Coarse, efficient decision-making

Michael Mandler*

Royal Holloway College, University of London

This version: November 2017

Abstract

To achieve efficiency in decision-making, an agent should use coarse criteria with a small number of categories to discriminate among alternatives. Coarse criteria reduce decision-making cost even though an agent, to maintain the number of choice distinctions, must use more criteria. The most efficient criteria are binary with two categories each, even in cases where the marginal cost of using additional categories diminishes to 0. That coarse criteria are used in practice can therefore be explained as a result of optimization rather than cognitive limitations. Binary criteria also generate choice functions that maximize rational preferences, thus linking efficiency to rational choice. The efficiency-of-binary-criteria principle applies to information storage: binary digits store information more efficiently than k-ary digits for any $k > 2$.

JEL codes: D01

Keywords: decision-making efficiency, criteria, rationality, information, bits.

*Address: Department of Economics, Royal Holloway College, University of London, Egham, Surrey TW20 0EX, United Kingdom. Email: m.mandler@rhul.ac.uk. I thank Sophie Bade, Salvador Barberà, Richard Bradley, Drew Fudenberg, Itzhak Gilboa, Raphaël Giraud, Kim Kaivanto, Todd Kaplan, David Kelsey, Michel De Lara, Christian List, Itai Sher, Tymon Tatur, and Peter Wakker for helpful comments and suggestions.

1 Introduction

In the psychology implicit in traditional preference theory, agents consider each pair of alternatives, gauge their reaction, and come to a preference judgment. If each pair of alternatives needs a separate costly judgment, a preference relation would be expensive to produce: a preference over n alternatives would require $\binom{n}{2} = \frac{n(n-1)}{2}$ judgments, a number that grows quickly as a function of n.[1]

This paper proposes a solution to this problem where agents make choices by consulting criteria that order categories of alternatives in various ways. To choose from a set of movies, for example, an agent could consult a first criterion that classifies and orders movies by genre (drama, comedy, documentary, all others), a second that classifies by type of director (commercial, arty, all others), a third by actor quality (famous, not famous), and so forth. Although we will not assume that criterion rankings or an agent's choices are rationally ordered, it turns out that if agents choose criteria efficiently then their choices will maximize a rational preference.

Sets of criteria enjoy the advantage that each criterion can discriminate within each set of alternatives that other criteria fail to rank, e.g., an actor criterion can distinguish among the movies of a particular genre and director type. Consequently the number of choice distinctions that criteria can generate equals the *product* of the number of criterion categories. If an agent uses a genre criterion with 4 categories, a director criterion with 3 categories, and an actor criterion with 2 categories, with decisions made by a weighted vote of the three criteria, then the agent will make $4 \times 3 \times 2 = 24$ distinctions among types of movies. But the number of pairs of categories that need to be ranked, $\binom{4}{2} + \binom{3}{2} + \binom{2}{2} = 10$, is small. If in contrast an agent had to compare each pair of 24 types of movies without the aid of criteria, $\binom{24}{2} = 276$ decisions would be required.

What sort of criteria should an agent use? If an agent's 'true' preferences could be inferred from a long list of fine criteria, then the agent would be better off when more of these criterion orderings are uncovered: the agent will be able to choose superior alternatives from more choice sets. But criterion rankings, like direct preference judgments, are costly to form or discover; they require time and mental attention. While an agent could undertake enough research to make a director criterion a fine discriminator – by reading reviews, checking whose movies show at Cannes, etc. – the benefits must be weighed against the cost of the research. In line with Herbert Simon's theories of bounded rationality, agents do not and should not try to extract all of the rankings that a criterion could in principle make.

[1] Even rational preferences require a number of judgments, $n \log_2 n$, that increases at a rate greater than n. While the cost of constructing a preference or choice function is distinct from the complexity of representing one, Apesteguia and Ballester (2010) provides consonant results.

Since increases in the number of preference or choice distinctions and reductions in decision-making cost are unambiguous benefits, an efficient arrangement consists of a set of criteria and corresponding choice function that are undominated with respect to these two goals. The first goal requires a new concept – choice classes – to measure the number of choice distinctions.

The preference-discovery problem seems to present a trade-off: fine criteria that make many rankings are more costly per-criterion but – fixing the number of choice distinctions – they allow the agent to employ fewer criteria. Should the agent use a small list of fine orderings or a large list of coarse orderings?

The answer will be that the trade-off should always be resolved in favor of coarse criteria, even when agents aim for a large number of choice distinctions. Optimality is therefore reached in the 'coarseness limit' where agents deploy only binary criteria, which partition alternatives into two categories. This result holds even if the marginal decision-making cost of using a category diminishes to 0: the additional categories of fine criteria could become asymptotically free and still it would be optimal to use the expensive categories of coarse criteria. To test the robustness of this conclusion, the paper's main result, I consider an alternative model where criteria vary in value and cost that covers the evaluations of criteria that utility maximizers would make. While optimality then does not require criteria to be binary, they must still be coarse. Even a high-value criterion with a marginal cost of categories that descends to 0 should not become too fine: it would be more efficient to use many low-value coarse criteria despite the greater marginal cost of categories that an agent would have to pay. And if we tilt the playing field by letting fine criteria have greater value than they would enjoy under utility maximization, the optimal criteria remain coarse.

The efficiency advantage of coarse criteria applies to seemingly distant questions including 'how many states or digits should obtain in information storage devices in order to store information efficiently?' I will show that, even when the marginal cost of using larger digits in a memory device converges to 0, the coarsest option – binary digits or bits – offers the cheapest storage method. Bits of course are what are used in practice.

In the preference setting, the efficiency of coarse criteria matches the psychological finding that people can readily manipulate only a small number of categories. Agents may find that even four categories, which requires six category comparisons, are unwieldy. Rather than an unfortunate limitation, this feature of human information-processing may be an outgrowth of optimization. Since our inability to handle more than a few categories forces us into efficiency, it may not have been vital to develop a capacity to manipulate many categories at once. More broadly, I hope to show that optimization in choice theory does not have to taken preferences as given; it applies to preference discovery as well.

Binary criteria lead to rational choices when criteria are aggregated in standard ways. For example, if an agent proceeds through binary criteria lexicographically[2] then the choice function that results maximizes a rational preference, as shown in Mandler et al. (2012). What these authors failed to understand is that the link between binary criteria and rational choice holds for a wide range of criterion-based choice procedures. I show here that a rational choice function will arise whenever criteria are binary and decisions satisfy axioms that model and generalize a weighted vote of criteria.[3]

Since criterion-based choice is a version of multicriterion decision-making,[4] our results offer a reply to Arrow and Raynaud's (1986) concern that aggregating criteria with ordinal voting rules will lead to irrational decisions. Multicriterion decision-making takes criteria to be exogenous, but if criteria are chosen to minimize decision-making cost then the problem of irrationality becomes less serious.

Coarser is better To illustrate the 'coarser is better' principle, suppose the number of choice distinctions – *choice classes* in our terminology – that an agent makes equals the product of the number of categories in the criteria deployed (which, as in the movie example, is the maximum number of choice classes that can be generated given the number of categories in each criterion). Suppose the jth of a set of N criteria uses $e_j > 2$ categories and we replace it with a criterion that uses $e_j - 1$ categories. If the number of categories in the other criteria remains unchanged, then, to avoid a drop in the number of choice classes, the agent must deploy an additional criterion. If the added criterion uses the minimum nontrivial number of categories, $e_{N+1} = 2$, then the difference between the number of choice classes created by the new set of criteria and the original set is

$$\left(\prod_{i \neq j} e_i\right)(e_j - 1)\,2 - \left(\prod_{i \neq j} e_i\right) e_j = \left(\prod_{i \neq j} e_i\right)(e_j - 2) > 0.$$

So the new set of criteria produces more choice distinctions than the original set.

What is the cost of the new, coarser set of criteria? Assuming that the first category of a criterion is costless (a single category requires no decisions and makes no choice distinctions), the total number of costly categories used in the two sets of criteria is the same. So if the marginal cost of using additional categories in a criterion is increasing, the shift to the coarser set of criteria strictly lessens costs.

[2] That is, the agent consults the criteria in sequence and at each stage eliminates from consideration any alternative that is defeated by some alternative still in contention.

[3] The rationality of choice that holds when binary criteria are lexicographically aggregated is in fact a special case of the general rationality result for weighted voting.

[4] See Figueira et al. (2005) and Bouyssou et al. (2006) for overviews.

Coarser criteria thus deliver two distinct benefits: they increase the number of choice distinctions and reduce costs. The benefits of binary criteria – the coarsest possible – persist even when the marginal cost of using additional categories is diminishing.

The binariness-rationality connection The Condorcet paradox provides a familiar example of how even rational criteria can lead to irrational choices. Let there be three alternatives x, y, z and three criteria C_1, C_2, C_2 defined as follows (with higher alternatives better than lower):

$$\begin{array}{ccc} \underline{C_1} & \underline{C_2} & \underline{C_3} \\ x & y & z \\ y & z & x \\ z & x & y \end{array}.$$

If the choice function c decides by a simple majority vote of the criteria, then choices will cycle on the pairs – $c(\{x, z\}) = \{z\}$, $c(\{y, z\}) = \{y\}$, $c(\{x, y\}) = \{x\}$ – and therefore cannot be rationalized. Now suppose instead that criteria are binary: each criterion ranks two of the alternatives above the remaining option, or ranks one alternative above the other two. Given a set of binary criteria and a choice set of alternatives, the option that lies in the greatest number of top categories will defeat any other alternative in a majority vote. Moreover, since the ordering that ranks each alternative a by the number of criteria that place a in the top category is complete and transitive, choices based on majority vote will maximize a rational preference.[5] I will generalize considerably in section 5.

The psychology of categorization The 'coarser is better' conclusion connects to the psychological literature on information processing, which finds that the number of categories that people can retain in working memory is quite small. In our setting, an agent who deploys a criterion has to hold in mind the category comparisons that the criterion requires. Miller (1956) famously concluded that the number of 'chunks' that an agent can hold in mind is roughly seven and since Miller the number has been steadily whittled down. Herbert Simon (1974) argued that five is more accurate. A binary comparison of categories qualifies as an object-file in the model of Kahneman et al. (1992), and Treisman (2006) judges that subjects can hold only three or four object-files in memory. An encyclopedic overview of the evidence, Cowan (2000), concludes that the 'magic number' that bounds working memory is four.[6] Since, for a criterion with e categories, the number of pairwise category comparisons is $\frac{e(e-1)}{2}$, a bound of four on the number of binary comparisons would give a bound of three on e. The psychological literature therefore suggests that a criterion

[5] This result, but not the generalizations in section 5, arises in the voting literature on dichotomous preferences. See Inada (1964), Vorsatz (2007), Ju (2011), and Maniquet and Mongin (2015)).

[6] See Luck and Vogel (1997) for a characteristic example of the research surveyed.

that needs to be manipulated in working memory could have at most three or four categories. Consider the movie example: if an agent wants to choose a movie with a genre criterion that divides movies into 5 categories then he or she would have to keep 10 category comparisons in mind, which indeed seems unwieldy.

Unlike the psychological literature, I will stress the efficiency advantage of coarse criteria. Since decision-making becomes more efficient as the number of categories per criterion shrinks, the cognitive constraints that limit the number of categories in decision-making might be the outcome of optimization or adaptation. The binary criteria that use two categories are especially prevalent in everyday decision-making, and their efficiency may help to explain this fact. Our results also give formal support to Gigerenzer et al.'s (1999) view on the superiority of frugal heuristics.

Related literature My emphasis on the absence of known ex ante preferences and the costs of decision-making owes a great debt to Herbert Simon (e.g., Simon (1972)). But one conclusion offered here diverges from the Simon program: paying attention to decision-making cost leads agents to rationality. This message complements Mandler (2015), which argues that if agents proceed lexicographically through criteria then only rational preferences can always be the outcome of a 'quick' sequence of criteria no matter how the numbers of categories in criteria are fixed. Agents in this paper choose their own categorization levels to minimize decision-making cost (and lexicography is dropped) and again rationality enjoys an efficiency advantage. Despite the common conclusion, the arguments used have no overlap.

Choice functions generated from a set of criteria have been extensively researched. See Apesteguia and Ballester (2010, 2013) (AB), Houy and Tadenuma (2009), Mandler et al. (2012), Mandler (2015), and Manzini and Mariotti (2007, 2012). The emphasis in AB (2010) on the cost of rational choice relates the most closely to the present paper. Some of the above work has a precedent in the lexicographic utility theory of Chipman (1960, 1971) and Fishburn (1974). Tversky and Simonson (1993) and Salant (2009) also link efficiency to rational decision-making.

2 Choice via criteria

We set a domain of alternatives X with at least two elements. A **criterion** C_i is an asymmetric binary relation on X. Each criterion that an agent uses can be 'irrational', i.e., fail to be negatively transitive. A **set of criteria**, denoted $\mathcal{C} = \{C_1, ..., C_N\}$, must have finitely many (typically N) criteria. Criterion indices do not indicate the order in which criteria are consulted.

To consider the efficiency of decisions based on criteria, we need measurement units for both criteria and choices. Beginning with criteria, we define a criterion's categories or equivalence classes to be the mutually exclusive subsets of X that C_i orders.

Definition 1 *The **equivalence relation** I_i of a criterion C_i is defined by*

$$x\, I_i\, y \Leftrightarrow \{z \in X : z\, C_i\, x\} = \{z \in X : z\, C_i\, y\} \text{ and } \{z \in X : x\, C_i\, z\} = \{z \in X : y\, C_i\, z\},$$

*and a C_i-**category** is a maximal $E \subset X$ such that, for all $x, y \in E$, $x\, I_i\, y$.*[7]

Alternatives x and y are C_i-equivalent if they are treated as indistinguishable by C_i: the C_i relationship between x and any z must coincide with the C_i relationship between y and z.

A criterion will typically divide X into fewer categories than the number of indifference classes of a preference (or the number of choice classes of a choice function). For a variation on the movie example, X could be a set of vacation destinations described by a list of attributes – e.g., climate, amenities available – with each attribute ordered by a criterion. Criteria can be 'incomplete' as well as intransitive: a C_i might not rank every pair of C_i-categories. Since counts of the number of categories in an incomplete binary relation can be controversial, readers may want to assume that each C_i ranks every pair of its categories; no changes in the paper would be introduced. Nonrational choice functions can arise even when each criterion is complete and transitive as well, as in the Condorcet example in the introduction.

Although the categories of a C_i and its rankings are formally intertwined, an agent would presumably begin by describing an attribute of X, for example, the climate of vacation destinations. The agent at this stage decides whether to do the research needed to distinguish between hot and sweltering, and between liable to rain and liable to pour, and so forth. The agent then orders these categories pair by pair. Outside of section 4.3, we will not require explicitly that each C_i orders a distinct attribute and that X equals a product of attributes, but each use of criteria in the paper is consistent with these assumptions.

We will use e_i or $e(C_i)$ to denote the number of categories in a criterion C_i and consider C_i to be *coarser than* C_j if $e_i < e_j$. As the formation of categories is costly, we require that each e_i is finite. Since criteria with a single category require no decisions and distinguish between no pair of alternatives, it is natural to assume they are costless.

Turning to the choices that criteria can generate, let c be a choice function defined on a family $\mathcal{F}$ of choice sets (subsets of X) that includes all the two-element sets: for every $A \in \mathcal{F}$, $c(A)$ is a nonempty subset of A. We use $x \in c(A)$ to mean *both* that x is in $c(A)$ and that $A \in \mathcal{F}$. The

[7] Equivalently, a nonempty $E \subset X$ is a C_i-category if, for all $x \in E$, $x\, I_i\, y \Leftrightarrow y \in E$. When C_i is transitive, see Fishburn (1970) and Mandler (2009) for discussions of I_i.

choice classes of c are defined comparably to the equivalence classes of criteria: two alternatives are in the same choice class if c treats them as interchangeable in every sense.[8]

Definition 2 *Given a choice function c, alternatives x and y are elements of the same* ***choice class*** *if and only if for all $A \subset X$,*

(i) if $\{x, y\} \subset A$ then $x \in c(A) \Leftrightarrow y \in c(A)$,

(ii) if $\{x, y\}$ does not intersect A then

$$\begin{aligned} x \in c(A \cup \{x\}) \quad &\Leftrightarrow \quad y \in c(A \cup \{y\}), \\ z \in c(A \cup \{x\}) \quad &\Leftrightarrow \quad z \in c(A \cup \{y\}), \text{ for all } z \in A. \end{aligned}$$

So x and y are in the same choice class if (i) when x is chosen and y is available then y is chosen too and (ii) when x is substituted for y then x is chosen if y was chosen previously with no effect on whatever other alternatives are chosen. When choices are determined by preferences, each choice class will be an indifference class. The important consequence of Definition 2 is that the choice classes form a partition of X (proved in the Appendix).

When a choice function is determined by criteria, selections must depend only on the distinctions the criteria make. In particular, if a pair of alternatives x and y is in the same criterion category for every C_i then the agent has no way to distinguish x and y and so the agent's choice function should deem x and y to be interchangeable, i.e., in the same choice class.

Definition 3 *A choice function c* ***uses the set of criteria*** *$\mathcal{C}$, which we denote by $(\mathcal{C}, c)$, if whenever $x, y \in X$ are contained in the same C_i-category for each $C_i \in \mathcal{C}$ there is a choice class of c that contains x and y.*

In the movie example at the beginning of the paper, two movies that fall into the same genre, director, and acting categories must be in the same choice class when choices use these criteria.

Example 1 The leading case of a choice function that uses a set of criteria compares alternatives via a weighted vote of the criteria. Given a set of criteria $\mathcal{C}$, for any pair $x, y \in X$, set

$$s_i(x, y) = \begin{cases} 1 & \text{if } x\, C_i\, y \\ -1 & \text{if } y\, C_i\, x \\ 0 & \text{otherwise} \end{cases},$$

and let the weight assigned to the criterion C_i equal ω_i. The sum of the weighted votes for alternative x, when $x \in A$, is given by $v(x, A) \equiv \sum_{y \in A} \sum_{i \in \{1,\ldots,N\}} \omega_i s_i(x, y)$. Then $v(x, A) = v(y, A)$ if

[8] I initially assumed that the concept of 'choice class' must already exist in the literature but I have not been able to find a precedent.

$x, y \in A$ and x and y are contained in the same C_i-category for each $C_i \in \mathcal{C}$. Consequently the choice function c defined by

$$x \in c(A) \Leftrightarrow (v(x, A) \geq v(y, A) \text{ for all } y \in A)$$

uses $\mathcal{C}$. The majority vote of criteria in the introduction amounts to a special case of this c in which all of the ω_i are equal; here some criteria can be more important than others by having larger weights. ■

The formal definition of weighted voting in section 5 will encompass Example 1 but generalize by remaining agnostic about what alternative is chosen when no alternative defeats all of its competitors in A in pairwise weighted votes.

3 The optimization problem

We identify an agent's 'true' preferences with the ordering of the choice classes the agent could form if he or she knew all of the possible distinctions among criterion categories. The agent will then be better off if he or she learns more criteria and categories: the agent can identify more of these choice classes and thus make better decisions from more choice sets. For example, when facing $\{x, y\}$ it is only when some criterion distinguishes x and y that the agent can place the items in different choice classes and choose the better option. Section 4.3.1 will link the goal of increasing the number of choice classes to classical utility maximization.

Our agents also have the second goal of decreasing their decision-making cost.

Let $\kappa(C_i)$ denote the cost of criterion C_i. We assume throughout that, for any criterion C_i, $\kappa(C_i) \geq 0$. Until an explicit warning to the contrary in section 4.3, costs will be determined by the number of C_i-categories: for all criteria C_i and $\widehat{C}_i$, $e(C_i) = e(\widehat{C}_i) \Rightarrow \kappa(C_i) = \kappa(\widehat{C}_i)$. The **cost of a set of criteria** $\mathcal{C} = \{C_1, ..., C_N\}$, denoted $\kappa[\mathcal{C}]$, is the sum $\sum_{i=1}^{N} \kappa(C_i)$.

Given a choice function c, let $n(c)$ be the number of choice classes in c. Remember that the notation $(\mathcal{C}, c)$ means that c uses $\mathcal{C}$.

Definition 4 *The pair $(\mathcal{C}, c)$ is* ***more efficient*** *than the pair $(\mathcal{C}', c')$ if*

$$n(c) \geq n(c') \text{ and } \kappa[\mathcal{C}] \leq \kappa[\mathcal{C}'],$$

and one of the above inequalities is strict. The set of criteria $\mathcal{C}$ is ***more efficient*** *than $\mathcal{C}'$ if there exists a c that uses $\mathcal{C}$ such that $(\mathcal{C}, c)$ is more efficient than $(\mathcal{C}', c')$ for any c' that uses $\mathcal{C}'$. A set of criteria $\mathcal{C}$ (resp. pair $(\mathcal{C}, c)$) is* ***efficient*** *if there does not exist a more efficient $\mathcal{C}'$ (resp. $(\mathcal{C}', c')$).*

The advantage of criteria is that each criterion can discriminate within every set of alternatives that the other criteria fail to rank, e.g., the genre criterion for movies will discriminate within each director category. The number of choice distinctions can then equal the product of the number of categories in the criteria (if that number does not outstrip the cardinality of X).

Definition 5 *The pair* $(\mathcal{C}, c)$ ***maximally discriminates*** *if the number of choice classes of c equals* $\min\left[\prod_{i=1}^{N} e(C_i), |X|\right]$.

Proposition 1 *If* $(\mathcal{C}, c)$ *is efficient then* $(\mathcal{C}, c)$ *maximally discriminates.*[9]

If $(\mathcal{C}, c)$ is efficient then $n(c) \geq n(c')$ must hold for any $(\mathcal{C}', c')$ that satisfies the constraints that $\mathcal{C}'$ has the same number of criteria as $\mathcal{C}$ and $e(C_i') = e(C_i)$ for all i (since then $\kappa[\mathcal{C}'] = \kappa[\mathcal{C}]$). To see when $n(c)$ reaches a maximum subject to these constraints, consider some $(\mathcal{C}, c)$ and the sets of alternatives equal to an intersection of C_i-categories, one category per criterion. Since alternatives in different choice classes must be distinguished by at least one criterion, the number of choice classes cannot exceed the number of these intersections. The number of intersections is in turn bounded by the product $\prod_{i=1}^{N} e(C_i)$ and criteria can always be chosen to reach this bound. Moreover the bound is necessarily achieved when X is a product of attributes and each C_i orders a distinct attribute. Our examples all enjoy this product feature. Recall that in the movie case, three criteria with 4, 3, and 2 categories can distinguish $24 = 4 \times 3 \times 2$ types of movies: each type equals the intersection of one genre category, one director category, and one actor category. Products of attributes in fact form the prototype of all cases of maximal discrimination: when criteria maximally discriminate, the alternatives can always redescribed as a product of attributes.[10] A choice function does not have to designate each intersection of C_i-categories to be a choice class – an agent might decide to ignore a criterion that, say, categorizes foods by color when the agent cares only about taste – but each intersection will form a distinct choice class when decisions are made generic weighted-voting choice functions.

An efficient $(\mathcal{C}, c)$ must maximally discriminate regardless of what assumptions are placed on the cost of criteria. Costs do come into play in the determination of the optimal number of criteria and their optimal coarseness, which we consider next.

[9] Proofs omitted from the text are in the Appendix.

[10] When some C_i does not order a distinct attribute or X is not a product of attributes, it is possible for the number of intersections of C_i-categories to be strictly less than $\min\left[\prod_{i=1}^{N} e(C_i), |X|\right]$, for example, when two criteria in $\mathcal{C}$ simply repeat each other.

4 The optimality of coarse criteria

Under mild assumptions, it is always more efficient to use coarser criteria that distinguish fewer categories. Maximum efficiency is therefore achieved by sets of criteria with just two categories each, the minimum nontrivial number, and for this result the needed assumptions are milder still. A criterion is **binary** if it has two categories.

The formation of a criterion requires an agent to partition X into categories and then order these categories. Since criteria with a single category require neither partitioning nor ordering and make no choice distinctions, we suppose for discussion purposes that they are free, and use this assumption explicitly in section 4.4. Any expansion of the number of categories e beyond 1 would incur both a partitioning and an ordering cost. While the former might be a linear function of the number of categories, ordering will require an agent to decide, for any pair of categories, if they are ranked and if so which is superior. If e is the number of categories, the number of pairs equals $\binom{e}{2} = \frac{e(e-1)}{2}$, a strictly convex function of e. So there is a strong case that the marginal cost of categories will be strictly increasing in e.

Since costs (until section 4.3) are determined by the number of categories in a criterion, we sometimes write costs as $\kappa(e)$, defined to equal $\kappa(C_i)$ for any C_i such that $e(C_i) = e$.

Increasing the number of categories e in a criterion seems to present a trade-off. While the affected criteria become more expensive to form, the creation of a given number of choice classes requires fewer criteria. Under mild assumptions, the first effect dominates the second: the cost of a larger e outweighs the advantage of using fewer criteria.

Example 2 To illustrate, consider choice functions with 9 choice classes. The minimal set of binary criteria that could lead to such a choice function must contain 4 criteria: the maximum number of choice classes that N binary criteria can generate is 2^N (see Proposition 1) and the minimum integer N such that 2^N is greater than or equal to 9 is 4, i.e., $\lceil \log_2 9 \rceil = 4$. The cost of using four binary criteria is therefore $4\kappa(2)$. Ternary criteria with 3 categories each would seem to be a better fit with 9 choice classes given that 9 is an exact multiple of 3. Generating a choice function with 9 choice classes requires 2 ternary criteria, which have a cost of $2\kappa(3)$. But assuming that the first category of a criterion is costless, the binary and ternary sets each employ the same number of costly categories, 4. So if marginal costs are increasing, the binary set must be strictly cheaper. Formally, if marginal costs are increasing then $\kappa(3) - \kappa(2) > \kappa(2) - \kappa(1) = \kappa(2)$ and hence $2\kappa(3) > 4\kappa(2)$. With constant marginal costs, the costs of the binary and ternary sets would tie but the binary set can generate an additional $7 = 2^4 - 9$ choice classes. Binary criteria thus enjoy both a cost and a number-of-choice-classes advantage over ternary criteria.

Both the binary and the ternary sets of criteria above employ markedly fewer categories than the 9 categories that a single criterion (in effect, a preference relation) would have to use to generate a choice function with 9 choice classes. Building choice distinctions from a nontrivial set of criteria, even when the set is not the most efficient possible, requires much less decision-making effort than making a separate decision for each pair of choice classes. ■

While Example 2 suggests that the advantage of binary criteria relies on the marginal costs of categories being at least weakly increasing, the optimality of binary criteria holds even when marginal costs are decreasing.

Let $\mathcal{X}$ denote a *set* of domains, with each $X \in \mathcal{X}$ associated with its own family of choice sets. We say that $\mathcal{C}$ **has a domain** in $\mathcal{X}$ if there is a $X \in \mathcal{X}$ such that each C_i in $\mathcal{C}$ is a binary relation on X.

Theorem 1 *Suppose that the set of domains $\mathcal{X}$ contains a X with m alternatives for all $m > 1$. The following two statements are then equivalent:*

- *for any $\mathcal{C}$ that is efficient and has a domain in $\mathcal{X}$, $\mathcal{C}$ contains only binary criteria,*
- *$\kappa(e) > \kappa(2) \lceil \log_2 e \rceil$ for all integers $e > 2$.*

The log cost condition is *very* weak: the marginal cost of additional categories can fall as e increases and even fall to 0 (as quickly as the derivative of any increasing power function falls to 0 but not quicker than $\frac{1}{e}$). For the reasoning behind half of Theorem 1, assume the second statement is satisfied: costs will then fall if any criterion C_j with $e > 2$ categories is replaced by $\lceil \log_2 e \rceil$ binary criteria and, since $2^{\lceil \log_2 e \rceil} \geq e$, the binary criteria will contribute at least as much to the product of the e_i as C_j did.[11]

Since our earlier arguments that the marginal cost of criterion categories should be increasing remain sound, Theorem 1 indicates that binary criteria will easily defeat finer criteria and that their optimality will withstand adjustments to our conceptual framework, if not too dramatic. For

[11] A violation of the inequality in Theorem 1, $\kappa(e) \leq \kappa(2) \lceil \log_2 e \rceil$ for some $e > 2$, does not imply that $\lceil \log_2 e \rceil$ binary criteria could be replaced by a single e-ary criterion without loss of efficiency. For instance, following Example 2, if $\kappa(3) = \kappa(2) \lceil \log_2 3 \rceil = 2\kappa(2)$ there will be an efficiency gain to using two binary criterion rather than a single ternary criterion as long as $n(c) < |X|$ when the ternary criterion is used (since $2^2 > 3$ and hence $\prod_{i=1}^{N} e_i$ will be greater with the binary pair). A natural way to ensure that $n(c) < |X|$ holds is to suppose that X is infinite, in which case the efficient sets of criteria are necessarily binary if and only if, for all integers $e > 2$,

$$\kappa(e) > \kappa(2) \log_2 e \text{ and } \kappa(e) \geq \kappa(2) \lceil \log_2 e \rceil .$$

This condition allows marginal costs to be constant from $e = 1$ to $e = 3$ and to be decreasing afterwards.

example, if a small discovery cost has to be paid whenever a new criterion is adopted it will remain optimal to use only binary criteria.

We turn to four applications and extensions of Theorem 1.

4.1 The costs of fine criteria and direct preference construction

When marginal category costs are increasing, Theorem 1 implies that the penalty for using fine criteria is sizable. If each criterion is constrained to have e categories, the minimum cost of a set of criteria that generates a choice function with n choice classes is $\lceil \log_e n \rceil \, \kappa(e)$ (since $\lceil \log_e n \rceil$ is the minimum integer r such that $e^r \geq n$). For approximation purposes, we ignore the difference between $\lceil \log_e n \rceil$ and $\log_e n$. The ratio of the minimum costs of a set of e-ary criteria and a set of binary criteria, when both generate n choice classes, is then

$$\frac{\kappa(e) \log_e n}{\kappa(2) \log_2 n} = \frac{\kappa(e)}{\kappa(2) \log_2 e}.$$

Recalling that $\frac{e(e-1)}{2}$ is a plausible cost function, suppose κ is linear or superlinear in e. Then, due to the $\log_2 e$ term in the denominator, the above ratio grows rapidly as a function of e: the penalty of using fine criteria becomes substantial as fineness increases. The costliest method of all lies at the extreme where a single criterion by itself determines all choice classes ($e = n$) which is the traditional account where agents make direct preference judgments. The penalty exacted by direct preference construction would become unsustainable as n increases: agents would have to turn to some cost-reduction strategy.

This estimate of the cost of fineness casts economic light on the empirical observation of psychologists, discussed in the introduction, that agents have only a limited ability to retain and manipulate concepts in working memory. These limitations seem to be a cognitive defect. But since these information-processing constraints force us into making choice discriminations more efficiently, there may never have been a pressing need for a capacity to handle many categories. Our limitations might even be the outcome of optimizing adaptations.

Binary criteria do not carry a special status in the above contest. Had we, for example, compared e-ary criteria with k-ary rather than binary criteria, the cost ratio of the former to the latter would equal $\frac{\kappa(e)}{\kappa(k) \log_k e}$ and we would conclude that, as e increases, k-ary criteria enjoy a rapidly increasing cost advantage.

Comparisons aside, the cost of using binary criteria, $\kappa(2) \lceil \log_2 n \rceil$, increases slowly as a function of n, as does the cost of using k-ary criteria. The problem introduced at the beginning of the paper, where the cost of preference construction increases on the order of n^2 (or on the order of

$n \log n$ for rational preferences) evaporates for criterion-based decision-making: the cost of a set of criteria of fixed coarseness k that makes n choice distinctions increases only on the order of $\log n$.

4.2 The efficiency of bits in information storage

Efficient decision-making can be understood as the efficient storage of information. The reinterpretation simplifies our set-up and clarifies its mathematics.

Suppose you want to store one out of n possible facts, which we can identify with the integers $0, ..., n-1$. Each of these n integers can be coded using k-ary digits and each digit can be stored in a device that attains k states. Since N k-ary digits can encode k^N integers, the number of devices needed to store n facts equals $\lceil \log_k n \rceil$, the least integer N such that $k^N \geq n$. Letting $\kappa(k)$ be the cost of a digit-storing device with k states, we therefore take the cost of storing one of n facts using k-ary devices to equal $\kappa(k) \lceil \log_k n \rceil$, the product of the cost one k-ary device and the number of devices required. To minimize this cost, there might at first glance seem to be an advantage to raising k and hence $\kappa(k)$ in order to lower $\lceil \log_k n \rceil$.

This cost-minimization problem is close to being a special case of the problem of selecting an efficient $\mathcal{C}$. If, in the decision-making model, each criterion is required to have the same number of categories e and we fix a target number of choice classes n then efficiency demands that we choose e to minimize $\kappa(e)N$ subject to the constraint that there is a c with n choice classes that uses N e-ary criteria. As Proposition 1 indicates, N must equal the least integer such that $e^N \geq n$, that is, $\lceil \log_e n \rceil$.

The correspondence between these two problems implies that if $(\mathcal{C}, c)$ is efficient in the decision-making model and every criterion in $\mathcal{C}$ has the same number of categories k^*, then k^* will solve the cost minimization problem for information storage if we set $n = n(c)$ (assuming that both problems share the same κ function). Theorem 1 therefore suggests that efficient information storage must use devices that attain just two states, i.e., that use binary digits (bits). Analogously to criterion-based choice, this conclusion will hold even when the marginal cost of using a larger digit diminishes to 0, a possibility that fits well with the increasing-returns character of digital technologies.

The facts that current-day computers store information in devices (transistors) that can assume two states and that binary storage devices are now extraordinarily cheap are presumably not coincidences. The question is why binary devices should enjoy a cost advantage.

Three non-notational differences distinguish this section from the decision-making model. First, to fit with information applications, we define efficiency as the minimization of the cost of storing one out of an exogenous number of facts, rather than jointly in terms of both cost and the number of facts to be stored. Second, the objects to be stored are associated with singleton facts rather than sets of items (as choice classes were). Finally, to fit the information-storage application, we consider digit-storing devices that share a common number of states k: each stores a k-ary digit. The decision-making model in contrast uses criteria where the number of categories can vary by criterion. If applied to information storage, our earlier set-up would allow facts to be stored as mixed-radix numerals (see Knuth (1997)).[12]

Although these differences mean that we need new definitions for the information-storage model, the proof of Theorem 2 is similar to the proof of Theorem 1; indeed the transferability of arguments across the two settings is one point of this section.

As in the decision model, we require that the cost of storing a k-ary digit is nonnegative, $\kappa(k) \geq 0$ for all $k \geq 1$. Fix the number of possible facts n and define (k, N) to be **feasible for** n if k is a positive integer such that $k^N \geq n$. Since the storage cost of (k, N) is $\kappa(k)N$, define k **to be the only base that stores one out of** n **facts efficiently** if, for some integer N, (k, N) is feasible for n and $\kappa(k)N < \kappa(k')N'$ for all (k', N') that are feasible for n with $k' \neq k$.

Theorem 2 *The following statements are equivalent:*

- *for all integers* $n > 1$*, the only base that stores one out of* n *facts efficiently is* 2*,*
- $\kappa(k) > \kappa(2) \lceil \log_2 k \rceil$ *for all integers* $k > 2$*.*

The storage cost κ is now easier to motivate since it represents the physical cost of building a device rather than psychological effort. But unlike categories in decision-making, the marginal cost of using an additional digit in a storage device is presumably decreasing in the number of digits k, due to economies of scale. So even though the log cost condition in Theorem 2 allows the marginal cost of additional digits to converge to 0, the condition might be violated for some k's. The cost $\kappa(2) \lceil \log_2 k \rceil$ is however likely to fall below $\kappa(k)$ for all sufficiently large k, since $\log_2 k$ increases so slowly, and it is therefore implausible that a large k would qualify as an efficient base. For small $k > 2$, however, in particular for $k = 3$, the log cost condition may well not apply. As it happens, ternary computers were built or emulated from the late 1950's to the early 1970's, and their advocates argued they were cost effective (Frieder (1972), Klimenko (1999), Parhami and McKeown (2013)).[13]

[12] There is no difficulty in providing a binary efficiency theorem comparable to Theorem 2 below for mixed-radix storage media.

[13] The case for ternary storage is sometimes based on 'radix economy', an engineering approximation of the cost of storing the integer n equal to $k \log_k n$, which is minimized at $k = e$, i.e., nearly at $k = 3$ (Engineering Research Associates (1950), Hurst (1984)). The implicit cost function in this approximation is $\kappa(k) = k$, which, even granting linearity in k, is difficult to justify: a $k = 1$ device, which does nothing, cannot store any number, and leaves the

4.3 Diverse criteria

Criteria can vary by how costly their categories are and by the value of their distinctions. With movies, for example, genre distinctions are presumably cheaper to discover than quality-of-director distinctions and, depending on the individual, have greater or lesser value. It might therefore seem that the efficient way to add choice classes would be to use the low-marginal cost categories of high-value criteria; and if the categories of the high-value criteria enjoy diminishing marginal costs then presumably those criteria should become ever finer. But in fact the coarse criteria continue to prevail: even when the marginal cost of categories is diminishing and no matter how many choice distinctions an agent wants to make, the agent should use only criteria that are coarser than some fixed ceiling.

To model diverse criteria, each criterion index will now identify a fixed attribute of the domain, e.g., genre or director-quality. A set of criteria $\mathcal{C}$ will now, for each attribute i, contain either one or no criterion that orders that attribute (rather than criteria always having indices $1, ..., N$). Let $\{C_i\}$ denote the **feasible criteria** for attribute i. To make sure that the conclusion that criteria should be coarser than a fixed ceiling is nontrivial, criteria must have the potential to be arbitrarily fine. Accordingly, we assume that there are one or more criteria in $\{C_i\}$ with e categories for any $e > 1$. To give agents the option to replace fine with coarse criteria no matter how many criteria are in use, we assume there is an attribute assigned to each positive integer i.

The value of an individual criterion $v(C_i)$ will lie in the interval $[\underline{v}, \overline{v}e(C_i)]$, where $1 < \underline{v} < \overline{v}$, when $e(C_i) > 1$ and will equal 1 when $e(C_i) = 1$. A set of criteria $\mathcal{C}$ and their values specify a **discrimination value function** V on sets of criteria defined by

$$V(\mathcal{C}) = \prod_{C_i \in \mathcal{C}} v(C_i),$$

which replaces the number of choice classes as the agent's discrimination objective.[14] The value of a criterion can increase without bound as its number of categories increases, a potential advantage for fine criteria (and a contrast to the utility-maximization model to come in section 4.3.1 which concludes that criteria will have bounded value). Coarse criteria in the present model can have negligible value since $v(C_i)$ can be near 1. Despite these biases that can favor fine criteria, the optimal decision is always to select coarser criteria.

The individual criterion values mix together the value per category of a criterion C_i and the number of categories in C_i. One way to disentangle the two effects is to assume that the value

objective function undefined would have positive cost. If we correct this feature by instead minimizing $(k-1)\log_k n$ s.t. $k \in \mathbb{Z}_+ \backslash \{1\}$ then $k = 2$ is optimal. Let me thank Itzhak Gilboa for telling me about ternary computers.

[14] Our original model is the special case where $v(C_i) = e(C_i)$ for all C_i.

per category, say w_i, is a function only of the number of categories: $v(C_i) = w_i(e(C_i))e(C_i)$ where $w_i(e(C_i))$ must lie in an interval $[\underline{w}, \overline{w}]$ such that $\frac{1}{2} < \underline{w} < \overline{w}$ and $\overline{w} > 1$. On the grounds of diminishing marginal utility, it would be natural to let $w_i(e)$ diminish in e, unlike our original model which in effect had $w_i(e) = 1$ for all e.

Both our original model and the present generalization determine the value of a set of criteria by multiplying the values of individual criteria. Recall that the advantage of criteria lies in the capacity of each criterion to distinguish within the categories of the remaining criteria. An increase in the distinctions of one criterion C_i therefore allow the *other* criteria to become more powerful since they can distinguish within more C_i-categories. This spillover of benefits can magnify when criteria have diverse values. To illustrate, consider a two-criteria world where C_1 has greater value than C_2, e.g., in the special case above, $w_1(e) > w_2(e')$ for all integers $e, e' > 1$. An expansion of e_2 would allow C_1 to distinguish more finely: each of the larger set of C_2-categories can be partitioned by C_1 into $e(C_1)$ distinct subsets. If, say, C_1 and C_2 are both initially binary an expansion of e_2 from 2 to 3 would allow C_1 to make its more valuable, binary distinction within three rather than two subsets of X. The greater value of C_1 therefore does *not* imply that an agent who decides to use a larger budget of categories should devote all of the increase to C_1; an increase in the number of C_2-categories could be more advantageous. This is more than an abstract possibility: Theorem 3 shows that using additional coarse criteria, even if they have small value, will be superior to making a highly valuable fine criterion yet more fine.

We retain our notation for the cost of criteria but now drop the assumption that the cost of a C_i is determined solely by $e(C_i)$. The set of criteria $\mathcal{C}$ is **efficient** if there does not exist a $\mathcal{C}'$ such that $V(\mathcal{C}') \geq V(\mathcal{C})$ and $\kappa[\mathcal{C}'] \leq \kappa[\mathcal{C}]$ with at least one of the inequalities strict.

To specify a well-behaved set of feasible criteria, we need to define when the costs of different sets of criteria are near to each other. Each sequence of criteria $\langle C_i^k \rangle$ for attribute i such that $e(C_i^k) = k$ for each positive k defines a cost function κ_i on $\mathbb{N}$ given by $\kappa_i(k) = \kappa(C_i^k)$. The feasible criteria for i thus define a set of feasible cost functions for i, denoted $\{\kappa_i\}$, one function for each possible $\langle C_i^k \rangle$. The **set of feasible cost functions is compact** if $\bigcup_{i=1}^{\infty} \{\kappa_i\}$ is compact in the topology of uniform convergence on $\mathbb{N}$.[15] Define κ_i to **dominate fractional power functions** if there exists a $0 < a < 1$ such that for any $\gamma > 0$ there is a $\bar{e}$ where $\kappa_i(e) > \gamma e^a$ if $e > \bar{e}$. This condition slightly weakens the log cost condition used earlier, but the marginal cost of categories can still descend 0 as e increases. For example, $\kappa_i(e) = \sqrt{e}$ satisfies the condition (set $a = \frac{1}{4}$).

Theorem 3 *If the feasible cost functions dominate fractional power functions and form a compact set then there is a ceiling b such that any efficient $\mathcal{C}$ contains only criteria with fewer than b*

[15] That is, a sequence κ_{i_n} converges if $\sup_{e \in \mathbb{N}} \left|\kappa_{i_n}(e) - \kappa_{i_{n+1}}(e)\right|$ converges to 0 as $n \to \infty$.

categories.

The thrust of Theorem 3 is that, even when aiming for a large number of choice distinctions and even when some criteria have value that grows without bound as the number of their categories increases, it is better to use more low-value coarse criteria than to let the high-value criteria become finer. Though binary criteria are not singled out in the assumptions above, the proof of Theorem 3 proceeds by showing that any criterion that is excessively fine can be efficiently replaced by binary criteria. Binary criteria do however enjoy a special status. If many criterion indices share the same cost functions and values then to achieve efficiency the criteria for *these* indices must all be binary, assuming that the log cost condition holds.

4.3.1 Utility-maximizing criteria

We show here that utility and discrimination value are compatible goals. Standard utility functions satisfy our assumptions on discrimination value with room to spare.

A criterion implicitly gives an agent the information to distinguish among alternatives by category. Without the information, the agent would not know how the categories and the labels of the alternatives were related. For example, it is only by learning a director criterion that an agent discovers how titles map to director categories. To show that discrimination value can order sets of criteria as utility maximizers would, we model explicitly the information that criteria convey.

Let X equal a product of n attributes $X = \prod_{i=1}^{n} X_i$ where n is large enough to accommodate Theorem 3 or counts only those attributes ordered by some criterion in use. An agent receives the utility $u(x) = \sum_{i=1}^{n} u_i(x_i)$ for $x \in X$. Each criterion orders only one attribute and so each C_i-category has the form $E_i \times \prod_{j \neq i} X_j$ where $E_i \subset X_i$. As before, there is a set of feasible criteria for each attribute i containing at least one criterion with e categories for each $e > 1$.

In this application, the agent will choose from choice sets A that equal a product $\prod_{i=1}^{n} A_i$ where $A_i \subset X_i$ for each i and A has finite cardinality $T > 1$ (which we could let be random). The tth element of A is written $x^t = (x_i^t)_{i=1}^{n}$.

An agent who uses a criterion C_i learns, for each $x^t \in A$, the C_i-category that contains x_i^t. The agent might know labels or indices for the alternatives in A – the movie titles – but not the x's or the utilities associated with those labels. A state s therefore specifies all of the alternatives in the choice set A the agent faces. Each x_i^t is thus a random variable as is $u_i(x_i^t)$, which we assume to be integrable. While the T draws that make up A are not independent or identically distributed, we assume that the distributions of the $u_i(x_i^t)$ are conditionally independent given the agent's information.[16]

[16] The agent's information is an event of the form $\prod_{i=1}^{n} \prod_{t=1}^{T} E_i^t$ where $E_i^t \times \prod_{j \neq i} X_j$ is the C_i-category such that

To give fine criteria an edge, we do *not* assume, as one normally would in expected utility theory, that the values that each u_i can attain are bounded. The model can thus allow a fine criterion to recognize the event that $u_i(x_i)$ surpasses any given threshold, no matter how large.

The conditional independence assumption implies that the choice of which element to select from A_i can be made separately for each attribute i. Since for the A_i decision the pertinent facts are the C_i-categories that contain the x_i^t, the agent's decision is conditioned on that information.[17] Let $u_{C_i}^t \equiv \mathbb{E}[\, u_i(x_i^t) \,|\, C_i \,]$ denote the random variable that at state s equals the conditional expected utility of x_i^t given the C_i-category that contains x_i^t when s obtains. We assume for all t that $u_{C_i}^t$ is greater than, and bounded away, from $\mathbb{E}\left[u_i(x_i^t)\right]$ for any C_i with two or more categories: there is a nontrivial expected gain to letting C_i distinguish categories. For example, if $X_i \subset \mathbb{R}$, u_i is strictly increasing, each C_i-category is an interval, and there are at least two C_i-categories that contain some x_i^t with positive probability, then $u_{C_i}^t > \mathbb{E}\left[u_i(x_i^t)\right]$.

Since the agent decides after receiving the C_i information, the optimal choice when s obtains is to pick a $x_i^t \in A_i$ such that $u_{C_i}^t(s)$ is maximal in $\{u_{C_i}^1(s), ..., u_{C_i}^T(s)\}$. Ex ante, the expected utility delivered by criterion C_i is given by

$$U_{C_i} \equiv \mathbb{E}\left[\max\left[u_{C_i}^1, ..., u_{C_i}^T\right]\right]$$

and the expected utility of using the set of criteria $\mathcal{C}$ is $U(\mathcal{C}) \equiv \sum_{C_i \in \mathcal{C}} U_{C_i}$. Call a U that can arise when our assumptions on the u_i are satisfied a **utility function for criteria**. Both a utility for criteria and a discrimination value function represent complete and transitive orderings on sets of criteria.[18]

Proposition 2 *Any utility function for criteria represents an ordering of sets of criteria that some discrimination value function also represents.*

The diverse criterion model allows $v(C_i)$ to grow without bound as $e(C_i)$ increases, while the proof of Proposition 2 shows that the values for individual criteria that stem from utility maximization are bounded: utility maximization fits the model with ease.

Example 3 To get a feel for actual numbers, suppose (1) for each choice set, only two options for each attribute are available (i.e., $|A_i| = 2$ and hence $T = 2^n$), (2) for each attribute i, draw t, and

$x_i^t \in E_i^t$, and where $E_i^t = X_i$ when $\mathcal{C}$ contains no criterion for attribute i.

[17] Due to conditional independence, it is immaterial for the A_i decision whether the agent conditions on an event of the form $\prod_{j \neq i} X_j \times \prod_{t=1}^T E_i^t$ (learning only a C_i-category) or an event of the form $\prod_{i=1}^n \prod_{t=1}^T E_i^t$ (learning a C_i-category for each $C_i \in \mathcal{C}$).

[18] Representation has its standard meaning: a discrimination value or utility W represents the binary relation $\succsim$ on sets of criteria if $W(\mathcal{C}) > W(\mathcal{C}') \Leftrightarrow \mathcal{C} \succsim \mathcal{C}'$.

state s, $u_i(x_i^t(s))$ lies in $[-\frac{1}{2}, \frac{1}{2}]$ and (3) for every C_i, t, and C_i-category, the conditional distribution of $u_i(x_i^t)$ given that x_i^t lies in the E_i specified by the category is uniform over an interval of length $\frac{1}{e(C_i)}$. Easy calculations show the following relationship between the number of C_i-categories and the expected utility that C_i delivers:

$e(C_i)$	U_{C_i}
1	0
2	.125
3	.148
4	.156
∞	.167

where the bottom row lists a perfectly discriminating criterion. Two or three categories deliver the lion's share of a criterion's potential value. ■

4.4 Arbitrary increases in coarseness

We have seen that first-best efficiency requires either that criteria are binary or, in a more general setting, have a bounded number of categories. In this section, we show that if the marginal cost of categories is increasing – a stronger assumption that the log cost condition – then *any* move from finer to coarser criteria brings a gain in efficiency.

In comparing the coarseness of sets of criteria with different numbers of criteria, only the costly categories are economically relevant. Since single-category criteria, which make no choice distinctions, should cost 0, the number of costly categories of a criterion C_i is given by $e_i^* = e_i - 1$ (where as usual $e_i = e(C_i)$). Call the vector of positive integers $(e_1, ..., e_N)$ the **discrimination vector** of $\mathcal{C}$ =$\{C_1, ..., C_N\}$. Following the analogy of first-order stochastic dominance, we consider $\mathcal{C}$ to be coarser than $\mathcal{C}'$ if the proportions of costly categories that are smaller than any given level is greater for the discrimination vector of $\mathcal{C}$ than for the discrimination vector of $\mathcal{C}'$. Given a discrimination vector $\mathbf{e} = (e_1, ..., e_N)$ with some $e_i > 1$, and given an integer $k \geq 1$, let $p_k(\mathbf{e})$ denote the proportion of $\sum_{i\in\{1,...,N\}} e_i^*$ that satisfies $e_j^* \leq k$:

$$p_k(\mathbf{e}) = \frac{\sum_{i\in\{j:e_j^*\leq k\}} e_i^*}{\sum_{i\in\{1,...,N\}} e_i^*}.$$

Formally, we define the set of criteria $\mathcal{C}$ with the discrimination vector $\mathbf{e}$ to be **coarser than** the set $\mathcal{C}'$ with the discrimination vector $\mathbf{e}'$ if, for each integer $k \geq 1$, $p_k(\mathbf{e}) \geq p_k(\mathbf{e}')$ and strict inequality obtains for some $k \geq 1$.[19]

[19] If we had defined coarseness using the distribution of the e_i rather than the e_i^*, we could take any discrimination

Greater coarseness does not by itself imply an increase in efficiency. First, coarseness measures the distribution of categories not their aggregate quantity: $\mathcal{C}$ could be coarser than $\mathcal{C}'$ but $\kappa(\mathcal{C})$ and $n(c)$ (for the c paired with $\mathcal{C}$) could be so large that $\mathcal{C}$ and $\mathcal{C}'$ are not efficiency ranked. A pure advantage of coarseness therefore can appear only when the number of costly categories in $\mathcal{C}$ and $\mathcal{C}'$ is the same. Second, the two potential advantages of coarseness need to find traction: as Example 2 illustrated, either marginal costs must be strictly increasing or there must be an opportunity to make more choice distinctions.

To deal with these points, we now return to the assumption that costs are determined by the number of categories and define marginal costs to be **increasing** if $\kappa(1) = 0$ and, for all positive integers e and e',

$$e' > e \Rightarrow \left(\kappa(e'+1) - \kappa(e') \geq \kappa(e+1) - \kappa(e)\right), \tag{MC}$$

and **strictly increasing** if they are increasing but with the latter inequality in MC strict. Let $\mathcal{C}$ and $\mathcal{C}'$ have the **same number of costly categories** if $\sum_{i=1}^{N} e_i^* = \sum_{i=1}^{N'} e_i'^*$ and **form a tight comparison** if *either* marginal costs are increasing and $\min\left[\prod_{i=1}^{N} e_i, \prod_{i=1}^{N} e_i'\right] < |X|$ *or* marginal costs are strictly increasing.

Theorem 4 ('Coarser is better') *Suppose $\mathcal{C}$ and $\mathcal{C}'$ form a tight comparison and have the same number of costly categories. If $\mathcal{C}$ is coarser than $\mathcal{C}'$ then $\mathcal{C}$ is more efficient than $\mathcal{C}'$.*

While Theorem 4 implies that an efficient set of criteria must be all-binary, the condition given in Theorem 1 for all-binary efficiency is much weaker.

Despite a notation-intensive induction step, the heart of the proof of Theorem 4 is simple. With $\mathcal{C}$ and $\mathcal{C}'$ as given in the Theorem, we can append enough single-category criteria to $\mathcal{C}'$ to equalize the number of criteria in $\mathcal{C}$ and $\mathcal{C}'$ without affecting the cost of $\mathcal{C}'$ or the number of choice classes of c''s that use $\mathcal{C}'$. Figure 1 illustrates with a finer (solid, blue) set $\mathcal{C}'$ of three criteria and a coarser (dashed, red) set $\mathcal{C}$ of six criteria, with criteria arranged so that the number of categories increases from left to right. The bottom graph adds single-category criteria to $\mathcal{C}'$ to equalize the number of criteria. Now compare the criteria in $\mathcal{C}$ and the amended version of $\mathcal{C}'$ with the greatest number of categories, then compare the criteria with the second greatest number of categories, and so on, i.e., move from right to left in the Figure. The greater coarseness of the criteria in $\mathcal{C}$ and the fact that $\mathcal{C}$ and $\mathcal{C}'$ have the same number of costly categories imply that at the first point where $\mathcal{C}$ and $\mathcal{C}'$ differ, it will be the criterion in $\mathcal{C}'$ – call it C'_{more} – that has more categories. Reduce the number of categories in this criterion by 1 and increase by 1 the number of categories in some

vector, append a large number of 1's to it, and thereby make it appear to be highly coarse even though its cost and discriminatory power would not have changed.

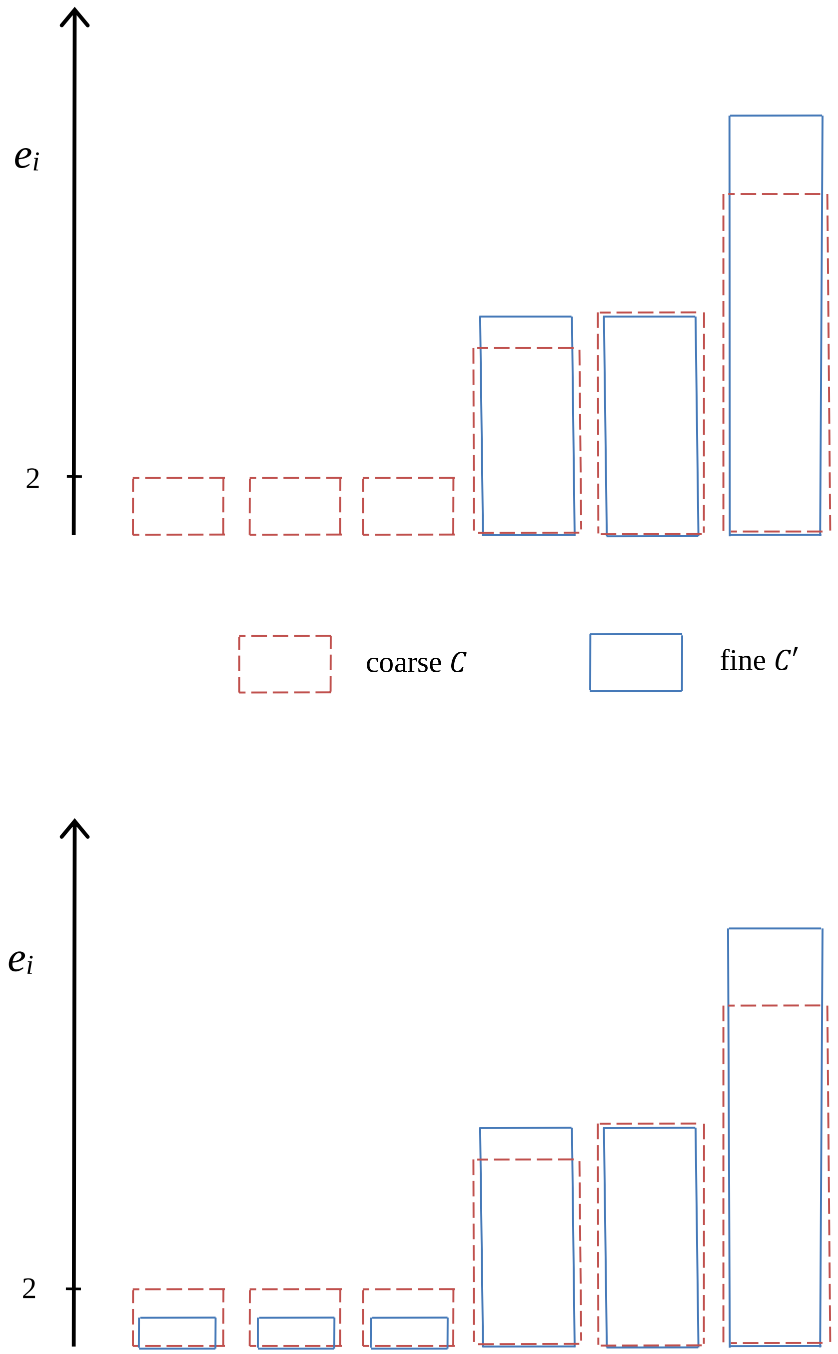

Figure 1: The move from a fine to a coarse distribution of categories

other criterion C'_k in $\mathcal{C}'$ that has at least two fewer categories than C'_{more} to create a new set $\mathcal{C}''$. In the bottom graph, this change would be a move of a category leftward from a point where the height of a solid column exceeds that of a dashed column. The change reduces costs, and strictly reduces costs if marginal costs are strictly increasing. Moreover, by a calculation similar to the one in the introduction, the product of the e'_i's will increase. Hence, the number of choice classes in some c'' that uses $\mathcal{C}''$ can weakly increase and can strictly increase if $\prod_{i=1}^{N} e'_i < |X|$. Due to the tight-comparison assumption, $\mathcal{C}''$ is strictly more efficient than $\mathcal{C}'$ (though it need not be coarser). Since the number of costly categories is the same in $\mathcal{C}$ and $\mathcal{C}'$, there is a sequence of such steps that terminate in a set of criteria with the same cost as $\mathcal{C}'$ with no accompanying drop in the number of choice classes; in the Figure, there is just enough mass where the solids exceed the dashes to fill in the locations where the dashes exceed the solids. Since each step is an efficiency increase, $\mathcal{C}$ must be more efficient than $\mathcal{C}'$.

In section 4.3, we noted that if many criterion indices share the same costs and values then efficiency requires that these criteria are all binary. The reasoning above lets us say more. If there are only a few criterion indices with shared costs and values then their numbers of categories should be smoothed under the assumptions of Theorem 4 (the numbers of categories for any two of these criteria should differ by at most one).

5 Efficiency leads to rationality

This section explains and axiomatizes the connection between binary criteria and choice functions that maximize a rational preference. For the choice functions that satisfy our axioms, rational choice is a necessary consequence of efficient decision-making. We begin with a direct and intuitive argument that shows that weighted voting, introduced in section 2, will generate a rational choice function when criteria are binary. Binariness is crucial: recall from the introduction that if three criteria rank three alternatives as voters do in the Condorcet paradox, then an equal-weight vote will cycle. Given the possibility results in voting models with dichotomous preferences (see footnote 5), one would expect that binary criteria aggregate well.

Definition 6 *Let $\mathcal{C}$ be a set of criteria and ω the criterion weights $(\omega_1, ..., \omega_N)$. Then c is a ω-**weighted-voting choice function that uses** $\mathcal{C}$ if c uses $\mathcal{C}$ and, for all $A \in \mathcal{F}$, $c(A)$ equals*

$$\left\{ x \in A : \sum_{i=1}^{N} \omega_i s_i(x, y) \geq 0 \text{ for all } y \in A \right\}$$

when this set is nonempty.[20]

A choice function c on the family of choice sets $\mathcal{F}$ is **rational** if there is a complete and transitive $\succsim$ on X such that, for all $A \in \mathcal{F}$, $c(A) = \{x \in A : x \succsim y \text{ for all } y \in A\}$.

Proposition 3 *Given any criterion weights ω, if each criterion in the set $\mathcal{C}$ is binary then the ω-weighted-voting choice c function that uses $\mathcal{C}$ is rational.*

Proof. For $\mathcal{C} = \{C_1, ..., C_N\}$, define $u_i : X \to \mathbb{R}$ by $u_i(x) = w_i$ if x is in the top category of C_i and $u_i(x) = 0$ otherwise. Since $u_i(x) - u_i(y) = w_i s_i(x, y)$ for any $x, y \in X$,

$$m(A) \equiv \left\{ x \in A : \sum_{i=1}^{N} u_i(x) \geq \sum_{i=1}^{N} u_i(y) \text{ for all } y \in A \right\} =$$

$$\left\{ x \in A : \sum_{i=1}^{N} w_i s_i(x, y) \geq 0 \text{ for all } y \in A \right\}$$

for all $A \in \mathcal{F}$. Since $\{\sum_{i=1}^{N} u_i(x) : x \in A\}$ is finite (it has at most 2^N elements), $\sum_{i=1}^{N} u_i(x)$ must reach a maximum as x varies in A and therefore $m(A)$ is nonempty. Given Definition 6, $m(A) = c(A)$ and, since the binary relation $\succsim$ on X defined by $x \succsim y$ if and only if $\sum_{i=1}^{N} u_i(x) \geq \sum_{i=1}^{N} u_i(y)$ is complete and transitive, c is rational. ■

The reach of Proposition 3 is fairly broad; for example when criteria are binary a seemingly unrelated choice procedure, the lexicographic rule of Manzini and Mariotti (2007), leads to a weighted-voting choice function. The emphasis in Mandler et al. (2012) that lexicographic compositions of binary criteria lead to rational choice functions therefore misleads a little: the key ingredient is that criteria are binary, not the lexicography.

For the general result, we continue to assume that alternatives that are winners against every element of a choice set A are selected from A. Given the choice function c, call $x \in A$ a **Condorcet winner in** A if $x \in c(\{x, y\})$ for all $y \in A$ and define c to satisfy the **Condorcet rule** if, for any $A \in \mathcal{F}$, whenever there is a Condorcet winner in A, $c(A)$ equals the set of Condorcet winners.

Fixing the set of criteria $\mathcal{C}$, let $U^{x,y} = \{C_i \in \mathcal{C} : x \, C_i \, y\}$ denote the set of criteria that rank x over y. Define the set of criteria U to be **decisive** against the set of criteria V, written $U \, D \, V$,

[20] Recall from section 2 that, for any pair x, y,

$$s_i(x, y) = \begin{cases} 1 & \text{if } x \, C_i \, y \\ -1 & \text{if } y \, C_i \, x \\ 0 & \text{otherwise} \end{cases}.$$

if, for all $a, b \in X$, $U^{a,b} = U$ and $U^{b,a} = V$ imply $a \in c(\{a, b\})$. So when U is decisive against V and a is recommended over b by the criteria in U and opposed by the criteria in V then a is chosen from $\{a, b\}$.

Definition 7 *A choice function c satisfies the* ***weighting axioms*** *with respect to the set of criteria* $\mathcal{C}$ *if, for all* $x, y \in X$, *and all subsets of criteria* U, V, U', V', *and* W *in* $\mathcal{C}$,

- $x \in c(\{x, y\}) \Rightarrow U^{x,y} \, D \, U^{y,x}$ *(decisiveness),*
- $U \, D \, V$, $U' \, D \, V'$, *and* $U \cap U' = \varnothing \Rightarrow (U \cup U') \, D \, (V \cup V')$ *(union),*
- $U \, D \, V$ *and* $W \subset (U \cap V) \Rightarrow (U \backslash W) \, D \, (V \backslash W)$ *(subtraction).*

Decisiveness says that if one set of criteria defeats another set with respect to some pair of alternatives then, whenever another pair of alternatives is backed by the same sets of criteria, the alternative backed by the previously victorious set will win: the victorious set of criteria has greater 'weight' than the defeated set. Union states that if each of two sets of disjoint criteria defeats another set of criteria, then the union of winners is decisive against the union of the losers. The disjointness is important: the union of overlapping sets of criteria might be no larger (or not much larger) than the sets of winners taken separately and it would not be reasonable to require such a union to defeat a more formidable set of criteria than each faced separately. Subtraction says that if we take away the same set of criteria from the winners and the losers then the winners remain decisive.

It is easy to confirm that a weighted-voting choice function that uses $\mathcal{C}$ satisfies the weighting axioms with respect to $\mathcal{C}$.

Theorem 5 *If a choice function c satisfies the weighting axioms with respect to a set of binary criteria and satisfies the Condorcet rule then c is rational.*

Subject to the stated provisos – the weighting axioms, the Condorcet rule, and our cost condition – Theorems 1 and 5 together show that efficiency in decision-making implies that choices will be rationally ordered.

6 Conclusion

To end with practical advice, suppose you want to use criteria to order in restaurants with the dual goals of discriminating sufficiently among meals and making the fewest decisions. Theorem 1 instructs you to use binary criteria and, to satisfy maximal discrimination, to set the binary distinction of each criterion to 'cut across' the distinctions made by the other criteria. To achieve these goals, you must be able to view the set of meals as a product of attributes, with one attribute for each criterion and with each criterion partitioning its attribute into two subsets, one better and one worse.

If an agent uses monotone attributes – in the case of meals, say, calorie count or cost – then building the needed criteria requires only that the agent choose a cutoff that divides each attribute into two parts with more and less respectively of the attribute. Some attributes that need not be monotone – meat versus vegetarian – may also happen to divide easily into two parts. The upshot of Theorem 1 is that a binary structure such as this, although it seems crude, is the most efficient way to partition meals into a given number of choice classes. If you can find the required binary distinctions, the efficiency advantage will remain even if on each evening you go out to eat your ranking of some of the binary categories changes – one day you prefer fish and the next you prefer meat.

This binary method may offer a good description of how people sometimes decide. Our analysis pushes Rubinstein (1996) one step further: not only do rational binary *relations* stand out in their usefulness but those binary relations that stem from binary *categories* end up being the cheapest way to make decisions.

A Appendix: Remaining results and proofs

Proposition 4 *For any choice function c, the choice classes of c form a partition of X.*

Proof. It is sufficient to show that the binary relation xRy defined by 'x and y are elements of the same choice class' is an equivalence relation. Reflexivity and symmetry are immediate. For transitivity assume that $xRyRz$.

Given $B \subset X$ and $a \in X$, let $B - a$ denote $B \backslash \{a\}$ and $B + a$ denote $B \cup \{a\}$.

Assume $x \in c(B)$ and $z \in B$. Suppose by way of contradiction that $z \notin c(B)$. Since yRz, $y \notin c(B)$. Since xRy, $y \notin B$. Letting $B - x$ play the role of A in Definition 2, the assumption that $x \in c(B)$ implies $y \in c(B - x + y)$ and hence, since yRz, $z \in c(B - x + y)$. But, letting $B - x$ again play the role of A, the assumption that $z \notin c(B)$ implies $z \notin c(B - x + y)$. So $x \in c(B)$ and $z \in B$ imply $z \in c(B)$.

Next assume $B \cap \{x, z\} = \varnothing$ and $x \in c(B + x)$. If $y \in B$ then, since xRy, $y \in c(B + x)$ and so, since yRz and letting $B + x$ play the role of A, $z \in c(B + x - y + z)$. Hence, since xRy and letting $B + z - y$ play the role of A, $z \in c(B + z)$. Alternatively if $y \notin B$ then, since xRy, $y \in c(B + y)$. Hence, given yRz, $z \in c(B + z)$. So $B \cap \{x, z\} = \varnothing$ and $x \in c(B + x)$ imply $z \in c(B + z)$.

Finally assume $B \cap \{x, z\} = \varnothing$, $w \in B$, and $w \in c(B + x)$. If $y \in B$ then yRz implies $w \in c(B + x - y + z)$. Hence, letting $B + x - y + z$ play the role of A and given that xRy, $w \in c(B + z)$. If $y \notin B$ then, letting $B + x$ play the role of A, xRy implies $w \in c(B + y)$ and hence, letting $B + y$ play the role of A and given that yRz, $w \in c(B + z)$. So $B \cap \{x, z\} = \varnothing$, $w \in B$, and $w \in c(B + x)$ imply $w \in c(B + z)$. ■

Definition 8 *Given a set of criteria $\mathcal{C}$, the **discrimination partition** $\mathcal{P}$ is the partition of X that, for any pair $x, y \in X$, places x and y in the same $P \in \mathcal{P}$ if and only if, for each $C_i \in \mathcal{C}$, x and y are contained in the same C_i-category.*

Proof of Proposition 1. Suppose $(\mathcal{C}, c)$ is efficient. Since c uses $\mathcal{C}$, $n(c) \leq |\mathcal{P}|$. Moreover, given $\mathcal{C}$ and hence $\mathcal{P}$, there exists a $\widehat{c}$ that uses C such that $n(\widehat{c}) = |\mathcal{P}|$. For example, assign a distinct number $r(P)$ to each $P \in \mathcal{P}$, set $R(x) = r(P)$ where $P \in \mathcal{P}$ satisfies $x \in P$, and let $\widehat{c}$ select from any choice set A only those alternatives $x \in A$ with the largest $R(x)$: $\widehat{c}(A) = \{x \in A : R(x) \geq R(y)$ for all $y \in A\}$. It is easy to confirm that $\widehat{c}$ uses $\mathcal{C}$, i.e., if x and y are elements of the same cell of $\mathcal{P}$ then x and y are elements of the same choice class. Conversely if x and y are elements of the same choice class then, since $\{x, y\} \in \mathcal{F}$, $\widehat{c}(\{x, y\}) = \{x, y\}$. Therefore $R(x) = R(y)$ and hence x and y are in the same cell of $\mathcal{P}$. Since therefore $n(\widehat{c}) = |\mathcal{P}|$, we must have $n(c) = |\mathcal{P}|$.

Let e_i indicate $e(C_i)$, $i = 1, ..., N$, for the remainder of the proof. To identify the maximum value of $|\mathcal{P}|$ for sets of N criteria $\widehat{\mathcal{C}}$ such that $e(\widehat{C}_i) = e_i$ for each i, assume first that $\prod_{i=1}^{N} e_i \leq |X|$. Set $\widehat{C}_1$ to have e_1 categories with cardinalities that, if X is finite, equal $\left\lfloor \frac{|X|}{e_1} \right\rfloor$ or $\left\lfloor \frac{|X|}{e_1} \right\rfloor + 1$ and otherwise are all infinite. Proceeding by induction, suppose that $\{\widehat{C}_1, ..., \widehat{C}_t\}$ with $e_1, ..., e_t$ categories determine, via Definition 3, a partition $\mathcal{P}_t$ of X into $\prod_{i=1}^{t} e_i$ cells such that the cardinalities of the cells of $\mathcal{P}_t$ are either all infinite or are finite and equal either $\left\lfloor \frac{|X|}{\prod_{i=1}^{t} e_i} \right\rfloor$ or $\left\lfloor \frac{|X|}{\prod_{i=1}^{t} e_i} \right\rfloor + 1$. Let m denote $\left\lfloor \frac{|X|}{\prod_{i=1}^{t} e_i} \right\rfloor$. To build $\widehat{C}_{t+1}$ in the finite case, define for each $P_t \in \mathcal{P}_t$ a partition $\mathcal{P}(P_t)$ of P_t into e_{t+1} cells that have either cardinality $\left\lfloor \frac{m}{e_{t+1}} \right\rfloor$ or $\left\lfloor \frac{m}{e_{t+1}} \right\rfloor + 1$. Since both m and $m + 1$ lie in the interval $\left[e_{t+1} \left\lfloor \frac{m}{e_{t+1}} \right\rfloor, e_{t+1} \left(\left\lfloor \frac{m}{e_{t+1}} \right\rfloor + 1 \right) \right]$, the partitioning is feasible. In the infinite case, let $\mathcal{P}(P_t)$ again consist of e_{t+1} cells, all with infinitely many alternatives. In both cases, let $1[P_t], ..., e_{t+1}[P_t]$ denote the cells of $\mathcal{P}(P_t)$. For $x \in X$, let $k_x \in \{1, ..., e_{t+1}\}$ denote the integer such that $x \in k_x[P_t]$ for some $P_t \in \mathcal{P}_t$. We set $\widehat{C}_{t+1}$ by $x \, \widehat{C}_{t+1} \, y \Leftrightarrow k_x \geq k_y$. Then $\widehat{C}_{t+1}$ has e_{t+1} categories and we set the partition $\mathcal{P}_{t+1}$ to equal $\bigcup_{P_t \in \mathcal{P}_t} \mathcal{P}(P_t)$. The cardinalities of the cells of $\mathcal{P}_{t+1}$ in the finite case equal, using the nested division rule for $\lfloor \cdot \rfloor$, either $\left\lfloor \frac{\left\lfloor \frac{|X|}{\prod_{i=1}^{t} e_i} \right\rfloor}{e_{t+1}} \right\rfloor = \left\lfloor \frac{|X|}{\prod_{i=1}^{t+1} e_i} \right\rfloor$ or $\left\lfloor \frac{|X|}{\prod_{i=1}^{t+1} e_i} \right\rfloor + 1$ and in the infinite case are all infinite. In both cases, $\mathcal{P}_{t+1}$ has $\prod_{i=1}^{t+1} e_i$ cells. Hence there is a set of criteria $\{\widehat{C}_1, ..., \widehat{C}_N\}$ with $e_1, ..., e_N$ categories that determine a partition $\mathcal{P}_N$ of X into $\prod_{i=1}^{N} e_i$

cells. Thus $|\mathcal{P}|$ can achieve a value of $\prod_{i=1}^{N} e_i$.

To see that $\prod_{i=1}^{N} e_i$ is the maximum for $|\mathcal{P}|$, let $\{C'_1, ..., C'_{N'}\}$ be a set of criteria with $e_1, ..., e_{N'}$ categories. Applying Definition 3, each $\{C'_1, ..., C'_t\}$ determines a partition $\mathcal{P}'_t$ of X and $|\mathcal{P}'_1| \leq e_1$. Suppose for $t \in \{1, ..., N-1\}$ that $|\mathcal{P}'_t| \leq \prod_{i=1}^{t} e_i$. Fix some $P_t \in \mathcal{P}_t$. Then ($P_{t+1} \in \mathcal{P}'_{t+1}$ and $P_{t+1} \subset P_t$) if and only if there is a C_{t+1}-category E_{t+1} such that $P_{t+1} = P_t \cap E_{t+1}$. Since there are at most e_{t+1} C_{t+1}-categories, P_t contains at most e_{t+1} cells of $\mathcal{P}'_{t+1}$. Hence $\left|\mathcal{P}'_{t+1}\right| \leq e_{t+1} \left(\prod_{i=1}^{t} e_i\right)$ and we conclude that $\left|\mathcal{P}'_{N'}\right| \leq \prod_{i=1}^{N'} e_i$.

When $\prod_{i=1}^{N} e_i > |X|$, terminate the recursion used to construct the $\widehat{C}_i$ at the largest t such that $\prod_{i=1}^{t} e_i \leq |X|$. To define $\widehat{C}_{t+1}$, let each partition $\mathcal{P}(P_t)$ have $|P_t|$ rather than e_{t+1} cells. The remainder of the construction of $\widehat{C}_{t+1}$ is unchanged, leading to a $\widehat{C}_{t+1}$ with $\max\{|P_t| : P_t \in \mathcal{P}_t\}$ categories, and a partition $\mathcal{P}_{t+1}$ with $|X|$ cells. The criteria $\widehat{C}_{t+2}, ..., \widehat{C}_N$ can be set arbitrarily. Of course the maximum value for $|\mathcal{P}|$ is $|X|$ in this case.

Since $\prod_{i=1}^{N} e_i$ is the maximum value for $|\mathcal{P}|$ (for $(\widehat{C}_1, ..., \widehat{C}_N)$ such that $(e(\widehat{C}_1), ..., e(\widehat{C}_N)) = (e_1, ..., e_N)$) when $\prod_{i=1}^{N} e_i \leq |X|$, and $|X|$ is the maximum value for $|\mathcal{P}|$ when $\prod_{i=1}^{N} e_i > |X|$, and since $n(c) = |\mathcal{P}|$, we conclude that $n(c) = \min\left[\prod_{i=1}^{N} e_i, |X|\right]$. ■

Proof of Theorem 1. Assume that $\kappa(e) > \kappa(2) \lceil \log_2 e \rceil$ for all $e > 2$ and suppose that, for some $X \in \mathcal{X}$, there is an efficient set $\mathcal{C}$ of N criteria defined on X that contains a C_i with $e > 2$ categories. Define a set $\mathcal{C}'$ of $N - 1 + \lceil \log_2 e \rceil$ criteria such that, for $j \in \{1, ..., N\} \backslash \{i\}$, $e(C'_j) = e(C_j)$ and where the remaining $\lceil \log_2 e \rceil$ criteria are binary. The proof of Proposition 1 shows that we may construct $(\mathcal{C}', c')$ so that the discrimination partition $\mathcal{P}'$ of $\mathcal{C}'$ satisfies $|\mathcal{P}'| = \min[\prod_{j=1}^{N-1+\lceil \log_2 e \rceil} e'_j, |X|]$ and $n(c') = |\mathcal{P}'|$. Since $\kappa[\mathcal{C}] = \sum_{j=1}^{N} \kappa(C_j)$ and $\kappa[\mathcal{C}'] = \sum_{j \in \{1,...,N\} \backslash \{i\}} \kappa(C_j) + \kappa(2) \lceil \log_2 e \rceil$,

$$\kappa[\mathcal{C}] - \kappa[\mathcal{C}'] = \kappa(e) - \kappa(2) \lceil \log_2 e \rceil > 0.$$

Let c use $\mathcal{C}$. By Proposition 1, $n(c) \leq \min[\prod_{j=1}^{N} e_j, |X|]$ whereas $n(c') = \min[\prod_{j=1}^{N-1+\lceil \log_2 e \rceil} e'_j, |X|]$. Since $2^{\lceil \log_2 e \rceil} \geq 2^{\log_2 e} = e$,

$$\left(\prod_{j=1}^{N-1+\lceil \log_2 e \rceil} e'_j \right) - \left(\prod_{j=1}^{N} e_j \right) = \left(\prod_{j \in \{1,...,N\} \backslash \{i\}}^{N} e_j \right) \left(2^{\lceil \log_2 e \rceil} - e \right) \geq 0,$$

and therefore $n(c') \geq n(c)$. Hence $(\mathcal{C}', c')$ is more efficient than $(\mathcal{C}, c)$ for any c that uses $\mathcal{C}$, a contradiction.

Conversely, assume that any efficient $\mathcal{C}$ on a domain in $\mathcal{X}$ contains only binary criteria and suppose that, for some $e > 2$, $\kappa(e) \leq \kappa(2) \lceil \log_2 e \rceil$. Set $X \in \mathcal{X}$ so that $|X| = e$. To see that there exists a minimum-cost $\widehat{\mathcal{C}}$ such that $\widehat{c}$ uses $\widehat{\mathcal{C}}$ and $n(\widehat{c}) = e$, notice that, for any $(\mathcal{C}'', c'')$ with $n(c'') = e$ and $e(C''_i) > 1$ for all i, $\mathcal{C}''$ cannot have minimum cost if $|\mathcal{C}''| > \lceil \log_2 e \rceil$: by eliminating a criterion we would have a $\mathcal{C}'''$ with $\prod_{j=1}^{N} e_j''' \geq e$ and the proof of Proposition 1 implies that there is then a $(\widetilde{\mathcal{C}}, \widetilde{c})$ with $\kappa[\widetilde{\mathcal{C}}] \leq \kappa[\mathcal{C}''']$ and $e(\widetilde{c}) = e$. We also cannot have $e(C_i'') > e$ for any i. Hence there exist only finitely many values of $\kappa[\mathcal{C}'']$ for $(\mathcal{C}'', c'')$ such that $n(c'') = e$. Hence there is an efficient $(\widehat{\mathcal{C}}, \widehat{c})$ such that $n(\widehat{c}) = e$. The set $\widehat{\mathcal{C}}$ is therefore efficient (if not there would be a $(\mathcal{C}'', c'')$ with $\kappa[\mathcal{C}''] < \kappa[\widehat{\mathcal{C}}]$ and $n(c'') = e$ and hence $(\widehat{\mathcal{C}}, \widehat{c})$ would not be efficient). By assumption, $\widehat{\mathcal{C}}$ contains only binary criteria and, given the proof of Proposition 1, $\left|\widehat{\mathcal{C}}\right| = \lceil \log_2 e \rceil$. So $\kappa[\widehat{\mathcal{C}}] = \kappa(2) \lceil \log_2 e \rceil$. But the $(\mathcal{C}', c')$, where $\mathcal{C}'$ consists of a single criterion with e categories and $n(c') = e$, satisfies $\kappa[\mathcal{C}'] = \kappa(e) \leq \kappa(2) \lceil \log_2 e \rceil$. There must therefore be an efficient $(\mathcal{C}'', c'')$ such that $n(c'') = e$ and where $\mathcal{C}''$ contains at least one nonbinary criterion. Since $\mathcal{C}''$ would also then be efficient, we have a contradiction. ■

Proof of Theorem 2. Assume that $\kappa(k) > \kappa(2) \lceil \log_2 k \rceil$ for all $k > 2$ and suppose, for the positive integer n, that (k', N') is feasible for n and $k' \neq 2$. Since $2^{\lceil \log_2 n \rceil} \geq n$, $(2, \lceil \log_2 n \rceil)$ is feasible for n.

Recursively define a sequence of vectors $v^1, ..., v^T$ beginning with the N'-vector v^1 of all k's and each stage removing one k and adding $\lceil \log_2 k \rceil$ 2's, thus arriving at a N-vector v^T of all 2's. For any $i = 1, ..., T-1$, let the number of 2's in v^i equal a and the number of k's equal b. Since $\kappa(2) \lceil \log_2 k \rceil < \kappa(k)$,

$$\kappa(2)(a + \lceil \log_2 k \rceil) + \kappa(k)(b - 1) < \kappa(2)a + \kappa(k)b.$$

Hence

$$\kappa(2) \left|\{j : v_j^{i+1} = 2\}\right| + \kappa(k) \left|\{j : v_j^{i+1} = k\}\right| < \kappa(2) \left|j : \{v_j^i = 2\}\right| + \kappa(k) \left|j : \{v_j^i = k\}\right|$$

and therefore $\kappa(2)N < \kappa(k)N'$.

Lemma. If $\kappa(k) > \kappa(2) \lceil \log_2 k \rceil$ and the $(a + b)$-vector w^1 consists of $a \geq 0$ 2's and $b > 0$ k's, then the $(a + \lceil \log_2 k \rceil + b - 1)$-vector w^2 that consists of $(a + \lceil \log_2 k \rceil)$ 2's and $b - 1$ k's satisfies $\prod_{j=1}^{a+\lceil \log_2 k \rceil + b-1} w_j^2 \geq \prod_{j=1}^{a+b} w_j^1$.

To prove the Lemma, note that since $2^{\lceil \log_2 k \rceil} \geq 2^{\log_2 k} = k$,

$$2^{a+\lceil \log_2 k \rceil} k^{b-1} - 2^a k^b = (2^a k^{b-1})(2^{\lceil \log_2 k \rceil} - k) \geq 0.$$

Beginning with v^1, apply the Lemma $T-1$ times to conclude that $2^N \geq k^{N'}$. Since by feasibility $k^{N'} \geq n$, we have $N \geq \log_2 n$ and, since N is an integer, $N \geq \lceil \log_2 n \rceil$. Since $\kappa(2)N < \kappa(k)N'$, $\kappa(2) \lceil \log_2 n \rceil < \kappa(k)N'$.

Conversely, assume for any positive integer n that 2 is the only base that stores one out of n facts efficiently and suppose, for some $k' > 2$, that $\kappa(k') \leq \kappa(2) \lceil \log_2 k' \rceil$. Set $n = k'$ and observe that $(k',1)$ is feasible for n. Since $\kappa(k')(1) \leq \kappa(2)\lceil \log_2 n \rceil$ and $(2, \lceil \log_2 n \rceil)$ is also feasible for n, 2 is not the only base that stores one out of n facts efficiently. ■

Proof of Theorem 3. We record that the topology on $\bigcup_{i\in\mathbb{N}}\{\kappa_i\}$ is a metric space with $\delta(\kappa_i, \kappa_j) = \min\left[1, \sup \bigcup_{e\in\mathbb{N}} \{|\kappa_i(e) - \kappa_j(e)|\}\right]$ serving as the metric.

We show, as a preliminary step, that there is a a such that $\inf \kappa_i(e) > \sup \kappa_j(2)\lceil e^a \overline{v} \rceil$ for all e sufficiently large, where $\inf \kappa_j(e)$ denotes $\inf\{\kappa_j(e) : \kappa_j \in \bigcup_{i\in\mathbb{N}}\{\kappa_i\}\}$, $\sup \kappa_j(e) = \sup\{\kappa_j(e) : \kappa_j \in \bigcup_{i\in\mathbb{N}}\{\kappa_i\}\}$. Since $2x \geq x+1 \geq \lceil x \rceil$ for all $x \geq 1$ and since $\overline{v} > 1$, $2e^a\overline{v} \geq \lceil e^a \overline{v} \rceil$ for $a > 0$ and $e \geq 1$. Hence $2 \sup \kappa_j(2) e^a \overline{v} \geq \sup \kappa_j(2) \lceil e^a \overline{v} \rceil$ for $a > 0$ and $e \geq 1$. For $\kappa_j \in \bigcup_{i\in\mathbb{N}}\{\kappa_i\}$ there is, by assumption, an $a > 0$ such that $\kappa_i(e) > 2\sup \kappa_j(2)\overline{v}e^a$ for all e sufficiently large. Hence $\kappa_i(e) > \sup \kappa_j(2)\lceil e^a \overline{v} \rceil$ for all e sufficiently large. To conclude this step, suppose to the contrary that for each $n \in \mathbb{N}$ there is an increasing sequence of natural numbers $\langle e_l^n \rangle$ that satisfies $\inf \kappa_i(e_l^n) \leq \sup \kappa_j(2)\left\lceil (e_l^n)^{\frac{1}{n}}\, \overline{v} \right\rceil$. Since the topology on $\bigcup_{i\in\mathbb{N}}\{\kappa_i\}$ is metrizable, the compactness assumption implies that for each n and e_l^n there is a $\kappa^{n,e_l^n} \in \bigcup_{i\in\mathbb{N}}\{\kappa_i\}$ such that $\kappa^{n,e_l^n}(e_l^n) \leq \sup \kappa_j(2)\left\lceil (e_l^n)^{\frac{1}{n}}\, \overline{v} \right\rceil$. Hence for each $n \in \mathbb{N}$ there is a $\widehat{e}^{\,n} \in \mathbb{N}$ and $\kappa^n \in \bigcup_{i\in\mathbb{N}}\{\kappa_i\}$ such that (1) $(\widehat{e}^{\,n})^{\frac{1}{n}}\, \overline{v} \to \infty$ (and therefore $\widehat{e}^{\,n} \to \infty$) and (2) $\kappa^n(\widehat{e}^{\,n}) \leq \sup \kappa_j(2)\left\lceil (\widehat{e}^{\,n})^{\frac{1}{n}}\, \overline{v} \right\rceil$. Due to compactness, there is a subsequence of natural numbers $\langle n_k \rangle$ and a $\overline{\kappa} \in \bigcup_{i\in\mathbb{N}}\{\kappa_i\}$ such that $\kappa^{n_k} \to \overline{\kappa}$. Since, for some $\overline{a} > 0$, $\overline{\kappa}(e) > 2\sup \kappa_j(2)\lceil e^{\overline{a}}\overline{v} \rceil$ for all e sufficiently large and since $\frac{1}{n_k} < \overline{a}$ for all n_k sufficiently large, there exist $\overline{e} > 0$ and $\overline{n} > 0$ such that $\overline{\kappa}(e) > 2\sup \kappa_j(2)\left\lceil e^{\frac{1}{n_k}}\overline{v} \right\rceil$ for all $e > \overline{e}$ and all $n_k > \overline{n}$. But due to (2), for each n_k,

$$\overline{\kappa}(\widehat{e}^{\,n_k}) \leq \sup \kappa_j(2)\left\lceil (\widehat{e}^{\,n_k})^{\frac{1}{n_k}}\, \overline{v} \right\rceil + \left(\overline{\kappa}(\widehat{e}^{\,n_k}) - \kappa^n(\widehat{e}^{\,n_k})\right),$$

and hence, since the metric δ implies $\kappa^n(\widehat{e}^{\,n_k}) - \overline{\kappa}(\widehat{e}^{\,n_k}) \to 0$ and since $(\widehat{e}^{\,n_k})^{\frac{1}{n_k}}\, \overline{v} \to \infty$, we conclude that

$$\overline{\kappa}(\widehat{e}^{\,n_k}) \leq \sup \kappa_j(2)\left\lceil (\widehat{e}^{\,n_k})^{\frac{1}{n_k}}\, \overline{v} \right\rceil + (\overline{\kappa}(\widehat{e}^{\,n_k}) - \kappa^n(\widehat{e}^{\,n_k})) \leq 2\sup \kappa_j(2)\left\lceil (\widehat{e}^{\,n_k})^{\frac{1}{n_k}}\, \overline{v} \right\rceil$$

for all n_k sufficiently large, a contradiction. Hence there exist $\overline{a} > 0$ and an integer $\overline{b} > 0$ such that $\inf \kappa_i(e) > \sup \kappa_j(2)\lceil e^{\overline{a}}\overline{v} \rceil$ for all $e > \overline{b}$.

To prove the Theorem, suppose the set of N' criteria $\mathcal{C}'$ contains a C'_k with e categories. Define $\overline{\mathcal{C}}$ to coincide with $\mathcal{C}'$ except that C'_k is replaced by $\lceil e^{\overline{a}}\overline{v} \rceil$ binary criteria with indices $N'+1, ..., N' + \lceil e^{\overline{a}}\overline{v} \rceil$. Due to the previous paragraph, $\kappa[\overline{\mathcal{C}}] < \kappa[\mathcal{C}']$ if $e > \overline{b}$. As for value,

$V(\mathcal{C}') = \prod_{j\in\{1,...,N'\}} v(C'_j)$ and $V(\overline{\mathcal{C}}) \geq \underline{v}^{\lceil e^{\overline{a}}\overline{v}\rceil} \left(\prod_{j=\{1,...,N'\}\backslash\{k\}} v(C'_j)\right)$. Hence

$$V(\overline{\mathcal{C}}) - V(\mathcal{C}') \geq \underline{v}^{\lceil e^{\overline{a}}\overline{v}\rceil} \left(\prod_{j=\{1,...,N'\}\backslash\{k\}} v(C'_j)\right) - \prod_{j\in\{1,...,N'\}} v(C'_j)$$
$$= \left(\prod_{j=\{1,...,N'\}\backslash\{k\}} v(C'_j)\right)\left(\underline{v}^{\lceil e^{\overline{a}}\overline{v}\rceil} - v(C'_k)\right).$$

Since $\underline{v}^{\lceil e^{\overline{a}}\overline{v}\rceil} \geq \underline{v}^{e^{\overline{a}}\overline{v}}$ and $\overline{v}e \geq v(C'_k)$, if $\underline{v}^{e^{\overline{a}}\overline{v}} > \overline{v}e$ for all e sufficiently large then there is an integer b' such that $V(\overline{\mathcal{C}}) - V(\mathcal{C}') > 0$ when $e > b'$. To conclude that $\underline{v}^{e^{\overline{a}}\overline{v}} > \overline{v}e$ for all e sufficiently large, it is sufficient that any of the following equivalent conditions

$$\frac{\underline{v}^{e^{\overline{a}}\overline{v}}}{\overline{v}e} \to \infty \iff \ln \underline{v}^{e^{\overline{a}}\overline{v}} - \ln \overline{v}e \to \infty \iff e^{\overline{a}}\overline{v}\ln \underline{v} - \ln \overline{v}e \to \infty$$

obtains. The last condition follows from $\frac{\ln e}{e^{\overline{a}}} \to 0$ and the implications

$$\frac{\ln e}{e^{\overline{a}}} \to 0 \Rightarrow \frac{\frac{1}{\overline{v}\ln \underline{v}}\ln e}{e^{\overline{a}}} \to 0 \Rightarrow \frac{\frac{\ln \overline{v}}{\overline{v}\ln \underline{v}} + \frac{1}{\overline{v}\ln \underline{v}}\ln e}{e^{\overline{a}}} \to 0 \Leftrightarrow \frac{\ln \overline{v} + \ln e}{e^{\overline{a}}\overline{v}\ln \underline{v}} \to 0.$$

So set b in the Theorem equal to $\max\{\overline{b}, b'\}$. ■

Proof of Proposition 2. Given that U is additively separable, we may normalize each u_i without changing the ordering represented. Set $\mathbb{E}\left[u_i(x_i^t)\right] = 0$ for $i = 1, ..., n$. Define $V(\mathcal{C}) = \exp(U(\mathcal{C}))$ and $v(C_i) = \exp(U_{C_i})$ for $C_i \in \mathcal{C}$. If $e(C_i) > 1$ then $u^t_{C_i} > 0$ and hence $U_{C_i} > 0$ and $v(C_i) > 1$. Setting $\underline{v} = \inf\{v(C_i) : C_i \text{ is feasible and } i \in \{1, ..., n\}\}$, we therefore have $\underline{v} \geq 1$. Given that $u^t_{C_i}$ is bounded away from 0 for feasible C_i with $e(C_i) > 1$, $v(C_i)$ is bounded away from 1 for the same C_i and hence $\underline{v} > 1$. Since

$$U_{C_i} \leq \mathbb{E}\left[\sum_{t=1}^T \left|u^t_{C_i}\right|\right] \leq \sum_{t=1}^T \mathbb{E}\left[\mathbb{E}[\left|u_i(x_i^t)\right| \,|\, C_i]\right] = \sum_{t=1}^T \mathbb{E}\left[\left|u_i(x_i^t)\right|\right],$$

and each $u_i(x_i^t)$ is integrable, for each i there is a well-defined constant $k_i = \sum_{t=1}^T \mathbb{E}\left[\left|u_i(x_i^t)\right|\right]$ such that $U_{C_i} \leq k_i$ for all feasible C_i. Hence $v(C_i) \leq \exp k_i$ for all feasible C_i. With the $v(C_i)$ and $\underline{v}$ already defined, by setting $\overline{v} = \max[\exp k_1, ..., \exp k_n, 2\underline{v}]$ we conclude that V qualifies as a discrimination value function. ■

Terminology for Lemmas 1 - 4 and Proof of Theorem 4. Given a cost function κ and a N-vector of positive integers $\mathbf{e}$, define $\kappa[\mathbf{e}]$ to equal $\sum_{i=1}^N \kappa(e_i)$. If $\mathbf{e}$ and $\mathbf{e}'$ are, respectively, N- and N'-vectors of positive integers, $\mathbf{e}$ is **weakly more efficient** than $\mathbf{e}'$ if $\prod_{i=1}^N e_i \geq \prod_{i=1}^{N'} e'_i$ and $\kappa[\mathbf{e}] \leq \kappa[\mathbf{e}']$ for all cost functions κ with increasing marginal costs, and is **more efficient** than $\mathbf{e}'$ if (i) $\kappa[\mathbf{e}] < \kappa[\mathbf{e}']$ for all κ with strictly increasing marginal costs, and (ii) $\prod_{i=1}^N e_i > \prod_{i=1}^{N'} e'_i$.

The vector $\mathbf{e}$ is **coarser than** $\mathbf{e}'$ if, for some domain X, there exist $\mathcal{C}$ and $\mathcal{C}'$ such that $\mathbf{e}$ is the discrimination vector of $\mathcal{C}$, $\mathbf{e}'$ is the discrimination vector of $\mathcal{C}'$, and $\mathcal{C}$ is coarser than $\mathcal{C}'$. We will follow the convention that, for any N-vector of integers $\mathbf{e}$, coordinate labels are chosen so that e_i increases in i: $e_i \geq e_{i-1}$ for $i = 2, ..., N$.

Lemma 1 Let $\mathbf{e}$ (resp. $\mathbf{e}^+$) be a vector of positive integers with N (resp. N^+) coordinates. If $\mathbf{e}$ is coarser than $\mathbf{e}^+$ and $\sum_{i=1}^{N^+} e_i^{+*} = \sum_{i=1}^{N} e_i^*$ then, for all integers $1 \leq x \leq \min[N, N^+]$, $\sum_{i=N^+-x+1}^{N^+} e_i^+ \geq \sum_{i=N-x+1}^{N} e_i$ and, for some integer $1 \leq x \leq \min[N, N^+]$, $\sum_{i=N^+-x+1}^{N^+} e_i^+ > \sum_{i=N-x+1}^{N} e_i$.

Proof. Suppose to the contrary that there is a smallest integer $1 \leq x \leq \min[N, N^+]$ such that $\sum_{i=N^+-x+1}^{N^+} e_i^+ < \sum_{i=N-x+1}^{N} e_i$. Since x is smallest, $e_{N-x+1} > e_{N^+-x+1}^+$. Since $\sum_{i=N^+-x+1}^{N^+} e_i^+ = \sum_{i=N-x+1}^{N} e_i$ and both are sums of x numbers, $\sum_{i=N^+-x+1}^{N^+} e_i^{+*} < \sum_{i=N-x+1}^{N} e_i^*$. Since $\sum_{i=1}^{N^+} e_i^{+*} \geq \sum_{i=1}^{N} e_i^*$, $\frac{\sum_{i=N^+-x+1}^{N^+} e_i^{+*}}{\sum_{i=1}^{N^+} e_i^{+*}} < \frac{\sum_{i=N-x+1}^{N} e_i^*}{\sum_{i=1}^{N^+} e_i^*}$. Hence $\frac{\sum_{i=1}^{N-x} e_i^*}{\sum_{i=1}^{N} e_i^*} < \frac{\sum_{i=1}^{N^+-x} e_i^{+*}}{\sum_{i=1}^{N^+} e_i^{+*}}$. Since $e_{N-x+1}^* > e_{N^+-x+1}^{+*}$,

$$\frac{\sum_{\{i:e_i^* \leq e_{N^+-x+1}^{+*}\}} e_i^*}{\sum_{i=1}^{N} e_i^*} \leq \frac{\sum_{i=1}^{N-x} e_i^*}{\sum_{i=1}^{N} e_i^*} < \frac{\sum_{i=1}^{N^+-x} e_i^{+*}}{\sum_{i=1}^{N^+} e_i^{+*}} < \frac{\sum_{\{i:e_i^{+*} \leq e_{N^+-x+1}^{+*}\}} e_i^{+*}}{\sum_{i=1}^{N^+} e_i^{+*}},$$

contradicting the fact that $\mathbf{e}$ is coarser than $\mathbf{e}^+$.

For the final claim note that if, for all integers $1 \leq x \leq \min[N, N^+]$, $\sum_{i=N^+-x+1}^{N^+} e_i^+ \geq \sum_{i=N-x+1}^{N} e_i$ then, since $\sum_{i=1}^{N^+} e_i^{+*} = \sum_{i=1}^{N} e_i^*$, we would have $\mathbf{e}^+ = \mathbf{e}$ which implies that $\mathbf{e}$ could not be coarser than $\mathbf{e}^+$. □

Lemma 2 Given the vector of positive integers $\overline{\mathbf{e}} = (\overline{e}_1, ..., \overline{e}_{\overline{N}})$, let $\widetilde{\mathbf{e}} = (\widetilde{e}_1, ..., \widetilde{e}_{\overline{N}})$ be defined by $\widetilde{e}_i = \overline{e}_i - 1$, $\widetilde{e}_j = \overline{e}_j + 1$, and $\widetilde{e}_k = \overline{e}_k$ for $k \neq i, j$. If $\overline{e}_i \geq \overline{e}_j + 2$ then $\prod_{k=1}^{\overline{N}} \widetilde{e}_k > \prod_{k=1}^{\overline{N}} \overline{e}_k$.

Proof. Since $e_i \geq e_j + 2$ implies $e_i - e_j - 1 > 0$ (and with the convention $\prod_{l \in \mathcal{I}} e_l = 1$ when $\mathcal{I} = \varnothing$),

$$\prod_{l=1}^{\overline{N}} \widetilde{e}_l = \left(\prod_{l \neq i,j} \overline{e}_l \right) (\overline{e}_i - 1)(\overline{e}_j + 1) = \left(\prod_{l=1}^{\overline{N}} \overline{e}_l \right) + \left(\prod_{l \neq i,j} \overline{e}_l \right) (\overline{e}_i - \overline{e}_j - 1) > \prod_{l=1}^{\overline{N}} \overline{e}_l. \quad \square$$

Lemma 3 Let the vector of positive integers $\mathbf{e}$ (resp. $\mathbf{e}'$) have N (resp. N') coordinates. If $\sum_{i=N'-x+1}^{N'} e_i' \geq \sum_{i=N-x+1}^{N} e_i$ for all integers $x \in \{1, ..., \min[N, N']\}$ and $\sum_{i=1}^{N'} e_i'^* = \sum_{i=1}^{N} e_i^*$, then there exists a vector of positive integers $\widehat{\mathbf{e}}$ with N' coordinates such that $\sum_{i=1}^{N'} \widehat{e}_i^* = \sum_{i=1}^{N'} e_i'^*$, $\widehat{e}_{N'-i+1} \geq e_{N-i+1}$ for $i \in \{1, ..., \min[N, N']\}$, and $\widehat{\mathbf{e}}$ is weakly more efficient than $\mathbf{e}'$.

Proof. To proceed by induction, set $\mathbf{e}^1 = \mathbf{e}'$. Given some $k \in \{1, ..., \min[N, N'] - 1\}$, suppose (1) $\sum_{i=1}^{N'} e_i^{k*} = \sum_{i=1}^{N} e_i^*$, (2) $e_{N'-i+1}^k \geq e_{N-i+1}$ for $i = 1, ..., k$, (3) $\sum_{i=N'-x+1}^{N'} e_i^k \geq \sum_{i=N-x+1}^{N} e_i$ for all $x \in \{1, ..., \min[N, N']\}$, and (4) $\mathbf{e}^k$ is weakly more efficient than $\mathbf{e}'$. These properties obtain for $k = 1$.

If $e^k_{N'-k} \geq e_{N-k}$ then set $\mathbf{e}^{k+1} = \mathbf{e}^k$. If $e^k_{N'-k} < e_{N-k}$, let m denote the smallest positive integer such that (i) $\sum_{i=N'-k}^{N'-k+m-1} e_i^k < \sum_{i=N-k}^{N-k+m-1} e_i$ and (ii) $\sum_{i=N'-k}^{N'-k+m} e_i^k \geq \sum_{i=N-k}^{N-k+m} e_i$. There must be such a m since we can set $m = k$ and $x = k+1$. Then set (A) $e^{k+1}_{N'-i} = e_{N-i}$ for $i = k-m+1, ..., k$, (B) $e^{k+1}_{N'-k+m} = \sum_{i=N'-k}^{N'-k+m} e_i^k - \sum_{i=N-k}^{N-k+m-1} e_i$ (or equivalently $e^{k+1}_{N'-k+m} = e_{N-k+m} + \sum_{i=N'-k}^{N'-k+m} e_i^k - \sum_{i=N-k}^{N-k+m} e_i$), and (C) $e_i^{k+1} = e_i^k$ for $i = 1, ..., N'-k-1$ and $i = N'-k+m+1, ..., N'$. In both cases, $\mathbf{e}^{k+1}$ is a N'-vector.

To conclude the induction, we show that properties (1) - (4) are satisfied for $k+1$. When $e^k_{N'-k} \geq e_{N-k}$ and therefore $\mathbf{e}^{k+1} = \mathbf{e}^k$, this is immediate. So assume $e^k_{N'-k} < e_{N-k}$. Property 1. Summing (A) and (B) yields $\sum_{i=N'-k}^{N'-k+m} e_i^{k+1} = \sum_{i=N-k}^{N-k+m} e_i + \sum_{i=N'-k}^{N'-k+m} e_i^k - \sum_{i=N-k}^{N-k+m} e_i = \sum_{i=N'-k}^{N'-k+m} e_i^k$. Given (C), $\sum_{i=1}^{N'-k-1} e_i^{k+1} = \sum_{i=1}^{N'-k-1} e_i^k$ and $\sum_{i=N'-k+m+1}^{N'} e_i^{k+1} = \sum_{i=N'-k+m+1}^{N'} e_i^k$. Therefore we have

$$\sum_{i=1}^{N'} e_i^{k+1} = \sum_{i=1}^{N'} e_i^k \tag{I}$$

and hence, due to (1), $\sum_{i=1}^{N'} e_i^{(k+1)*} = \sum_{i=1}^{N} e_i^*$. Property 2. We have $e^{k+1}_{N'-i+1} \geq e_{N-i+1}$ for $i = k-m+2, ..., k+1$ by (A), for $i = k-m+1$ by (ii) and (B), and for $i = 1, ..., k-m$ by (2) and (C). Property 3. Due to (C) and (3), we have

$$\sum_{i=N'-x+1}^{N'} e_i^{k+1} \geq \sum_{i=N-x+1}^{N} e_i \tag{II}$$

for $x = 1, ..., k-m$. Due to (B) and (ii), $e^{k+1}_{N'-k+m} \geq e_{N-k+m}$. So, given that II holds for $x = 1, ..., k-m$, II holds for $x = k-m+1$. Similarly, due to (A) and given that II holds for $x = 1, ..., k-m+1$, II holds for $x = k-m+2, ..., k+1$. Finally, due to (C) and I, $\sum_{i=j}^{N'} e_i^{k+1} = \sum_{i=j}^{N'} e_i^k$ holds for $j = 1, ..., N'-k-1$. Hence $\sum_{i=N'-x+1}^{N'} e_i^{k+1} = \sum_{i=N'-x+1}^{N'} e_i^k$ for $x > k+1$ and therefore (3) implies that II holds for $x > k+1$.

Property 4. We build recursively a sequence $\langle \mathbf{e}(j) \rangle$ of $(m+1)$-vectors, each with the coordinate labels $N'-k, ..., N'-k+m$, and beginning with $\mathbf{e}(0) = (e^k_{N'-k}, ..., e^k_{N'-k+m})$. If $e_{N'-k}(j) < e_{N'-k}$ and there exists a coordinate $l \in \{N'-k+1, ..., N'-k+m\}$ with $e_l(j) > e_l^{k+1}$ then set $e_{N'-k}(j+1) = e_{N'-k}(j) + 1$, $e_l(j+1) = e_l(j) - 1$, and $e_r(j+1) = e_r(j)$ for all other coordinates r. Otherwise the sequence terminates with $\mathbf{e}(j)$. By substituting (A) and (B) into the identity

$$e_{N'-k} - e^k_{N'-k} = \sum_{i=N'-k+1}^{N'-k+m-1} e_i^k - \sum_{i=N'-k+1}^{N'-k+m-1} e_i + e^k_{N'-k+m} - \left(e_{N'-k+m} + \sum_{i=N'-k}^{N'-k+m} e_i^k - \sum_{i=N'-k}^{N'-k+m} e_i \right),$$

we get

$$e^{k+1}_{N'-k} - e^k_{N'-k} = \left(\sum_{i=N'-k+1}^{N'-k+m-1} e_i^k - \sum_{i=N'-k+1}^{N'-k+m-1} e_i^{k+1} \right) + e^k_{N'-k+m} - e^{k+1}_{N'-k+m}. \tag{III}$$

Given our assumption that $e_{N'-k} > e^k_{N'-k}$ and (A), $e^{k+1}_{N'-k} > e^k_{N'-k}$. Due to (2) and (A), $e^{k+1}_i \leq e^k_i$ for $i = N' - k + 1, ..., N' - k + m - 1$. Combining (A) and (B) gives $\sum_{i=N'-k}^{N'-k+m} e^k_i = \sum_{i=N'-k}^{N'-k+m} e^{k+1}_i$ while combining (A) and (i) gives $\sum_{i=N'-k}^{N'-k+m-1} e^k_i < \sum_{i=N'-k}^{N'-k+m-1} e^{k+1}_i$. Hence $e^{k+1}_{N'-k+m} < e^k_{N'-k+m}$. Condition III therefore implies that, for some positive integer t, $\mathbf{e}(t) = (e^{k+1}_{N'-k}, ..., e^{k+1}_{N'-k+m})$ (at which point $\langle \mathbf{e}(j) \rangle$ terminates).

For $j = 0, ..., t - 1$, $e_{N'-k}(j) < e_{N'-k}$ and, using (2), $e_l(j) > e^{k+1}_l \geq e_l$. Since $e_{N'-k} \leq e_l$, we have $e_l(j) \geq e_{N'-k} + 2$. By weakly increasing marginal costs, $\kappa[\mathbf{e}(j+1)] \leq \kappa[\mathbf{e}(j)]$ for $j = 0, ..., t-1$ and therefore $\kappa[\mathbf{e}^{k+1}] \leq \kappa[\mathbf{e}^k]$.

Applying Lemma 2,

$$\left(\prod_{l=1}^{N'-k-1} e^k_l\right)\left(\prod_{l=N'-k+m+1}^{N'} e^k_l\right)\left(\prod_{l=N'-k}^{N'-k+m} e_l(j+1)\right) > \left(\prod_{l=1}^{N'-k-1} e^k_l\right)\left(\prod_{l=N'-k+m+1}^{N'} e^k_l\right)\left(\prod_{l=N'-k}^{N'-k+m} e_l(j)\right)$$

for $j = 0, ..., t - 1$. Hence $\prod_{l=1}^{N'} e^{k+1}_l > \prod_{l=1}^{N'} e^k_l$. Therefore $\mathbf{e}^{k+1}$ is weakly more efficient than $\mathbf{e}^k$ and hence weakly more efficient than $\mathbf{e}'$, concluding the argument for Property 4.

With the induction complete, we conclude by setting $\widehat{\mathbf{e}} = \mathbf{e}^{\min[N,N']}$. $\square$

Lemma 4 If for the N-vector $\mathbf{e}$ and the $\widehat{N}$-vector $\widehat{\mathbf{e}}$ (i) each e_i and $\widehat{e}_i$ is an integer greater than 1, (ii) $N > \widehat{N}$, (iii) $\widehat{e}_{\widehat{N}-i+1} \geq e_{N-i+1}$ for $i = 1, ..., \widehat{N}$, and (iv) $\sum_{i=1}^{\widehat{N}} \widehat{e}^*_i = \sum_{i=1}^{N} e^*_i$, then $\mathbf{e}$ is more efficient than $\widehat{\mathbf{e}}$.

Proof. Define $\widetilde{\mathbf{e}} = (1, ..., 1, \widehat{e}_1, ..., \widehat{e}_{\widehat{N}})$, where the number of 1's equals $N - \widehat{N}$. Note that $\sum_{i=1}^{N} \widetilde{e}^*_i = \sum_{i=1}^{\widehat{N}} \widehat{e}^*_i = \sum_{i=1}^{N} e^*_i$ and $\widetilde{e}_{N-i+1} \geq e_{N-i+1}$ for $i = 1, ..., \widehat{N}$. We can therefore build a sequence of N-vectors $\langle \mathbf{e}(j) \rangle$ with $\mathbf{e}(1) = \widetilde{\mathbf{e}}$ and terminal vector $\mathbf{e}(t) = \mathbf{e}$ such that, for $j = 1, ..., t-1$, $e_k(j + 1) = e_k(j) + 1 \leq e_k$ for some $k = 1, ..., N - \widehat{N}$, $e_{k'}(j + 1) = e_{k'}(j) - 1 \geq e_{k'}$ for some $k' = N - \widehat{N} + 1, ..., N$, and $e_l(j + 1) = e_l(j)$ for $l \in \{1, ..., N\} \backslash \{k, k'\}$. Suppose κ has strictly increasing marginal costs. Then, since

$$e_k(j) < e_k(j + 1) \leq e_k \leq e_{k'} \leq e_{k'}(j + 1) < e_{k'}(j),$$

$e_{k'}(j) > e_k(j) + 1$ and therefore $\kappa[\mathbf{e}(j + 1)] < \kappa[\mathbf{e}(j)]$. Due in addition to Lemma 2, $\mathbf{e}(j + 1)$ is more efficient than $\mathbf{e}(j)$. Due to (ii), $t \geq 2$. Since the final $\widehat{N}$ coordinates of $\widetilde{\mathbf{e}}$ equal the vector $\widehat{\mathbf{e}}$ and the remaining coordinates equal 1, $\mathbf{e}(1)$ is weakly more efficient than $\widehat{\mathbf{e}}$. Hence $\mathbf{e}$ is more efficient than $\widehat{\mathbf{e}}$. $\square$

Proof of Theorem 4. Without loss of generality, we may assume that the discrimination vector $\mathbf{e}$ of $\mathcal{C}$ and $\mathbf{e}'$ of $\mathcal{C}'$ contain only integers greater than 1. Due to Lemma 1, $\sum_{i=N'-x+1}^{N'} e'_i \geq$

$\sum_{i=N-x+1}^{N} e_i$ for all $x \in \{1,...,\min[N,N']\}$ and therefore, by Lemma 3, there exists a vector of positive integers $\widehat{\mathbf{e}}$ with N' coordinates such that $\sum_{i=1}^{N'}(\widehat{e}_i - 1) = \sum_{i=1}^{N'}(e'_i - 1)$, $\widehat{e}_{N'-i+1} \geq e_{N-i+1}$ for $i = 1,...,\min[N,N']$, and $\widehat{\mathbf{e}}$ is weakly more efficient than $\mathbf{e}$.

Suppose that $N' > N$. Then, since $\widehat{e}_{N'-i+1} \geq e_{N-i+1}$ for $i = 1,...,\min[N,N']$ and since each $e'_i > 1$, $\sum_{i=1}^{N'}(\widehat{e}_i - 1) > \sum_{i=1}^{N}(e_i - 1)$. Since $\sum_{i=1}^{N'}(\widehat{e}_i - 1) = \sum_{i=1}^{N'}(e'_i - 1)$, $\sum_{i=1}^{N'}(e'_i - 1) > \sum_{i=1}^{N}(e_i - 1)$, which contradicts $\sum_{i=1}^{N'}(e'_i - 1) = \sum_{i=1}^{N}(e_i - 1)$.

Suppose that $N' = N$. Since $\mathbf{e}$ is coarser than $\mathbf{e}'$, $\mathbf{e} \neq \mathbf{e}'$. Since $\sum_{i=N'-x+1}^{N'} e'_i = \sum_{i=N-x+1}^{N} e_i$ for all $x \in \{1,...,N\}$ implies $\mathbf{e} = \mathbf{e}'$, Lemma 1 implies there is a $\widehat{x} \in \{1,...,N\}$ such that $\sum_{i=N'-\widehat{x}+1}^{N'} e'_i > \sum_{i=N-\widehat{x}+1}^{N} e_i$.

Next we show that for all $y \in \{1,...,N\}$, $\sum_{i=1}^{y} e'_i \geq \sum_{i=1}^{y} e_i$. If to the contrary there is a minimal $y \in \{1,...,N\}$ such that $\sum_{i=1}^{y} e'_i < \sum_{i=1}^{y} e_i$ then $\sum_{i=1}^{y-1} e'_i \geq \sum_{i=1}^{y-1} e_i$ and $e'_y < e_y$. Hence

$$\frac{\sum_{\{i: e^*_i \leq e'^*_y\}} e^*_i}{\sum_{i=1}^{N} e^*_i} \leq \frac{\sum_{i=1}^{y-1} e^*_i}{\sum_{i=1}^{N} e^*_i} \leq \frac{\sum_{i=1}^{y-1} e'^*_i}{\sum_{i=1}^{N'} e'^*_i} < \frac{\sum_{\{i: e'^*_i \leq e'^*_y\}} e'^*_i}{\sum_{i=1}^{N'} e'^*_i},$$

contradicting $\mathbf{e}$ being coarser than $\mathbf{e}'$.

Using this fact, we conclude that $\sum_{i=1}^{\widehat{x}} e'_i \geq \sum_{i=1}^{\widehat{x}} e_i$, which when combined with $\sum_{i=N'-\widehat{x}+1}^{N'} e'_i > \sum_{i=N-\widehat{x}+1}^{N} e_i$, yields $\sum_{i=1}^{N'} e'_i > \sum_{i=1}^{N} e_i$. But since $N = N'$ implies $\sum_{i=1}^{N'} e'_i = \sum_{i=1}^{N} e_i$, we have a contradiction.

We conclude that $N > N'$. Apply Lemma 4 to conclude that $\mathbf{e}$ is more efficient than $\widehat{\mathbf{e}}$ and hence more efficient than $\mathbf{e}'$.

Let c use $\mathcal{C}$ and maximally discriminate (the proof of Proposition 1 shows such a c exists) and let c' use $\mathcal{C}'$. Proposition 1 implies $n(c) = \min\left[\prod_{i=1}^{N} e_i, |X|\right]$ and $n(c') \leq \min\left[\prod_{i=1}^{N'} e'_i, |X|\right]$. Given that $\mathbf{e}$ is more efficient than $\mathbf{e}'$, $\prod_{i=1}^{N} e_i > \prod_{i=1}^{N'} e'_i$ and therefore $\min\left[\prod_{i=1}^{N} e_i, |X|\right] \geq \min\left[\prod_{i=1}^{N'} e'_i, |X|\right]$. Hence $n(c) \geq n(c')$.

Since $\mathcal{C}$ and $\mathcal{C}'$ form a tight comparison, either $\min\left[\prod_{i=1}^{N} e_i, \prod_{i=1}^{N} e'_i\right] < |X|$ or marginal costs are strictly increasing. In the first case, the fact that $\prod_{i=1}^{N'} e'_i < \prod_{i=1}^{N} e_i$ implies

$$n(c') \leq \min\left[\prod_{i=1}^{N} e'_i, |X|\right] = \prod_{i=1}^{N} e'_i < \min\left[\prod_{i=1}^{N} e_i, |X|\right] = n(c).$$

Regarding costs, since $\mathbf{e}$ is more efficient than $\mathbf{e}'$, $\kappa[\mathbf{e}] < \kappa[\mathbf{e}']$ for all κ with strictly increasing marginal costs. If κ fails to have strictly increasing marginal costs then there is a sequence $\langle \widehat{\kappa}_n \rangle$ where each $\widehat{\kappa}_n$ has strictly increasing marginal costs and $(\widehat{\kappa}_n(1),...,\widehat{\kappa}_n(N)) \to (\kappa(1),...,\kappa(N))$. Hence $\kappa[\mathbf{e}] \leq \kappa[\mathbf{e}']$ in all cases and so $(\mathcal{C}, c)$ is more efficient than $(\mathcal{C}', c')$. Alternatively, suppose marginal costs are strictly increasing. Then, since $\mathbf{e}$ is more efficient than $\mathbf{e}'$, $\kappa[\mathbf{e}] < \kappa[\mathbf{e}']$ and, since $n(c) \geq n(c')$, $(\mathcal{C}, c)$ is again more efficient than $(\mathcal{C}', c')$. ■

Proof of Theorem 5. Let $x\,R\,y$ mean $x \in c(\{x, y\})$. R is complete. We show that if criteria are binary then R is transitive.

Suppose $x\,R\,y\,R\,z$ and, for any $B \subset \{x, y, z\}$, let $\mathcal{C}(B)$ denote the set of criteria C_i such that B is contained in the top C_i-category and $\{x, y, z\}\backslash B$ is contained in the bottom C_i-category.[21] Since criteria are binary, if a criterion C_i does not place x, y, and z in the same C_i-category then C_i must be an element of one of the following six sets: $\mathcal{C}(\{x\})$, $\mathcal{C}(\{x, z\})$, $\mathcal{C}(\{y\})$, $\mathcal{C}(\{x, y\})$, $\mathcal{C}(\{z\})$, and $\mathcal{C}(\{y, z\})$. Therefore

$$\begin{aligned} U^{xy} &= \mathcal{C}(\{x\}) \cup \mathcal{C}(\{x, z\}), U^{yx} = \mathcal{C}(\{y\}) \cup \mathcal{C}(\{y, z\}), \\ U^{yz} &= \mathcal{C}(\{y\}) \cup \mathcal{C}(\{x, y\}), U^{zy} = \mathcal{C}(\{z\}) \cup \mathcal{C}(\{x, z\}), \\ U^{xz} &= \mathcal{C}(\{x\}) \cup \mathcal{C}(\{x, y\}), U^{zx} = \mathcal{C}(\{z\}) \cup \mathcal{C}(\{y, z\}). \end{aligned}$$

Since $x\,R\,y\,R\,z$, the union assumption implies

$$(\mathcal{C}(\{x\}) \cup \mathcal{C}(\{x, z\}) \cup \mathcal{C}(\{y\}) \cup \mathcal{C}(\{x, y\}))\, D\, \mathcal{C}(\{y\}) \cup \mathcal{C}(\{y, z\}) \cup \mathcal{C}(\{z\}) \cup \mathcal{C}(\{x, z\}).$$

So, by the subtraction assumption, $(\mathcal{C}(\{x\}) \cup \mathcal{C}(\{x, y\}))\; D\; (\mathcal{C}(\{y, z\}) \cup \mathcal{C}(\{z\}))$ and hence $U^{x,z}\, D\, U^{z,x}$. Therefore $x\,R\,z$.

Since R is complete and transitive and any $A \in \mathcal{F}$ is finite, the set $M(A) = \{x \in A : x\,R\,y$ for all $y \in A\}$ is nonempty. By the Condorcet rule, $M(A) = c(A)$. ■

References

[1] Apesteguia, J. and M. Ballester, 2010, 'The computational complexity of rationalizing behavior' *Journal of Mathematical Economics* 46: 356-363.

[2] Apesteguia J. and M. Ballester, 2013, 'Choice by sequential procedures', *Games and Economic Behavior* 77: 90-99.

[3] Arrow, K. and H. Raynaud, 1986, *Social Choice and Multicriterion Decision-Making.* MIT Press: Cambridge.

[4] Bouyssou, D., T. Marchant, M. Pirlot, A. Tsoukias, and P. Vincke, 2006, *Evaluation and Decision Models with Multiple Criteria: Stepping Stones for the Analyst.* Springer: New York.

[5] Chipman, J., 1960, 'The foundations of utility' *Econometrica* 28: 193-224.

[21] For a binary C_i, a C_i-category E is *top* (resp. *bottom*) if there exists $x \in E$ and $y \in X$ such that $x\,C_i\,y$ (resp. $y\,C_i\,x$).

[6] Chipman, J., 1971, 'On the lexicographic representation of preference orderings' in J. Chipman, L. Hurwicz, M. Richter, and H. Sonnenschein (eds.) *Preferences, Utility, and Demand*. Harcourt: New York.

[7] Cowan, N., 2000, 'The magical number 4 in short-term memory: a reconsideration of mental storage capacity,' *Behavioral and Brain Sciences* 24: 87-185.

[8] Engineering Research Associates, Inc., 1950, *High-Speed Computing Devices*. McGraw-Hill: New York.

[9] Figueira, J., S. Greco, and M. Ehrgott, 2005, *Multiple Criteria Decision Analysis: State of the Art Surveys*. Springer: Boston.

[10] Fishburn, P., 1970, *Utility Theory for Decision Making*. Wiley: New York.

[11] Fishburn, P., 1974, 'Lexicographic orders, utilities and decision rules: a survey' *Management Science* 20: 1442-1471.

[12] Frieder, G., 1972, 'Ternary computers, part 1: motivation and part 2: emulation' *Proceedings of the 5th Workshop on Microprogramming*, 83-89.

[13] Gigerenzer, G., P. Todd, and the ABC Research Group, 1999, *Simple Heuristics That Make Us Smart*. Oxford University Press, New York.

[14] Houy, N., and K. Tadenuma, 2009, 'Lexicographic compositions of multiple criteria for decision making' *Journal of Economic Theory* 144: 1770-1782.

[15] Hurst, S., 1984, 'Multiple-valued logic its status and its future' *IEEE Transactions on Computers* 33: 1160-1179.

[16] Inada, K, 1964, 'A note on the simple majority decision rule' *Econometrica* 32: 525-531.

[17] Ju, B., 2011, 'Collectively rational voting rules for simple preferences' *Journal of Mathematical Economics* 47: 143-149.

[18] Kahneman, D., A. Treisman, and B. Gibbs, 1992, 'The reviewing of object files: object specific integration of information' *Cognitive Psychology* 24: 175-219.

[19] Klimenko, S., 1999, 'Computer science in Russia: a personal view' *IEEE Annals of the History of Computing* 29: 16-30.

[20] Knuth, D., 1997, *The Art of Computer Programming, Vol. 2: Seminumerical Algorithms*. Addison-Wesley: Reading MA.

[21] Luck, S. and E. Vogel, 1997, 'The capacity of visual working memory for features and conjunctions' *Nature* 390: 279-281.

[22] Mandler, M., P. Manzini, and M. Mariotti, 2012, 'A million answers to twenty questions: choosing by checklist' *Journal of Economic Theory* 147: 71-92.

[23] Mandler, M., 2009, 'Indifference and incompleteness distinguished by rational trade' *Games and Economic Behavior* 67: 300-314.

[24] Mandler, M., 2012, 'The lexicographic method in preference theory,' mimeo.

[25] Mandler, M., 2015, 'Rational agents are the quickest' *Journal of Economic Theory* 155: 206-233.

[26] Maniquet, F. and P. Mongin, 2015, 'Approval voting and Arrow's impossibility theorem' *Social Choice and Welfare* 44: 519-532.

[27] Manzini, P. and M. Mariotti, 2007, 'Sequentially rationalizable choice' *American Economic Review* 97: 1824-1839.

[28] Manzini, P. and M. Mariotti, 2012, 'Choice by lexicographic semiorders' *Theoretical Economics* 7: 1-23.

[29] Miller, G., 1956, 'The magical number seven, plus or minus two: some limits on our capacity for processing information' *Psychological Review* 63: 81-97.

[30] Parhami, B., and M. McKeown, 2013, 'Arithmetic with binary-encoded balanced ternary numbers' *Proceedings of the 47th Asilomar Conference on Signals, Systems, and Computers*, 1130-1133.

[31] Rubinstein, A., 1996, 'Why are certain properties of binary relations relatively more common in natural language?' *Econometrica* 64: 343-355.

[32] Salant, Y., 2009, 'Procedural analysis of choice rules with applications to bounded rationality' *American Economic Review* 101: 724-748.

[33] Simon, H., 1972, 'Theories of bounded rationality,' in C. McGuire and R. Radner (eds.) *Decision and Organization*. North-Holland, Amsterdam.

[34] Simon, H., 1974, 'How big is a chunk?' *Science* 183: 482-488.

[35] Treisman, A., 2006, 'Object tokens, binding and visual memory, in H. Zimmer, A. Mecklinger, and U. Lindenberger (eds.) *Handbook of Binding and Memory: Perspectives from Cognitive Neuroscience*. Oxford University Press, New York.

[36] Tversky, T., and I. Simonson, 1993, 'Context-dependent preferences' *Management Science* 39: 1179-1189.

[37] Vorsatz, M., 2007, 'Approval voting on dichotomous preferences' *Social Choice and Welfare* 28:127-141.