Redefining heuristics in multi-attribute decisions: A probabilistic framework

Percy K. Mistry[1] **(pkmistry@uci.edu)**
Jennifer S. Trueblood[2] **(jennifer.s.trueblood@vanderbilt.edu)**
Department of Cognitive Sciences, University of California Irvine
[2]Psychological Sciences, Vanderbilt University

Abstract

In this paper, we highlight the shortfall of conventionally described heuristics in multi-attribute decision theory, and propose recasting these heuristics within a novel probabilistic framework. This framework is based on defining a psychological feature space, with rule-based heuristics represented as prototypical representations within this space. We provide various examples of meaningful heuristics that can be constructed under this representation, including recasting probabilistic versions of popular heuristics such as take-the-best. Next, we propose an evaluation framework to measure the effectiveness of a consideration set of heuristics. This framework measures whether the set of heuristics are sufficient to describe, predict and infer strategy selection and learning behavior. We propose that this is a step towards a robust framework within which models of strategy selection and learning should be evaluated. The framework aspires to develop a consideration set of heuristics that can be represented as a mathematically well-posed inference problem. We show that the heuristics redefined under our probabilistic framework generally perform better than conventional heuristics under this evaluation. We conclude with a discussion on the possible applications of this framework.

Introduction

Gigerenzer and Gaissmaier (2011) defined heuristics as *"a strategy that ignores part of the information, with the goal of making decisions more quickly, frugally, and/or accurately than more complex methods."*. They highlighted three key building blocks towards a theoretical framework for how heuristics are constructed:

1. Search rules: How do people explore the search space?
2. Stopping rules: When do people stop searching?
3. Decision rules: How do people make a final decision?

They also proposed that people learn to adaptively select appropriate heuristics depending on the environment. There have been various proposals to formalize this adaptive learning and strategy selection, such as reinforcement-learning of strategies (Rieskamp & Otto, 2006; Rieskamp, 2008; Erev & Barron, 2005) or rational metareasoning (Lieder & Griffiths, 2015). In such approaches, the learning model is implemented by making inferences about how people update their belief about the effectiveness of one or more heuristics. This effectiveness can be based on some measure of accuracy, cost, time, effort, constraint satisfaction, or a combination of such measures. When feedback is available, some measure of accuracy is an especially important factor.

Within the domain of multi-attribute decision making, to be able to predict how people will make decisions, react to feedback, change their decision strategies, or react to changes in the choice architecture, an important step is to infer the generating process that leads to sequential learning and selection of strategies. Most quantitative approaches to strategy selection and learning rely on the assumption that people use and test certain heuristics, and then update their belief about the effectiveness of these heuristics. For a researcher interested in modeling this behavior, one needs to infer (a) which heuristic was used on each decision, and (b) how belief about the effectiveness of this heuristic was updated by the decision maker. The first is an inference problem, where the cause or generating process, h (or to be accurate, an approximation to the true cognitive process) needs to be inferred from observed measurements, x, *on each trial of an experiment*. A brief survey of methods used to infer this in models of learning and strategy selection reveal the following commonly used approaches:

1. By simply checking if the final choices are compatible with the heuristics (Rieskamp, 2008).
2. Based on whether the choices are compatible, and whether all the necessary cues proposed by the heuristic have been searched, but allowing for the search of any extra cues (Rieskamp & Otto, 2006).
3. Based on whether the choices are compatible, and whether the observed cue search pattern *exactly matches* the search proposed by the heuristic.
4. By analyzing aspects such as response time or process tracing (Bergert & Nosofsky, 2007).

In fact, the first method is often the most common approach used, and essentially ignores the information search patterns, which are a key component of how the heuristics are defined. The approach used may indicate either that none of the heuristics in the consideration set were applied, may provide weak evidence for multiple heuristics, or strong evidence for a single heuristic. The learning mechanisms in such models then assume that belief about the heuristic inferred to have been used is updated, based on its effectiveness. This inference about belief updating is only robust if on each trial there is strong evidence for a single heuristic being used. If there is no evidence for any of the heuristics, or weak evidence for multiple heuristics, the model effectively fails to gain any information about the learning on that particular trial. Since the learning process is cumulative over trials, having a large proportion of trials with ineffective learning inferences is likely to significantly hurt a model of learning. It follows that the effectiveness of inferences depends significantly on what set of heuristics a researcher decides to include in their model, and the method used to infer use of heuristics on any particular trial. We show a brief meta-analysis of 7 published experiments involving multi-attribute decision making, in table 1. For each experiment, using we take a consideration

Table 1: Meta-analysis showing the proportion of trials on which inference about TTB and WADD yielded either no match, a match for both heuristics, and a unique match. The analysis is conducted using two of the popular inference methods using minimum cues acquired, and only based on choice compatibility.

		Minimum cues acquired + choice			Compatible choice only		
Experiment	%cues	No match	Multiple match	*Unique match*	No match	Multiple match	*Unique match*
Broder et. al. 1	42%	63%	2%	*35%*	11%	79%	*10%*
Broder et. al. 2	42%	52%	3%	*45%*	9%	84%	*7%*
Mistry et. al. 1	48%	77%	7%	*16%*	11%	79%	*10%*
Mistry et. al. 2	31%	90%	0%	*10%*	10%	80%	*10%*
Rieskamp et. al.	98%	2%	48%	*50%*	10%	45%	*45%*
Lee et. al. 1	84%	9%	37%	*54%*	2%	74%	*24%*
Lee et. al. 2	65%	19%	10%	*71%*	2%	76%	*22%*

set of two heuristics, take-the-best (TTB) and weighted average (WADD). We use two of the most commonly employed inference methods, (1) infer if a heuristic was used based on the minimum cues required for the heuristic being used, plus having a compatible choice selection, and (2) just having a compatible choice selection based on the cues selected, without checking if the cue search satisfied the heuristic requirement. We report the percentage of trials on which inferences could not match either heuristic (no match), those on which both heuristics were deemed compatible (multiple match), and where only one of the two heuristics was inferred to be used (unique match). From a learning model perspective, only the trials with unique match are able to provide any meaningful inference about the strategy learning process. In a set of simulations not included in this paper, we found that inferences about learning process were almost impossible to make when less than 50% of the trials yielded a unique match. Across these 7 experiments, unique matches were found only on about 40% of the trials using the first method (range from 10% to 70%), and on only 16% using the second method (range from 10% to 45%).

In the rest of this paper, we suggest that traditional rule-based heuristics need to be redefined in a probabilistic sense to provide a more meaningful analysis of the cognitive process involved in multi-attribute decision making. We suggest a generalized framework, provide examples, and also provide an evaluation framework for how any consideration set of heuristics should be measured, to ensure it provides a reasonable descriptive and predictive exposition of multi-attribute decisions.

Notation and assumptions

For the purpose of this paper, the task set is restricted to multi-attribute decision making, where an individual selects one of multiple choice options based on some criteria, which depends on a set of multiple attributes associated with each option. For a particular decision, we define n_A as the number of attributes available, n_O as the number of choice options. Thus, there are $n_A n_O$ cues in the choice task. The key observable measures for the m^{th} decision include which cues are searched (s_m), and which choice option is selected (y_m). Note that this is not a comprehensive list, but the framework can be expanded to include other observed measures, for instance, the time taken for each choice. Observed behavior x_m is defined as $x_m = s_m \cap y_m$. Since each cue can either be searched or not, there are $N_s = 2^{n_A n_O}$ unique cue search patterns possible. Thus, $s_m \in \{s_j : j \in [1 : N_s]\}$. Similarly, we have $y_m \in \{y_k : k \in [1 : n_O]\}$. Finally, the use of a latent heuristic is denoted as h_m, where $h_m \in \{h_i : i \in [1 : N_h]\}$, where N_h is the number of distinct heuristics in the consideration set. The probability of observed behavior ($x_m = x_{kj} = s_j \cap y_k$) on the m^{th} decision, conditional on the use of a particular heuristic h_i is defined as:

$$p(x_{kj}|h_i) = p(y_k \cap s_j|h_i) = p(y_k|s_j,h_i)p(s_j|h_i) \qquad (1)$$

Note that $p(y_k \cap s_j|h_i)$ is just a probabilistic representation of the *decision rule* of the heuristic h_i, and $p(s_j|h_i)$ is a probabilistic representation of the *search and stop rules* of the heuristic. It is possible that some observed behavioral patterns are not compatible with any of the heuristics in the considerations set, that is, $\exists(y_k,s_j) : p(y_k \cap s_j|h_i) = 0 \;\forall\; i$. In such cases, observations of such behavioral patterns have been designated either to erroneous application of heuristics, or to a guessing strategy. We define an indicator $I_{kj} = 1$ if the observed pattern (y_k,s_j) is not compatible with any of the heuristics. In the realm of traditionally defined heuristics, we introduce an application error ε_i, which is the probability of making an error conditional on using the heuristic h_i. An error results in selecting a particular cue search and decision pattern that is not compatible with any of the heuristics. If an error is made, the probability of any incompatible behavior is uniformly distributed over all possible incompatible behaviors. Hence we adapt equation 1 to:

$$p(x_{kj}|h_i) = (1-\varepsilon_i I_{kj})p(y_k|s_j,h_i)p(s_j|h_i) \;+\; \frac{\varepsilon_i \, I_{kj}}{\Sigma_{k'}\Sigma_{j'} \, I_{k'j'}} \qquad (2)$$

An alternate way to accommodate incompatible behavioral patterns using traditional heuristics is a guessing strategy (G), which can be defined as having uniform probability over all patterns instead of just over incompatible behavioral patterns:

$$p(x_{kj}|G) = \frac{1}{N_s \, n_O} \qquad (3)$$

We denote $p(G) = g$ and assume equal prior probability of each heuristic, $p(h_i) = \frac{1-g}{N_h}$, and equal application error ($\varepsilon_i = \varepsilon$). Note that these assumptions are not necessary, but enable more compact and intuitive expressions of the properties, and we will continue to use these assumptions in the formal specification. The overall probability of observing a particular behavioral pattern x_{kj} is given by:

$$p(x_{kj}) = \left[\frac{(1-\varepsilon I_{kj})(1-g)}{N_h}\Sigma_i\left\{p(y_k|s_j,h_i)p(s_j|h_i)\right\}\right] + \left[\frac{\varepsilon(1-g)I_{kj}}{\Sigma_{k'}\Sigma_{j'}I_{k'j'}} + \frac{g}{N_s\, n_O}\right] \quad (4)$$

In equation 4, the terms in the first square brackets give the probability of observed behavior based on error-free heuristics, whereas the terms in the second brackets give the probability on account of errors or guessing. For ease of reference, we group the first $p(x'_{kj})$, and second $p(x''_{kj})$ set of terms:

$$p(x_{kj}) = p(x'_{kj}) + p(x''_{kj}) \quad (5)$$

Redefining heuristics

A prototypical rule-based heuristic such as take-the-best may be a reasonable approximation to an underlying cognitive process, however, most implementations assume that any deviation from the rule arises randomly, or based on a uniform distribution of error. Instead, we recast such rule-based heuristics probabilistically, in a new n-dimensional psychological feature space. Each individual search pattern can be represented as a point in this n-dimensional space. We propose that the features defining this space should reflect the cognitive primitives that are represented in any such heuristic. These features could be statistical properties (e.g. proportion of cues searched, sensitivity to validity, variability in cues searched across attributes, search density within selected attributes, etc.), process measures (e.g. time spent, search orders, type 1 versus type 2 search transitions between attribute and option wise search), or other psychological constructs that define search behavior (e.g. confidence, effort, contextual features). In any possible problem, we may consider using a subset of these (or other) features. Behavior arising from the heuristic rule is represented as a prototype point in this feature space. Any other behavioral patterns can be defined in terms of the distance of such a pattern from the prototype within this space. We can then apply kernel based solutions to classify behavior depending on relative proximity to the heuristic prototypes. For this paper, we focus primarily on defining the space for information or cue search behavior, that is, to define $p(s_j|h_i)$, with the final aim of inferring $p(h_i|s_j)$.

Once we select any subset of n features, our next objective is then to define the prototype for each heuristic within this space. If we apply a rule-based heuristic, such as take-the-best, or tallying, we simply calculate the features based on the search behavior proposed by exactly following the heuristic rule. This is straight forward for defining statistical properties, but may require some subjective estimates for aspects such as process measures or psychological constructs. We denote the vector of n features as f, where $f : \{F_1, ...F_n\}$, and the prototype for a particular heuristic h_i is defined by $\bar{f}_i : \{F_1(s_i), ...F_n(s_i)\}$, where s_i is the search pattern obtained by strict application of the heuristic rule. For any behavioral pattern s_j, we define a Gaussian kernel K

$$K(s_j,h_i) = exp\left(\frac{-||f_j - \bar{f}_i||^2}{2\sigma_i^2}\right) \quad (6)$$

This kernel defines a similarity measure in the range $[0,1]$, with a maximum value of 1 when $f_j = \bar{f}_i$. The parameter σ_i defines how rapidly the similarity measure drops off as the search point moves away from the prototype for heuristic h_i in the feature space. Depending on the purpose, this parameter can be designed a priori or post hoc after looking at the data. In the first case, we can select σ by selecting the values that optimize the performance measures P_1 to P_4 (introduced in the subsequent section). In the latter case, σ can be treated as a free parameter to be inferred from the data. This latter treatment is similar to the treatment of kernel based machine learning methods where σ is treated as a free parameter, inferred to minimize the loss function. We then define a probability distribution for observing search patterns in this feature space, conditional on this prototype being used, as in equation 7. Note that this distribution depends on the parameter σ_i. Selecting an infinitesimally small value for $sigma_i$ reduces this representation to the rule-based heuristic, since in this case, $p(s_j|h_i) = 1$ if s_j is the exact search pattern under the heuristic rule, and 0 otherwise. As we increase the value of σ_i, this defines a radial error distribution in the feature space.

$$p(s_j|h_i) = \frac{K(s_j,h_i)}{\Sigma_l K(s_l,h_i)} \quad (7)$$

The next step is to define $p(h_i|s_j)$, or the probability of inferring a heuristic h_i was used based on observing a search behavior s_j. Using Bayes rule, we obtain equation 8, where the subscript k indicates all the different heuristics within the consideration set.

$$p(h_i|s_j) = \frac{1}{1 + \Sigma_{k(k\neq i)}\left\{\frac{K(s_j,h_k)}{K(s_j,h_i)}\frac{\Sigma_l K(s_l,h_i)}{\Sigma_l K(s_l,h_k)}\frac{p(h_k)}{p(h_i)}\right\}} \quad (8)$$

Note that equation 8 presents a straightforward avenue for Bayesian regularization if we apply this recursively for each sequential decision. That is, the ratios $p(h_k)/p(h_i)$ define the ratio of prior probabilities of expecting the use of heuristic h_k versus h_i. On every trial t, we recursively update a prior probability $p(h_{i,t})$ based on observations from trials 1 to $(t-1)$. However, note that $p(h_{i,t}) \neq p(h_{i,\,t-1}|s_{j,\,t-1})$, unless no learning is assumed. Under a learning and adaptive strategy selection model M, we assume that the prior probability on any trial t will be based on some updating of the effectiveness of the heuristic on the previous trial: $p(h_{i,t}) \sim M(x_1 : x_{t-1}, \theta_M)$.

Measurement Framework

A systematic approach to researcher decisions on what consideration set of heuristics and inference rules to use, we propose a measurement framework that formalizes the idea of well-posedness of inference problems of strategy selection learning based on mathematical theory. We specify the operator A (the approximate cognitive process) that maps the space of cognitive heuristics H to the space of observed behaviors X. It is desirable that the operator A^{-1} is strictly a well-posed inverse operator, although it is always going to be an uncertain inference problem. problem faced by a researcher. Kabanikhin (2008) defines an inverse problem as ill-posed, if any one of the conditions below are not met:

1. Existence of a solution: $\exists\,(h \in H)\,\forall\,(x \in X)$. This requires any possible observable behavior (x) should be compatible and explained by at least one of the heuristics ($h \in H$).
2. Uniqueness of the solution: $\exists\, A^{-1} : X \to H$. This requires that each observable behavior (x) should be compatible with only one heuristic (h) within the set H.
3. Stability of the solution: $O(\delta_h) \approx O(\delta_x)$ *"Arbitrarily small errors in the measurement data",* δ_x should not *"lead to indefinitely large errors in the solutions"* δ_h. This requires that small changes in observed behavior (δ_x) should not result in significant changes to the inferred heuristic (h).

This is a hard set of constraints, and would impossible, for any set of cognitive heuristics to satisfy, and A^{-1} is never expected to be a strict well-posed inverse operation. Instead, we treat these as properties that a consideration set of cognitive heuristics should try to maximize. We specify a formal but relaxed interpretation of these criteria for how a set of heuristics should be evaluated:

1. *Existence property (P_1):* We propose a measure of the average probability of observing a behavioral pattern based on error-free application of heuristics, compared to the overall probability of observing it, including on account of errors or guessing, integrated over all possible behaviors and over all possible heuristics in the consideration set. This measure will vary from 0 to 1 with higher values desirable, and a value of 1 implying strict compliance with the existence property:

$$P_1 = \frac{1}{N_s\, n_O} \Sigma_k \Sigma_j \left[\frac{p(x'_{kj})}{p(x'_{kj}) + p(x''_{kj})} \right] \tag{9}$$

2. *Uniqueness property (P_2):* We propose using the generalized Jensen-Shannon divergence (Lin, 1991) for multiple distributions to measure uniqueness of a set of heuristics (equation 10). This divergence is bounded by $[0, log(1+N_h)]$, so we adapt this to the range $[0,1]$. In equation 10, $\mathbb{H}_n(p_n) = -\Sigma_n \left[p_n\, log(p_n)\right]$, refers to the Shannon entropy.

$$P_2 = \frac{1}{log(1+N_h)} \left\{ \mathbb{H}_{kj}\left(\Sigma_i \left[p(h_i) p(x_{kj}|h_i) + \frac{g}{N_s\, n_O} \right] \right) - \Sigma_i \left[p(h_i) \mathbb{H}_{kj}\left(p(x_{kj}|h_i) \right) \right] - g \mathbb{H}_{kj}\left(\frac{1}{N_s\, n_O} \right) \right\} \tag{10}$$

The advantage over commonly used measures of divergence such as the KullbackLeibler divergence, is that $P2$ is smoothed, symmetrical, and can be simultaneously applied over multiple distributions. Further, it also linked directly to both the lower and upper bound on the Bayes probability of error (*BPE*, the lowest possible error rate of a classifier). In our context, this provides the lower and upper bounds for the lowest possible irreducible error in inferring the correct cognitive heuristic.

$$\frac{(\mathbb{H}_p - P_2\, log(1+N_h))^2}{4\, N_h} \leq BPE \leq \frac{\mathbb{H}_p - P_2\, log(1+N_h)}{2} \tag{11}$$

$$\mathbb{H}_p = -\left[(1-g)\, log\left(\frac{1-g}{N_h} \right) + g\, log(g) \right] \tag{12}$$

3. *Stability property (P_3):*
We define a distance metric $d_{kj_1j_2}$ between any pairs (x_{kj_1}, x_{kj_2}) of observed behavioral patterns, as the Euclidean distance based on a value of 1 if a cue is searched and 0 if a cue is not searched, measured over all $(n_A n_O)$ cues. We identify all pairs $\bar{\bar{x}}_{kj_1kj_2}$ that have the lowest possible distance.

$$\bar{\bar{x}}_{kj_1kj_2} = argmin_{(x_{kj_1}, x_{kj_2})}\, d_{kj_1,kj_2} \tag{13}$$

Stability is defined in terms of the mean absolute change in inferring probability of use of a heuristic, given a change in the behavioral pattern, integrated over all heuristics in the consideration set, and measured over all pairs of behavioral patterns that belong to the set $\bar{\bar{x}}_{kj_1kj_2}$, where $n(\bar{\bar{x}}_{kj_1kj_2})$ refers to the cardinality of this set. Stability is in the range [0,1], with higher values indicating higher stability of inferences made about latent heuristics.

$$P_3 = 1 - \frac{\Sigma_{\bar{\bar{x}}_{kj_1kj_2}} \Sigma_i \left| p(h_i|x_{kj_1}) - p(h_i|x_{kj_2}) \right|}{N_h\, n(\bar{\bar{x}}_{kj_1kj_2})} \tag{14}$$

4. *Predictiveness property (P_4):*
In addition to be well-posed, the heuristics should have a high degree of predictive capability, in that, *once it is inferred which heuristics is being used*, it should be capable of making strong inferences about what search patterns and choice options will be selected. To illustrate, the guessing strategy above has no predictive capability, since it accords equal probabilities to all observed behaviors. To measure predictiveness, we base it on the within-heuristic entropy across all possible search and decision patterns, integrated across all heuristics. This also yields a value in the range $[0,1]$. Note that this measure has to be read in tandem with P_1 and P_2.

$$P_4 = \frac{(1-g)}{N_h} \Sigma_i \left[1 + \frac{\Sigma_k \Sigma_j \left(p(x_{kj}|h_i)\, log(p(x_{kj}|h_i)) \right)}{log(N_s\, n_O))} \right] \tag{15}$$

Examples of probabilistically redefined heuristics

Example 1: Single dimensional feature space

We start with the simplest example, using only one feature of cue search patterns. We base this on the experimental paradigm used in Lee, Newell, and Vandekerckhove (2014). Here, participants on each trial have access to 9 different cue attributes for two choice options. The cue validities are known to participants and attributes can only be selected in order of their validity, and selecting an attribute reveals the value for both choice options. Thus, there are only 9 different possible cue search patterns, from selecting 1 attribute to 9 attributes. Traditional TTB search patterns would imply selecting the minimum numbers of attributes required to discriminate between the two choices, and WADD would imply selecting all the attributes. The original paper represents the proportion of extra cues (PEC) searched incremental to the first discriminating cue (FDC).

$$PEC = \frac{N_{cues} - FDC}{N_{total} - FDC} \tag{16}$$

We define the 1-dimensional psychological space in terms of this PEC feature, which varies discretely in the range [0,1]. A value of 0 is representative of TTB, and 1, of WADD. Hence, for a particular search pattern s_j, we can write:

$$K(s_j, h_{ttb}) = exp\left(\frac{-||PEC_j||^2}{2\sigma_{ttb}^2}\right) \tag{17}$$

$$K(s_j, h_{wadd}) = exp\left(\frac{-||PEC_j - 1||^2}{2\sigma_{wadd}^2}\right) \tag{18}$$

Example 2: Multidimensional feature space

We define a 3-dimensional psychological feature space, $f_j : [F_1(s_j), F_2(s_j), F_3(s_j)]$, so that any search pattern s_j can be expressed in terms of these features and represented as f_j.
F_1 : Proportion of cues searched
F_2 : Sensitivity to cue validity
F_3 : Variability in cues selected across attributes
We use the experimental structure of (Bröder & Schiffer, 2003), and take the cue search pattern suggested by an original heuristic, for instance, TTB, s_{ttb}), and calculate the feature vector corresponding to that search pattern, denoting this as f_{ttb}. The squared Euclidean distance $||f_j - f_{ttb}||^2$ is calculated as $\Sigma_{n=1:3}\left(F_n(s_j) - F_n(s_{ttb})\right)^2$.

$$K(s_j, h_{ttb}) = exp\left(\frac{-\Sigma_{n=1:3}\left(F_n(s_j) - F_n(s_{ttb})\right)^2}{2\sigma_{ttb}^2}\right) \tag{19}$$

For example, for a cue space involving 4 attributes and 3 choice options, we get $f_{ttb} = [0.25, 1.00, 0.50]$ and $f_{wdd} = [1.00, 0.25, 0.00]$. In table 2 we show the evaluation metrics for a pair of heuristics defined in the sense of TTB and WDD. The first half of the table shows the metrics for conventionally defined heuristics, with different values of ε and g. The second half shows the kernel based probabilistic heuristics sets, with different values of σ. Conventional heuristics find it hard to find a balance between existence P_1 and uniqueness P_2. Probabilistic heuristics as defined here show a better balance, and overall improvement in scores.

Note that the kernel specification essentially assumes equal weights on the three features. However, it is entirely feasible that individuals pay differential attention to these features. We can accommodate such individual differences by defining a mixture of kernels. Here the subscript p refers to the p^{th} individual, and w_{np} refers to the weight placed on the n^{th} feature by the p^{th} individual. These can be treated as free parameters during the inference process. This shows how this framework can be used to infer individual differences in attention to different aspects of the search space. Effectively, this skews the n-dimensional space by scaling each feature dimension by its corresponding weight, allowing for different scaling by individuals.

$$K(s_j, h_{pi}) = \Sigma_{n=1:3}\left[w_{np}\ exp\left(\frac{-\left(F_n(s_j) - F_n(s_i)\right)^2}{2\sigma_i^2}\right)\right] \tag{20}$$

Table 2: Evaluation metric for traditional and probabilistic pairs of (TTB,WADD) heuristics. Exact, minimum, and choice refer to the method of inference about heuristics explained in the introduction. The ε and gparameters used here are highly optimistic, and post hoc error rates are likely to be higher, leading to a further worsening of evaluation metrics.

		P_1	P_2	P_3	P_4	$\Sigma_n P_n/4$
Conventionally defined TTB - WADD heuristics						
exact	$\varepsilon = 0.01$	0.0	0.61	0.99	0.98	*0.65*
min	$\varepsilon = 0.01$	0.02	0.55	0.99	0.74	*0.58*
choice	$\varepsilon = 0.01$	0.40	0.06	1.0	0.12	*0.40*
exact	$g = 0.01$	0.0	0.67	0.99	0.98	*0.66*
min	$g = 0.01$	0.02	0.60	0.99	0.75	*0.59*
choice	$g = 0.01$	0.40	0.09	1.0	0.12	*0.40*
Based on the proposed kernel density framework						
prob	$\sigma = 0.01$	0.03	0.63	1.0	1.0	*0.67*
prob	$\sigma = 0.02$	0.86	0.63	0.82	1.0	*0.83*
prob	$\sigma = 0.05$	0.99	0.63	0.80	0.99	*0.86*
prob	$\sigma = 0.20$	0.99	0.59	0.80	0.35	*0.69*

Example 3: Context specific feature space

These example were based on traditional heuristics and statistical properties of the cue search patterns. However, since we generalize heuristics in terms of a psychological feature space, we can define heuristics in a context-specific manner. For instance in (Mistry & Trueblood, 2015), two of the attributes were financial metrics, while the remaining two were advisory recommendations. It is perfectly reasonable that people search the attribute space based on their own level of expertise and prior perceptions of advisory versus financial cue attributes. We can define feature in terms of proportion of advisory cues searched compared to total cues searched. A full advisory heuristics will be represented by a feature value of 1, and a full financial heuristic by a feature value of 0. The kernel and probability calculations can then proceed as in the previous examples. Results of evaluation metrics for context specific heuristics are provided in table 3.

Table 3: Evaluation metric for traditional and probabilistic pairs of context defined (Financial,Advisory) heuristics.

		P_1	P_2	P_3	P_4	$\Sigma_n P_n/4$
Conventionally defined context - heuristics						
exact	$\varepsilon = 001$	0.0	0.13	0.99	0.97	*0.52*
min	$\varepsilon = 0.01$	0.0	0.13	0.99	0.97	*0.52*
choice	$\varepsilon = 0.01$	0.52	0.11	1.0	0.19	*0.45*
exact	$g = 0.01$	0.0	0.18	0.99	0.97	*0.54*
min	$g = 0.01$	0.02	0.18	0.99	0.97	*0.54*
choice	$g = 0.01$	0.52	0.13	1.0	0.19	*0.46*
Based on the proposed kernel density framework						
prob	$\sigma = 0.01$	0.43	0.63	1.0	0.65	*0.68*
prob	$\sigma = 0.02$	0.82	0.63	0.81	0.65	*0.73*
prob	$\sigma = 0.05$	0.82	0.63	0.81	0.58	*0.71*
prob	$\sigma = 0.20$	0.82	0.53	0.82	0.23	*0.60*

Conclusions

The evaluation framework should be seen as a first step towards a unified and systematic approach to defining strategy selection and learning models. The probabilistic framework is generalized enough to be applicable to a variety of experimental and empirical designs and heuristics. It can easily be incorporated with existing approaches to learning, such as rational meta-reasoning, reinforcement-learning, and cost-benefit based or cognitive effort based frameworks. Importantly, it has the potential to unify rule-based and exemplar based heuristic models. The heuristics described under our framework generally perform better than conventional heuristics under the proposed evaluation measures. This framework raises a lot of possibilities in terms of future work, including experimental design based on maximizing information gain, and generating a new class of heuristics based on context specific, process driven, or exemplar measures. The evaluation framework allows us to calculate the a priori performance measures of a set of kernel based heuristics, for each particular configuration of cues. This evaluation method can be used to for experimental design to select cue configurations that allow for the strongest possible inference given a particular set of heuristics to be tested, that is, by selecting configurations that maximize measures P_1 to P_4 for the heuristics to be tested. Note that P_1 and P_2 often competethe first two conditions often compete. Since these are all measured on the same scale $[0,1]$, we can use an objective function that optimizes a weighted average of the four measures. The examples here rely on defining heuristic based on a prototypical feature set that is derived from the rule-based heuristics. It has been proposed that people may also approach multi-attribute decisions using exemplars (Juslin, Olsson, & Olsson, 2003). The feature space lends itself naturally to defining multiple exemplars, with the kernel specification defining the typicality of a cue search pattern from a particular exemplar. Since our heuristics are defined in terms of kernel densities, we can simply used kernel-based clustering mechanisms popular in machine learning literature to identify clusters within a feature space (Girolami, 2002), with each cluster corresponding to an exemplar. In future work, this method can be used to identify common search exemplars used by people without a priori definition of heuristics.

References

Bergert, F. B., & Nosofsky, R. M. (2007). A response-time approach to comparing generalized rational and take-the-best models of decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(1), 107.

Bröder, A., & Schiffer, S. (2003). Bayesian strategy assessment in multi-attribute decision making. *Journal of Behavioral Decision Making*, *16*(3), 193–213.

Erev, I., & Barron, G. (2005). On adaptation, maximization, and reinforcement learning among cognitive strategies. *Psychological review*, *112*(4), 912.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual review of psychology*, *62*, 451–482.

Girolami, M. (2002). Mercer kernel-based clustering in feature space. *IEEE Transactions on Neural Networks*, *13*(3), 780–784.

Juslin, P., Olsson, H., & Olsson, A.-C. (2003). Exemplar effects in categorization and multiple-cue judgment. *Journal of Experimental Psychology: General*, *132*(1), 133.

Kabanikhin, S. I. (2008). Definitions and examples of inverse and ill-posed problems. *Journal of Inverse and Ill-Posed Problems*, *16*(4), 317–357.

Lee, M. D., Newell, B. R., & Vandekerckhove, J. (2014). Modeling the adaptation of search termination in human decision making. *Decision*, *1*(4), 223.

Lieder, F., & Griffiths, T. L. (2015). When to use which heuristic: A rational solution to the strategy selection problem. In *Proceedings of the 37th annual conference of the cognitive science society* (pp. 1362–1367).

Lin, J. (1991). Divergence measures based on the shannon entropy. *IEEE Transactions on Information theory*, *37*(1), 145–151.

Mistry, P. K., & Trueblood, J. S. (2015). Reconstructing the bayesian adaptive toolbox: Challenges of a dynamic environment and partial information acquisition. *X*, *X*(X), X.

Rieskamp, J. (2008). The importance of learning when making inferences. *Judgment and Decision Making*, *3*(3), 261–277.

Rieskamp, J., & Otto, P. E. (2006). Ssl: a theory of how people learn to select strategies. *Journal of Experimental Psychology: General*, *135*(2), 207.