cue evidence as it searches will always make the same decision, regardless of how many cues it searched, when cumulative dominance exists. We assume that the appropriate additive measure of evidence is the log-odds of the cue validities, and that cues are searched in order of decreasing validity. Using those assumptions, the π_d and π_c measures in Table 1 show the proportion of stimulus pairs in which one dominates and cumulatively dominates, respectively. Dominance ranges from about 10 % to about 30 % of all stimulus pairs across the environments. Cumulative dominance is, as expected, higher, at around 50 % of the stimulus pairs for three of the environments, but about 24 % for the Hamburgers environment.

For each of the cues in each environment, Table 1 also lists the standard measures of validity and discriminability. To help understand the correlational structure of each

Table 1 Cues for the four new environments, and their validities v , discriminabilities d , and positive d^+ and negative d^- discriminabilities. Also listed for each environment is the proportion of dominant π_d and cumulatively dominant π_c stimulus pairs

Environment	π_d	π_c	Cue	v	d	d^+	d^-
Bridges	0.12	0.50	toll	0.51	0.47	0.22	0.25
			capital	0.66	0.20	0.07	0.12
			guinness	0.96	0.14	0.03	0.11
			long	0.54	0.46	0.25	0.21
			viaduct	0.78	0.33	0.15	0.18
			box-girder	0.68	0.43	0.15	0.28
			cable-stayed	0.52	0.46	0.25	0.21
			suspension	0.54	0.14	0.06	0.08
			steel	0.50	0.47	0.23	0.24
			concrete	0.51	0.36	0.21	0.15
			rail	0.63	0.31	0.17	0.13
			car	0.82	0.17	0.09	0.08
Hamburgers	0.30	0.55	meal deal	0.59	0.33	0.20	0.11
			drive through	0.98	0.52	0.48	0.04
			24 hours	0.87	0.48	0.40	0.08
			calorie menu	0.52	0.46	0.21	0.24
			self-serve drinks	1.00	0.52	0.47	0.05
			build burger	0.56	0.29	0.22	0.07
Theme Parks	0.17	0.23	trademark	0.68	0.27	0.19	0.07
			water-rides	0.86	0.19	0.16	0.03
			neighboring	0.77	0.51	0.44	0.06
			seasonal	0.73	0.39	0.25	0.14
			hotels	0.81	0.52	0.43	0.09
			multiple sites	0.89	0.10	0.05	0.05
US Universities	0.23	0.49	parades	0.61	0.51	0.36	0.14
			public	0.66	0.50	0.19	0.31
			research	0.67	0.09	0.04	0.05
			smoke free	0.52	0.46	0.27	0.19
			ncaa d1	0.52	0.38	0.18	0.20
			capital	0.56	0.29	0.14	0.15
			rotc	0.57	0.34	0.16	0.18
			quarter	0.66	0.34	0.11	0.24
			president	0.82	0.33	0.07	0.26
			online	0.52	0.42	0.23	0.19

environments, the positive and negative discriminability measures d^+ and d^- are also shown. These are relatively new measures, introduced by Lee and Zhang (2012). The basic motivation for the d^+ and d^- measures is to examine how often the information provided by one cue is consistent with information provided by other cues. That is, when a cue discriminates, does it provide evidence consistent in favor of the same alternative as the overall evidence provided by the other cues? The positive discriminability d^+ is the proportion of times, when a cue does discriminate, it favors the same alternative as the summed evidence from the other cues. The negative discriminability d^- is the proportion of times, when a cue does discriminate, the summed evidence

from the other cues favors the other alternative.² In this way, relatively greater positive discriminability can be interpreted as meaning that cues provide evidence for the same alternative, consistent with a correlated environment. The pattern of d^+ and d^- measures suggest that the Hamburgers and Theme Parks have this sort of correlated structure, but the Bridges and US Universities environments consist of cues that tend to provide more independent evidence.

²Lee and Zhang (2012) considered d^+ and d^- as partitioning d , which is a useful way to conceive of the measures. The cases in Table 1 for which $d^+ + d^- \neq d$ are because of rounding, or because cases in which the remaining cues provide equal evidence for both alternatives are not included in either measure.

Fig. 1 The discriminability and validity of each cue in each of the four new environments

Figure 1 shows the discriminability and validity of each cue, in each of the four new environments. In each environment, the cues collectively span a wide range of the possible discriminability values from 0 to 0.5, and possible validity values from 0.5 to 1. As is standard in this literature (e.g., Gigerenzer & Goldstein, 1996), cues are coded in terms of the meaning of presence or absence so that all validities are at least 0.5. For three of the environments—bridges, theme parks, and US universities—there is an evident negative correlation between discriminability and validity. This follows expectations, because it is natural for high validity cues to be rare, and for cues that discriminate well to correlate only loosely with the criterion. As concrete examples, the low discriminating national capital cue in the German cities environment (post re-unification) is highly valid, because only Berlin is the national capital, and it has the largest population, while the Former East Germany cue is highly discriminating but has low validity, because around half the cities are from the former east Germany, but there are small and large population cities in both the former east and former west Germany. More generally, the combinatorics of ordering suggest that it is more likely cues with lower discriminabilities will have higher validities. What is needed for a cue to have a high validity is that all the stimuli with the cue have high values on the criterion, and so are

“clustered” at the top of the order (or all the stimuli that do not have the cue are clustered at the bottom). This is more likely to happen if there are relatively few stimuli with the cue, which corresponds to a cue with low discriminability. According to this analysis, it should be expected that validity and discriminability are negatively correlated. The fast food environment provides an interesting exception to this general expectation, with a number of highly-discriminating and highly-valid cues.

Strategy performance

Figure 2 summarizes the performance of take-the-best, and the compensatory tally and WADD strategies, on the four new environments. The left panel relates to accuracy, showing the percentage of correct decisions made over all possible stimulus pairs. Accuracies range from about 65 % to about 75 %, and are very similar for all three strategies. The right hand panel relates to frugality, or the extent of search. For each environment, the distribution of the number of cues searched by take-the-best is shown, as well as the average number of cues searched across all stimulus pairs. For the compensatory strategies, the number of cues searched is simply the number of cues in the environment. It is clear from this analysis that take-the-best searches fewer than half

Fig. 2 Performance of take-the-best, tally, and WADD on the new environments. The left panel shows accuracy in comparing all possible pairs of stimuli. The right panel shows the average number of

cues searched for all decision strategies (lines with markers), and the distribution of the number of cues searched by take-the-best (shaded circles)

the cues in each environment, on average, and often makes a decision based on the first cue it encounters.

We also conducted a test of the performance of each strategy on each environment using a train-test approach, which provides an out-of-sample evaluation. We considered training set sizes ranging from two stimuli to all but two of the stimuli in each environment. A total of 1000 training sets were chosen at random for each set size, and the remaining stimuli formed the test set. For the take-the-best strategy, cue validities were learned from the training set, re-coding the cue if needed to make its validity at least 0.5. The strategy was then applied with the resultant validities and cue codings to all of pairs of stimuli in the test set. The WADD strategy is not affected by these re-codings, but does depend on the validities learned from the training set. The tally strategy does not involve validities, and so was simply applied to all of pairs of stimuli in the test set.

Figure 3 summarizes the results of this evaluation. Each panel corresponds to an environment, and the distribution of accuracy—the mean, and 2.5 % to 97.5 % range—for each strategy at each training set size is shown. The mean number of cues search by take-the-best at each training set size is also shown. All four environments show similar pattern of results, with all three strategies performing approximately equally well for each training set size. All of the strategies naturally show a wide distribution of performance when the training set size is large, and the test set size is correspondingly small. There is a suggestion in all four environments that the average performance of the WADD strategy is slightly worse for very small training set sizes. For the two environments with the most stimuli—the bridges and US universities—there is also a decline in the

Fig. 3 Out-of-sample performance of take-the-best, tally, and WADD on the new environments. Each panel corresponds to an environment. The x-axis corresponds to the number of stimuli in a training set. The lines with markers show the mean accuracy, against the left-hand y-axis, of each strategy on the remaining test set items, with error bars showing the 2.5 % to 97.5 % range of the distribution of accuracy over 1000 repetitions. The broken line shows, against the right-hand y-axis, the mean number of cues searched by take-the-best on the test set

mean accuracy of the tally strategy for large training set sizes. The mean number of cues searched by take-the-best is relatively large for small training set sizes, but is stable around one-third of the total number of cues once the training set size is moderately large. It is clear from this analysis that the take-the-best strategy performs as well or better than the tally and WADD strategies across the entire range of out-of-sample tests, and inherently needs fewer cues to achieve this level of accuracy.

Changing environments

To test the robustness of take-the-best and other decision strategies in changing environments, we developed data sets describing German, Italian, US, and UK city domains at different points in time. The cities and cues used for Germany were taken from the original data set in Gigerenzer and Goldstein (1996), while the cities and cues used for Italy, the US, and the UK were taken from the original data sets in Lee and Zhang (2012). The sets of years chosen for each country varied, and were based on ready availability of the necessary information about city populations and cue properties. For Germany, we generated four new data sets for the years 1950, 1990, 2000, and 2010. For Italy, we generated five new data sets for the years 1981, 1991, 1995, 2000, and 2005. For the US, we generated five new data sets for the years 1900, 1950, 1990, 2000, and 2005. Finally, for the UK, we generated five new data sets for the years 1981, 1990, 1995, 2001, and 2005. We did not include the original years considered by Gigerenzer and Goldstein (1996) and Lee and Zhang (2012) because we wanted to use consistent sources, using the same definitions of city populations, which can vary depending on whether or not a greater metropolitan area is included. Thus, we used the original data sets only to determine which cities and cues were used to represent the environments over the years considered.

All of these 19 data sets are archived using the Open Science Framework at <https://osf.io/yd7mw/>. While they were all generated carefully, using multiple sources and researchers to try and resolve ambiguities in cue values, it is likely that some errors remain. At a minimum, many of the cue definitions require some degree of interpretation involving partially arbitrary decisions. Exactly what structure is significant enough to be a “rail station” or a major “airport”, for example, is not always obvious, and changes in administrative definitions over time can affect reported cue values. Some cues also change their meaning over time: the “former East Germany” cue is not “former” in 1950, for example, and “intercity trainline” only became

a concept in Germany in the 1960s. We do not believe these sorts of discrepancies affect the major conclusions of our analyses, and think our new data sets are a useful resource. But, we caution against treating the data sets as completely accurate characterizations of the environments they are meant to represent.

Environment properties

Figure 4 shows the pattern of change in cue validities, for each cue in each country over time. It is clear that some validities are very stable. For example, Rome has been the largest city in Italy over the time periods considered, and so the national capital cue always has validity one. Other cue validities change over time. This change can either be steady or sudden, and can be caused by changes in whether cities have the cue, the relative populations of the cities, or a combination of both of these factors. For example, Washington has always been the national capital of the US, but other cities have progressively grown to have larger populations, leading to a steady decrease in the validity of the national capital cue. The airport cue in the US, in contrast, changes suddenly from 1900 to 1950, rising from a validity of one-half, because there were no airports, to a validity near 0.8, because, as airports were built, they tended to be located in larger cities.

Figure 5 shows the pattern of change in cue discriminabilities. It is clear that they tend to more stable than validities over time. This makes sense, since cue discriminabilities depend only on whether or not cities have cues, and not on the relative order of their populations. The discriminabilities that shows continual change are for cues like sporting teams (sport team in the US, premier league in the UK, serie A and serie B in Italy, and soccer team in Germany) that do change regularly. Other cues show more of a step-change in discriminability, for building projects or events, such as the airport cue in the US. Many cues have constant discriminability over the entire time period, and there are relatively few changes in the order of cue discriminabilities.

Strategy performance

The theoretical motivation for considering changing environments is to examine the robustness of decision strategies to these changes. Figure 6 summarizes the accuracy of the take-the-best, WADD, and tally decision strategies over time in the changing environments for the four countries. Countries correspond to rows, while strategies correspond to columns, with take-the-best on the left, WADD in the middle, and tally on the right. The solid line corresponds to

Fig. 4 Pattern of change in cue validities over time with respect to city populations in four countries

the accuracy of each strategy, applied to the data set for each year. All of these accuracies lie in a 60–80 % range, and typically are close to 70 %. They change relatively little, in almost all cases, over the different years.

For the take-the-best and WADD strategies, an extra test of robustness to change is possible, because these

strategies depend on cue validity. Take-the-best orders the search of cues according to validity, and WADD combines cue validities. Thus, it is interesting to consider how each performs when applied to one environment, using cue validities learned in a different environment. The accuracy of the take-the-best and WADD strategies under this

Fig. 5 Pattern of change in cue discriminabilities over time with respect to city populations in four countries

Fig. 6 Accuracy of the take-the-best, WADD, and tally decision strategies in the changing environments. Solid lines show the accuracy for strategies with cue validities based on the year to which the strategy

is being applied. Broken lines show the accuracy for strategies with cue validities based on different years. The gray histograms show the distribution of accuracy over all possible search orders for the cues

approach—learning validities on every possible year, and applying them to every other possible year—are shown by the broken lines and open markers. It is clear that this results in, at most, very small decreases in accuracy. Intuitively, this stronger test of robustness corresponds to assuming information about cue validity is not continually adapted as environments change, but is learned at some point and then applied even though the changes shown in Fig. 4 have occurred.

For take-the-best, Fig. 6 provides one additional level of analysis. The gray histograms show the distribution of accuracy over *all* possible cue search orders for each environment in each year. It is clear that the observed performance of take-the-best using a decreasing-validity search order—whether those validities are based on the current year or a different year—are among the very best in the distribution,

Thus, the results in Fig. 6 make it clear that all of the strategies are robust to the changing cue values in the country data sets, and take-the-best and WADD are additionally robust to the changes in validity over time.

Figure 7 summarizes the frugality of take-the-best over the changing environments. Both the WADD and tally strategies, of course, use all of the available cues for every decision. It is clear from Fig. 7 that take-the-best requires many fewer cues, and that this frugality is generally robust across environmental change. Once again, this robustness is evident both for changes in cues over different years, and whether the cue validities are based on the current environment, or one from another year. The only significant change in frugality is for the US from 1900 to 1950 to 1990. As is clear from Fig. 5, the discriminability of cues generally rose over this period, as sports teams were established,

Fig. 7 Frugality of the take-the-best in the changing environments. Solid lines show the accuracy for strategies with cue validities based on the year to which the strategy is being applied. Broken lines show the

accuracy for strategies with cue validities based on different years. The gray histograms show the distribution of the number of cues searched over all possible search orders for the cues

airport and metro stations were built, and so on. The relatively larger numbers of cues searched in these early years is presumably a result of this lower discriminability, meaning more cues have to be examined to find a reason to choose one city over another.

The gray histograms show the distribution of the mean number of cues searched for all possible search orders. Typically, take-the-best requires about half of the available cues, on average, to make a decision. Interestingly, the observed frugality of take-the-best is in the center of this distribution for the United States and United Kingdom environments, but among the least frugal—and sometimes in the extreme tail of the longest searches—for the Italy and Germany environments.

Discussion

We have provided two sorts of evaluations of take-the-best as a decision-making strategy. Our first evaluation involved testing its performance in a set of four new environments. These environments have the advantage of having naturally binary cues, so that both cue validity and cue

discriminability are meaningful. We found, consistent with previous evaluations, that take-the-best is as accurate as WADD and tally strategies that consider all the cues, and is inherently more frugal. These conclusions were supported by a basic analysis of accuracy and the number of cues searched by take-the-best on each new environment in its entirety, and in an extended train-test analysis that considered out-of-sample performance. Our second evaluation involved testing the robustness of take-the-best to environmental change, including both changes in cue values and the order of cues according to their validities. Our testing involved the generation of 19 new environments, supplementing existing representations of the cities in four countries. We found that take-the-best, along with the WADD and tally strategies, was robust in the accuracy of their decisions. Take-the-best, however, maintained its inherent advantage of being much more frugal, typically requiring only half the number of cues.

While these findings answer the questions set out in our current aims, there are deeper questions that could be asked, and more detailed analyses that could be done to try and answer them. The four new environments we tested provide more evidence of the generality of take-the-best as

an accurate and frugal decision-making strategy, but leave open the question of exactly what environmental properties are required. This is an active area of study (e.g., Hogarth & Karelaia, 2007; Katsikopoulos & Martignon, 2006; Lee & Zhang, 2012; Martignon & Hoffrage, 2002), and the four additional environment data sets we have generated should provide a valuable resource for studying the relationship between effective decision strategies and the environments in which they operate. One obvious use for these environments is to test other decision strategies, beyond the take-the-best, tally, and WADD strategies. Possibilities include other plausible process of models of cognition, such as non-compensatory strategies that use more than one discriminating cue, or strategies that use both discriminability and validity to structure their cue search (e.g., Lee & Newell, 2011; Lee & Zhang, 2012; Newell et al., 2004) as well as other normative models, including more complete Bayesian benchmarks than the WADD strategy. A second use of these environments is to consider the relationship between strategy performance on cues that are inherently binary, as we have considered, and cues that are inherently continuous and so require representational assumptions to make binary, as has been the focus of most previous studies (e.g., Czerlinski et al., 1999; Şimşek, 2013). The development and evaluation of models for the representational processes that lead to binary representations is also an active area of research (Luan, Schooler, & Gigerenzer, 2014; Şimşek & Buckmann, 2015).

Finally, for the question of environmental change, deeper analysis is needed to understand exactly why take-the-best robustly maintains its accuracy. This is a challenging problem, because whether or not the stimuli have various cues changes along with their ordering with respect to the decision criterion. One possible answer is suggested by the work of Martignon and Hoffrage (2002) and Todd and Dieckmann (2005), who present evidence that take-the-best often produces answers that are relatively insensitive to the order in which cues are searched. This property would explain why changes in cue order—which are not extremely frequent, but certainly occur, as seen in Fig. 4—do not affect accuracy. Our results suggest, however, this can only be a part of the answer, since the distribution of accuracy for all possible search orders showed that take-the-best's use of search orders based on validity did provide a clear benefit. Ultimately, a characterization of what sorts of environmental changes lead to robustness, and what sorts might make take-the-best or other strategies more fragile, will require a detailed analysis of the inter-related changes between the decision criterion and cues. Ideally, this analysis would be based on some sort of organizing taxonomy for the different possible types of environment change. Intuitively, it seems like some changes are punctate and discrete, as when a city acquires a train station, some are cyclical, as when

a team moves between the Bundesliga and lower divisions from season to season, and others may be gradual, as when a city steadily grows in population. We are a long way from understanding what different change patterns exist, and how they differentially affect different decision strategies. Thus, the 19 new data sets we have generated should provide a valuable resource for studying how decision strategies are affected by environmental change, but are only a first step.

Fast and frugal heuristics are based on theoretical assumptions about the need for quick and robust decisions in a competitive and changeable world, and are designed to use the structure of environments in which stimuli are defined by binary cues to achieve this speed and robustness. We have expanded the set of naturally binary environments that can be used to evaluate take-the-best, and other heuristics and strategies, and have developed a first set of changing environments to extend the types of evaluations and analyses that can be considered. We hope these new environments sharpen our theoretical understanding of decision making, and its relationship to environmental structures.

Acknowledgments We thank Peter Todd for helpful discussions, and Konstantinos Katsikopoulos and two anonymous reviewers for helpful comments on an earlier version of this paper.

References

- Baucells, M., Carrasco, J. A., & Hogarth, R. M. (2008). Cumulative dominance and heuristic performance in binary multiattribute choice. *Operations Research*, 56, 1289–1304.
- Brighton, H. (2006). Robust inference with simple cognitive models. In C. Lebiere & R. Wray (Eds.), *AAAI spring symposium: Between a rock and a hard place: Cognitive science principles meet AI-hard problems* (pp. 17–22). Menlo Park, CA: American Association for Artificial Intelligence.
- Czerlinski, J., Gigerenzer, G., & Goldstein, D. G. (1999). How good are simple heuristics? In G. Gigerenzer, P. M. Todd, & The ABC Research Group (Eds.), *Simple heuristics that make us smart*, (pp. 97–118). London: Oxford University Press.
- Dhami, M. K. (2003). Psychological model of professional decision-making. *Psychological Science*, 14, 175–180.
- Dhami, M. K., & Ayton, P. (2001). Bailing and jailing the fast and frugal way. *Journal of Behavioral Decision Making*, 14, 141–168.
- Dhami, M. K., & Harries, C. (2001). Fast and frugal versus regression models of human judgement. *Thinking & Reasoning*, 7, 5–27.
- Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science*, 1, 107–143.
- Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 103(4), 650–669.
- Gigerenzer, G., Todd, P. M., & the ABC Group. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.
- Hilbig, B. E., & Moshagen, M. (2014). Generalized outcome-based strategy classification: Comparing deterministic and probabilistic choice models. *Psychonomic Bulletin & Review*, 21, 1431–1443.

- Hogarth, R. M., & Karelaia, N. (2007). Heuristic and linear models of judgment: Matching rules and environments. *Psychological Review*, 114, 733–758.
- Katsikopoulos, K. V., & Martignon, L. (2006). Naive heuristics for paired comparisons: Some results on their relative accuracy. *Journal of Mathematical Psychology*, 50, 488–494.
- Lee, M. D. (2016). Bayesian outcome-based strategy classification. *Behavior Research Methods*, 48, 29–41.
- Lee, M. D., & Cummins, T. D. R. (2004). Evidence accumulation in decision making: Unifying the “take the best” and “rational” models. *Psychonomic Bulletin & Review*, 11, 343–352.
- Lee, M. D., & Newell, B. R. (2011). Using hierarchical Bayesian methods to examine the tools of decision-making. *Judgment and Decision Making*, 6, 832–842.
- Lee, M. D., & Zhang, S. (2012). Evaluating the process coherence of take-the-best in structured environments. *Judgment and Decision Making*, 7, 360–372.
- Lichman, M. (2013). UCI machine learning repository. Retrieved from <http://archive.ics.uci.edu/ml>.
- Luan, S., Schooler, L. J., & Gigerenzer, G. (2014). From perception to preference and on to inference: An approach–avoidance analysis of thresholds. *Psychological Review*, 121, 501–525.
- Martignon, L., & Hoffrage, U. (2002). Fast, frugal and fit: Simple heuristics for paired comparison. *Theory and Decision*, 52, 29–71.
- Newell, B. R., Rakow, T., Weston, N. J., & Shanks, D. R. (2004). Search strategies for decision making: The success of ‘success’. *Journal of Behavioral Decision Making*, 17, 117–130.
- Serwe, S., & Frings, C. (2006). Who will win Wimbledon? The recognition heuristic in predicting sports events. *Journal of Behavioral Decision Making*, 19, 321–332.
- Şimşek, O. (2013). Linear decision rule as aspiration for simple decision heuristics. In *Advances in Neural Information Processing Systems* 26. Curran Associates, Inc.
- Şimşek, O., & Buckmann, M. (2015). Learning from small samples: An analysis of simple decision heuristics. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* 28 (pp. 3159–3167). Curran Associates, Inc.
- The University of Toronto. (1996). *DELVE dataset repository*. Retrieved from <http://www.cs.toronto.edu/~delve/data/datasets.html> (The University of Toronto).
- Todd, P. M., & Dieckmann, A. (2005). Heuristics for ordering cue search in decision making Heuristics for ordering cue search in decision making. In L. K. Saul, Y. Weiss, & L. Bottou (Eds.), *Advances in Neural Information Processing Systems* (Vol. 17, pp. 1393–1400). Cambridge, MA: MIT Press.