Coarse, efficient decision-making

Michael Mandler*
Royal Holloway College, University of London

This version: June 2016

Abstract

To achieve efficiency, an agent who uses criteria to make choice decisions should employ the coarse criteria that have a small number of categories. 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; this result holds even if the marginal cost of using more categories declines 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, Todd Kaplan, David Kelsey, Michel De Lara, Christian List, Itai Sher, and Tymon Tatur 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 studies a solution to this problem in which 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, a second that classifies by type of director, a third by quality of the actors, 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.

A set of criteria enjoys 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 \textit{product} of the number of criterion categories. If say 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 typically 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, \( \binom{24}{2} = 276 \) 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.

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

\(^1\)Even rational preferences require a number of judgments, \( n \lceil \log_2 n \rceil \), 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.
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 the binary criteria that partition alternatives into two categories. This result holds even if the marginal decision-making cost of using an additional category diminishes to 0: the categories of fine criteria could become asymptotically free and still it would be optimal to use the expensive categories of coarse criteria. To probe the robustness of this surprising conclusion, I consider an alternative model where criteria vary in value and cost. While optimality then does not require that criteria are 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 better to use many low-value coarse criteria even though the marginal cost of the categories employed would then be greater.

The efficiency advantage of coarse criteria applies to seemingly distant questions including ‘which digits should a memory device use 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 offers the cheapest storage method. Binary digits (bits) are therefore the efficient choice, and of course 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 construction as well.

Binary criteria lead to rational choices when criteria are aggregated in standard ways. If, say, an agent proceeds through binary criteria lexicographically then the choice function that results

\[2\] 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.
maximizes a rational preference. See Mandler et al. (2012) and Mandler (2015). 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: for example, a rational choice function will arise whenever an agent uses a weighted vote of binary criteria to determine choice. ³

To show that the binariness-rationality connection runs deep, I provide an axiomatic treatment. 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, let the number of choice distinctions – *choice classes* in our terminology – that an agent makes equal the product of the number of categories in the criteria deployed. (This is the maximum number of choice classes that can be generated given the number of categories in each criterion.) Suppose the *j*th of a set of *N* criteria uses *e*ₗ > 2 categories and we replace it with a criterion that uses *e*ₗ − 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. And if the marginal cost of using additional categories in a criterion is increasing – which is likely since the number of ranking decisions that a criterion requires is a convex function of the number of categories – the shift to the coarser set of criteria strictly lessens costs.

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

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.
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 \(C_1, C_2, C_3\) defined as follows (with higher alternatives better than lower):

\[
\begin{array}{ccc}
C_1 & C_2 & C_3 \\
x & y & z \\
y & z & x \\
z & x & y \\
\end{array}
\]

If the choice function \(c\) decides by a simple majority vote of the criteria, then choices (or revealed preferences) will cycle on the pairs – \(c(\{x, y\}) = \{x\}\), \(c(\{y, z\}) = \{y\}\), \(c(\{x, z\}) = \{z\}\) – 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 list of such criteria and any choice set of alternatives \(A\), 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.  \(^{6}\) 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.  \(^{7}\) 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 that needs to be manipulated in working memory could have at most three or possibly four categories. Consider the movie example: if an agent wants to choose a movie with a genre criterion

\(^{5}\)This example employs rational criteria, which our general model does not require.

\(^{6}\)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)).

\(^{7}\)See Luck and Vogel (1997) for a characteristic example of the research surveyed.
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 (see, 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. Let $\mathcal{C} = \{C_1, ..., C_N\}$ denote a set of criteria. The number $N$ will denote the number of criteria in $\mathcal{C}$ (and $N'$ the number of criteria in $\mathcal{C}'$). Criterion indices do not indicate the order in which criteria are consulted.

To measure coarseness, we define a criterion’s categories or equivalence classes, 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

\[
x I_i y \iff \{ z \in X : z \, C_i \, x \} = \{ z \in X : z \, C_i \, y \} \text{ and } \{ z \in X : x \, C_i \, z \} = \{ z \in X : y \, C_i \, z \},
\]

and a \( C_i \)-category is a maximal \( E \subset X \) such that, for all \( x, y \in E \), \( x I_i y \).\(^8\)

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, that is, 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, e.g., in the Condorcet example in the introduction where each \( C_i \) is in fact fully rational.

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. While we will not need to require explicitly that each \( C_i \) orders a distinct attribute and that \( X \) equals a product of attributes, each use of criteria in the paper is consistent with these assumptions (with one adjustment in footnote 12).

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. A criterion with a single category requires no decisions and distinguishes between no pair of alternatives and so we assume it is 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 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.\(^9\)

\(^8\)Equivalently, a nonempty \( E \subset X \) is a \( C_i \)-category if, for all \( x \in E \), \( x I_i y \iff y \in E \). When \( C_i \) is transitive, see Fishburn (1970) and Mandler (2009) for discussions of \( I_i \).

\(^9\)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.
Definition 2 Given a choice function \( c \), alternatives \( x \) and \( y \) are elements of the same choice class if and only if for all \( A \subseteq X \),

(i) if \( \{x, y\} \subseteq A \) then \( x \in c(A) \iff y \in c(A) \),

(ii) if \( \{x, y\} \) does not intersect \( A \) then, for all \( z \in A \),

\[
x \in c(A \cup \{x\}) \iff y \in c(A \cup \{y\}),
\]

\[
z \in c(A \cup \{x\}) \iff z \in c(A \cup \{y\}).
\]

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 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 \( \mathcal{P} \in \mathcal{P} \) if and only if, for each \( C_i \in \mathcal{C} \), \( x \) and \( y \) are contained in the same \( C_i \)-category. A choice function \( c \) uses the set of criteria \( \mathcal{C} \) if the partition of choice classes is coarser than \( \mathcal{P} \) (if \( x, y \in \mathcal{P} \in \mathcal{P} \) then there is a choice class \( E \) of \( c \) such that \( x, y \in E \)).

In the movie example at the beginning of the paper, \( \mathcal{P} \) has 24 cells, each equal to an intersection of one of the 4 genre categories, one of the 3 director categories, and one of the 2 actor categories.

We use the notation \((\mathcal{C}, c)\) to indicate that \( c \) uses \( \mathcal{C} \).

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 \( C_i \), 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 \( w_i \). The sum of the weighted votes for alternative \( x \) when \( x \) is in choice set \( A \) is then given by \( v(x, A) \equiv \sum_{y \in A} \sum_{i \in \{1, \ldots, N\}} w_i s_i(x, y) \). It is easy to check that the choice function \( c \) defined by

\[
x \in c(A) \iff (v(x, A) \geq v(y, A) \text{ for all } y \in A)
\]
uses $C$. The majority vote of criteria in the introduction amounts to a special case of this $c$ in which all of the $w_i$ 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

Our agents pursue two goals. First, they seek choice functions with larger numbers of choice classes: if choice classes reflect an agent’s true preference distinctions, an increase in their number will let the agent choose a superior alternative from more choice sets. In section 4.2, we link this goal to classical utility maximization. Second, they wish to decrease 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, costs will be determined by the number of $C_i$-categories: for all criteria $C_i$ and $\tilde{C}_i$, $e(C_i) = e(\tilde{C}_i) \Rightarrow \kappa(C_i) = \kappa(\tilde{C}_i)$. 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 all $c'$ that use $\mathcal{C}'$. 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 actor 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.\(^{10}\)

If \((\mathcal{C}, c)\) is efficient then \(n(c) \geq n(c')\) must hold for any \((\mathcal{C}', c')\) that satisfies the constraint 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 this constraint, consider some \((\mathcal{C}, c)\) and the inequalities

\[
n(c) \leq |\mathcal{P}| \leq \min \left[ \prod_{i=1}^{N} e(C_i), |X| \right], \tag{3.1}
\]

where \(\mathcal{P}\) is the discrimination partition that identifies the cells of alternatives that are not distinguished by any criterion (Definition 3). Since the alternatives in different choice classes must be distinguished by at least one criterion, the number of choice classes cannot exceed \(|\mathcal{P}|\) and hence the first inequality above holds. While \(n(c) = |\mathcal{P}|\) is not guaranteed – an agent might decide to ignore a criterion that, say, categorizes foods by color when the agent cares only about taste – equality will hold for generic weighted-voting choice functions. As for the second inequality, since each cell of \(\mathcal{P}\) equals an intersection of \(C_i\)-categories (one category per criterion), \(|\mathcal{P}|\) is bounded by the product \(\prod_{i=1}^{N} e(C_i)\). Equality moreover will obtain 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 2, 3, and 4 categories can distinguish 24 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 be redescribed as a product of attributes.\(^{11}\)

Since both inequalities in (3.1) can be satisfied with equality, 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.

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.

\(^{10}\) Proofs omitted from the text are in the Appendix.

\(^{11}\) More generally, to ensure that \(|\mathcal{P}| = \min \left[ \prod_{i=1}^{N} e(C_i), |X| \right]\) holds it is sufficient to choose \(C_i\) so that each intersection of categories (one per criterion) is nonempty. Cases where \(|\mathcal{P}| < \min \left[ \prod_{i=1}^{N} e(C_i), |X| \right]\) holds are certainly possible however, for example, when two criteria in \(\mathcal{C}\) simply repeat each other.
Since we assume for now that the cost of a criterion $\kappa(C_i)$ is determined by its number of categories, we can use the notation $\kappa(e)$, which is defined to equal $\kappa(C_i)$ for any criterion $C_i$ such that $e(C_i) = e$. We assume throughout that costs are increasing: $e' > e \Rightarrow \kappa(e') > \kappa(e)$.

The formation of a criterion requires an agent to partition $X$ into categories and then order these categories. A criterion with a single category requires neither partitioning nor ordering and makes no choice distinctions. We therefore assume it is free: henceforth, for any criterion $C_i$, $e(C_i) = 1$ implies $\kappa(C_i) = 0$. But any further expansion of the number of categories $e$ incurs both types of costs. While the partitioning cost might be a linear function of $e - 1$, an ordering of categories requires an agent to decide, for any pair of categories, if they are ranked and if so which is superior. Since the number of pairs equals $\binom{e}{2} = \frac{e(e-1)}{2}$, a strictly convex function of $e$, there is a strong case that the marginal cost of categories will be strictly increasing in $e$.

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 will require 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 inequality (3.1)) 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 $2\kappa(3)$. Since 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 in fact 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 can even fall to 0 (as quickly as the derivative of any increasing power function falls to 0 but not quicker than $\frac{1}{e}$). Since our earlier arguments that the marginal cost of criterion categories should be increasing remain sound, Theorem 1 indicates that adjustments to our conceptual framework, if not too dramatic, will not overturn the optimality of binary criteria. For 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.

---

11 Example 2 showed that with a fixed domain the inequality in Theorem 1 can hold with equality when $\prod_{i=1}^{N} e(C_i) < |X|$ obtains. To cover fixed domains, the inequality in Theorem 1 would be replaced by:

$$\kappa(e) > \kappa(2) \log_2 e$$

for all integers $2 \leq e \leq |X|$. This allows marginal costs to be constant from $e = 1$ to $e = 3$ and to be decreasing afterwards. When $X$ is a product of attributes and criteria order attributes and are not maximally fine then $\prod_{i=1}^{N} e(C_i) < |X|$ will obtain and the above qualification applies.

12 To outline the proof of Theorem 1, suppose the second statement is violated for some $e > 2$. Then, when $|X| = e$, it would be at least weakly cheaper to use a single criterion $C_1$ with $e$ categories than to use $\lceil \log_2 e \rceil$ binary criteria. Since $\lceil \log_2 e \rceil$ is the smallest integer $r$ such that $2^r \geq e$, fewer than $\lceil \log_2 e \rceil$ binary criteria must generate fewer than $e$ choice classes (see inequality (3.1)). So criterion $C_1$ on its own, joined to a $c$ with choice classes that coincide with the $C_1$-categories, cannot be dominated by any $(\mathcal{C}, c)$ with only binary criteria. Conversely, if the second statement is satisfied then costs will fall if any criterion with $e > 2$ categories is replaced by $\lceil \log_2 e \rceil$ binary criteria and a simple calculation shows that the number of choice classes can then be weakly increased.
4.1 The costs of fine criteria and direct preference construction

When marginal category costs are increasing, Theorem 1 suggests that the penalty for using fine criteria is substantial. If each criterion is constrained to have \( k \) categories, the minimum cost of a set of criteria that generates a choice function with \( n \) choice classes is \([\log_k n] \kappa(k)\) (since \([\log_k n]\) is the minimum integer \( r \) such that \( k^r \geq n \)). For estimation purposes, we ignore the difference between \([\log_k n]\) and \( \log_k n \). The ratio of the minimum costs of a set of \( k \)-ary criteria and a set of binary criteria, when both generate \( n \) choice classes, is then approximately

\[
\frac{\kappa(k) \log_k n}{\kappa(2) \log_2 n} = \frac{\kappa(k)}{\kappa(2) \log_2 k}.
\]

Recalling that \( \frac{k(k-1)}{2} \) is a plausible cost function, suppose \( \kappa \) is linear or superlinear in \( k \). Then, due to the \( \log_2 k \) term in the denominator, the above ratio grows rapidly as a function of \( k \): 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 \((k = n)\) which is the traditional account where agents make direct preference judgments. The penalty exacted by direct preference construction would eventually become unsustainable as \( n \) increases: agents would have to turn to some cost-reduction strategy.

This estimate of the cost of fineness casts an 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 may even be the outcome of optimizing adaptations.

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

Comparisons aside, the cost of using binary criteria, \( \kappa(2) \) \([\log_2 n]\), increases slowly as a function of \( n \), as does the cost of using \( l \)-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 \( l \) that makes \( n \) choice distinctions increases only on the order of \( \log n \).
4.2 Diverse criteria

Criteria can vary by how costly their categories are and by the value of the distinctions their categories make. With movies, for example, genre distinctions are presumably cheaper to discover than quality-of-director distinctions and, depending on the individual, may have greater or lesser value. It might then 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 those criteria would become finer and finer. But in fact the advantage of coarse criteria continues to prevail in this setting: 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 criterion values, each criterion index will now identify a fixed attribute of the domain, e.g., genre or director-quality. An agent in this section will select a set of criteria $C$ that, for each attribute $i$, contains either one or no criterion with that attribute (rather than criteria always having indices $1, \ldots, N$). Let $\{C_i\}$ denote the feasible criteria for attribute $i$. To make sure that the program of showing that criteria should be coarser than a fixed ceiling is a nontrivial goal, 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 $e > 1$.

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 equal 1 otherwise. A set of criteria $C$ then has the discrimination value

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

which replaces the number of choice classes as the agent’s objective function.\(^{14}\) The value of a criterion can increase without bound as its number of categories increases, a potential advantage for the fine criteria; a classical utility-maximization model (see the end of this section) would, in contrast, lead to a $V$ where criteria have bounded value. Coarse criteria on the other hand can have negligible value since $v(C_i)$ can be near 1. Despite these biases in favor of 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 special case occurs when these two forces can be disentangled and the value 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 had $w_i(e) = 1$ for all $e$).

\(^{14}\)Our original model is the special case where $v(C_i) = e(C_i)$ for all $C_i$. 13
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 allows 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 2 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 $\kappa_i$ 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 $\bigcup_{i \in \mathbb{N}} \{\kappa_i\}$ is compact in the topology of uniform convergence on $\mathbb{N}$. Define $\kappa_i$ 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 2** 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 $b$ categories.

The message of Theorem 2 is surprising: even when aiming for a large number of choice dis-

\[ \text{sup}_{e \in \mathbb{N}} |\kappa_{i_n}(e) - \kappa_{i_{n+1}}(e)| \text{ converges to 0 as } n \to \infty. \]
tinctions 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 2 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 efficiency will require that the criteria for these indices are all binary, assuming that the log cost condition holds.

Utility-maximizing criteria Maximization of a standard utility fits into the diverse criteria model and our assumptions on discrimination value will then be satisfied with room to spare. Let $X$ be a product of $n$ attributes $X = \prod_{i=1}^{n} X_i$ where we take $n$ to be large enough to accommodate Theorem 2 or include only those attributes ordered by some criterion in use. An agent receives a utility for $x \in X$ given by $u(x) = \sum_{i=1}^{n} u_i(x_i)$ which will be a random variable. To ensure that the agent’s preferences are compatible with expected utility theory, we assume that the values that each $u_i$ can attain are bounded. A state specifies $u_i(x_i)$ for all $i$ and $x_i$, and, for each available criterion $C_i$, the partition of $X$ into $C_i$-categories. An agent who uses a set of criteria $C$ learns the partitions of the $C_i$ in $C$, which we indicate by conditioning on $C$.

Each criterion orders only one attribute and so each $C_i$-category has the form $\overline{X}_i \times \prod_{j \neq i} X_j$ where $\overline{X}_i \subset X_i$. We identify a $C_i$-category with its $\overline{X}_i$. To make sure that alternatives that are not distinguished by any criterion are expected-utility equivalent and that a criterion for one attribute does not provide information about other attributes, we assume that, when an agent uses a set of criteria $C$ that contains $C_i$, the alternatives within a $C_i$-category all deliver an expected utility level equal to what could be inferred from $C_i$ alone. Formally, for each $C$ and $C_i \in C$ and each $C_i$-category $\overline{X}_i$, there is a constant, written $E [u_i(\overline{X}_i)|C_i]$, such that $E [u_i(x_i)|C] = E [u_i(x_i)|C_i] = E [u_i(\overline{X}_i)|C_i]$ for all $x_i \in \overline{X}_i$.\footnote{We here let $C_i$ also denote the set of criteria that contains only $C_i$.}

If there is no criterion for attribute $i$ in $C$ this requirement is still required to hold, with $C_i$ and $\overline{X}_i$ replaced by $\emptyset$ and $X_i$. We assume that each available $C_i$ is useful: if $e(C_i) > 1$ then there is a pair of distinct $C_i$-categories $\overline{X}_i$ and $\overline{X}'_i$ such that $E [u_i(\overline{X}_i)|C_i] \neq E [u_i(\overline{X}'_i)|C_i]$.

In this application we take $X$ itself to be the choice set. Given $C$ and a $C_i \in C$, let $\overline{X}_i$ denote a utility-maximizing $C_i$-category, i.e., $E [u_i(\overline{X}'_i)|C_i] \geq E [u_i(\overline{X}_i)|C_i]$ for each $C_i$-category $\overline{X}_i$. A utility-maximizing way to choose is then to let $c(X)$ equal $\overline{X}_1^* \times ... \times \overline{X}_n^*$ or any of its subsets and we therefore define the maximum utility achievable with $C$ by $U(C) = \sum_{i=1}^{n} E [u_i(\overline{X}_i^*)|C] = \sum_{i=1}^{n} E [u_i(\overline{X}_i)|C_i]$. \footnote{We here let $C_i$ also denote the set of criteria that contains only $C_i$.}
\[ \sum_{i=1}^n \mathbb{E} \left[ u_i(X_i^*)|C_i \right]. \] With \( X \) as the choice set, the agent has no use for any category of \( C_i \) besides \( X_i^* \) but the research needed to discover \( X_i^* \) would reveal various other categories: the information that lets you identify the top 10% of movie directors would let you pin down several other brackets.

Letting \( U_i(C_i) = \mathbb{E} \left[ u_i(X_i^*)|C_i \right] \), simple monotonic transformations of the \( U_i \) and \( U \) then confirm that \( U \) orders sets of criteria as a discrimination value function \( V \) can. These transformations generate a value for an individual criterion, \( v(C_i) \), that is bounded above.

To get a feel for actual numbers, suppose for each attribute \( i \) that \( u_i(x_i) \) lies in \([-\frac{1}{2}, \frac{1}{2}]\) and that, for every \( C_i \) and \( x_i \in X_i \), \( u_i(x_i) \) is uniformly distributed over an interval of length \( \frac{1}{e(C_i)} \) (which implies that \( u_i(x_i) \) and \( u_i(x_i') \) have the same support when \( x_i \) and \( x_i' \) lie in the same \( C_i \)-category).

Easy calculations show the following relationship between the number of \( C_i \)-categories and the expected utility that \( C_i \) delivers:

<table>
<thead>
<tr>
<th>( e(C_i) )</th>
<th>( U_i(C_i) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>.25</td>
</tr>
<tr>
<td>3</td>
<td>.30</td>
</tr>
<tr>
<td>4</td>
<td>.31</td>
</tr>
<tr>
<td>( \infty )</td>
<td>.33</td>
</tr>
</tbody>
</table>

where the bottom row lists a perfectly discriminating criterion. Thus in this example two or three categories deliver the lion’s share of a criterion’s potential discrimination gains. ■

4.3 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 identify with the integers \( 0, \ldots, n-1 \). Each of the \( n \) integers can be coded using \( k \)-ary digits and each digit can therefore be stored in a device that attains \( k \) states. We assume, uncontroversially, that the cost of a digit-storing device, labeled \( \kappa(k) \), is increasing in the number of states \( k \) it can attain. Since \( N \ k \)-ary digits can encode \( k^N \) integers, the number of devices needed to store \( n \) facts equals \( \lfloor \log_k n \rfloor \), the

\[ \text{Subtract the constant } \mathbb{E} \left[ u_i(X_i)|\emptyset \right] \text{ from } u_i, \text{ thus defining } \tilde{u}_i, \tilde{U}_i, \text{ and } \tilde{U}. \text{ Then set } V(C) = \exp \left( \tilde{U}_i(C) \right). \text{ The corresponding values for individual criteria are then given by } v(C_i) = \exp \left( \tilde{U}_i(C_i) \right). \text{ The utility delivered by a perfectly discriminating criterion, } \sup_{x \in X} \tilde{u}_i(x), \text{ must be at least as great as } \tilde{U}_i(C_i) \text{ for any available } C_i; \text{ and since each } C_i \text{ is useful, } \tilde{U}_i(C_i) > \mathbb{E} \left[ \tilde{u}_i(X_i)|\emptyset \right] = 0 \text{ when } e(C_i) > 1. \text{ So } v(C_i) \text{ is bounded above, greater than 1 when } e(C_i) > 1, \text{ and equal to 1 otherwise: } V \text{ thus qualifies as a discrimination value function. To conclude that, as the number of criteria in use increases, } v(C_i) \text{ is bounded away from 1 when } e(C_i) > 1, \text{ we would need to assume that } \mathbb{E} \left[ u_i(X_i)|C_i \right] - \mathbb{E} \left[ u_i(X_i)|\emptyset \right] \text{ is bounded away from 0.} \]
least integer $N$ such that $k^N \geq n$. We therefore take the cost of storing one of $n$ integers 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 $C$. If, in the decision-making model, each criterion is required to have the same number of categories $k$ and we fix a target number of choice classes $n$ then efficiency demands that we choose $k$ to minimize $\kappa(k)N$ subject to the constraint that there is a $c$ with $n$ choice classes that uses $N$ $k$-ary criteria. As the proof of Proposition 1 shows, $N$ must equal the least integer such that $k^N \geq n$, that is, $\lceil \log_k n \rceil$.

The correspondence between these two problems implies that if $(C, c)$ is efficient in the decision-making model and every criterion in $C$ 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 $\kappa$ 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, which fits well with the increasing-returns character of data storage technologies.

Our analysis sets aside the empirical fact that current-day computers store information in devices (transistors) that assume two states and that binary storage devices are now extraordinarily cheap. The question is: why is it that binary devices 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 $n$, rather than jointly in terms of both cost and the number of integers to be stored. The storage cost $\kappa$ should now be easier to motivate since it represents the physical cost of building a device rather than psychological effort. Second, the objects to be stored are integers 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. When applied to information storage, our earlier set-up would thus allow integers to be represented and stored as mixed-radix numerals (see Knuth (1997)).

Although these differences mean that we need new definitions for the information-storage model, the proof of Theorem 3 is similar to the proof of Theorem 1; indeed the transferability of arguments

---

18 There is no difficulty in providing a binary efficiency theorem comparable to Theorem 3 below for mixed-radix storage media.
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 $n$ integers 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 3** The following statements are equivalent:

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

In contrast to the cost of categories in decision-making, the marginal cost of storing an additional digit is presumably decreasing in the number of digits $k$, due to economies of scale. So even though the log cost condition in Theorem 3 allows the marginal cost of additional digits to converge to 0, the condition might nevertheless be violated for some $k$’s. However, since $\log_2 k$ increases so slowly, $\kappa(2) \lceil \log_2 k \rceil$ is likely to fall below $\kappa(k)$ for all sufficiently large $k$ and it is therefore implausible a large $k$ would qualify as an efficient base. For small $k > 2$, 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)).

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 more demanding assumption that the log cost condition – then any move from finer to coarser criteria brings a gain in efficiency.

To compare the coarseness of sets of criteria with different numbers of criteria, it is the distribution of the number of costly categories that is economically relevant. Since single-category criteria, which make no choice distinctions, cost 0, we label the number of costly categories of a criterion $C_i$ by $e^*_i = e_i - 1$ (where as usual $e_i = e(C_i)$). Call the vector of positive numbers $(e_1, ..., e_N)$ the discrimination vector of $C = \{C_1, ..., C_N\}$. Following the analogy of first-order stochastic dominance, we consider $C$ to be coarser than $C'$ if the proportions of costly categories that are smaller than any given level is greater for the discrimination vector of $C$ than for the

c\footnote{Let me thank Itzhak Gilboa for telling me about ternary computers.}
discrimination vector of $C'$. Given a discrimination vector $e = (e_1, ..., e_N)$ such that some $e_i > 1$, and given the integer $k \geq 1$, let $p_k(e)$ denote the proportion of $\sum_{i=1}^{N} e_i^*$ that satisfies $e_j^* \leq k$:

$$p_k(e) = \frac{\sum_{i \in \{i | e_i^* \leq k\}} e_i^*}{\sum_{i=1}^{N} e_i^*}.$$

We define the set of criteria $C$ with the discrimination vector $e$ to be coarser than the set $C'$ with the discrimination vector $e'$ if, for each integer $k \geq 1$, $p_k(e) \geq p_k(e')$ and strict inequality obtains for some $k \geq 1$.\(^{20}\)

Greater coarseness does not by itself imply an increase in efficiency. First, coarseness measures the distribution of categories not their aggregate quantity: $C$ could be coarser than $C'$ but the cost of $C$ could be so large that $C$ and $C'$ are not efficiency ranked. A pure advantage of coarseness therefore can appear only when the number of costly categories in $C$ and $C'$ 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 provide terminology to deal with these points, define marginal costs to be increasing if, for all positive integers $e$ and $e'$,

$$e' > e \Rightarrow (\kappa(e') > \kappa(e) \text{ and } \kappa(e' + 1) - \kappa(e') \geq \kappa(e + 1) - \kappa(e)),$$

and strictly increasing if they are increasing and the weak inequality in MC is replaced by a strict inequality. We say that $C$ and $C'$ 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

$$\min \left[ \prod_{i=1}^{N} e_i, \prod_{i=1}^{N'} e_i' \right] < |X| \text{ or marginal costs are strictly increasing.}$$

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

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 $C$ and $C'$ as given in the Theorem, we can append enough single-category criteria to $C'$ to equalize the number of criteria in $C$ and $C'$ without affecting the cost of $C'$ or the number of choice classes of $c'$'s that use $C'$. Figure 1 illustrates with a finer (solid, blue) set $C'$ of three criteria and a coarser

\(^{20}\)If we defined coarseness using the distribution of the $e_i$, we could take any discrimination 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.
(dashed, red) set $C$ 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 $C'$ to equalize the number of criteria. Now compare the criteria in $C$ and the amended version of $C'$ 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 $C$ and the fact that $C$ and $C'$ have the same number of costly categories imply that at the first point where $C$ and $C'$ differ, it will be the criterion in $C'$ – call it $C'_{\text{more}}$ – that has more categories. Reduce the number of categories in this criterion by 1 and increase by 1 the number of categories in some other criterion $C'_k$ in $C'$ that has at least two fewer categories than $C'_{\text{more}}$ to create a new set $C''$. 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 $C''$ can weakly increase and can strictly increase if $\prod_{i=1}^{N} e'_i < |X|$. Due to the tight-comparison assumption, $C''$ is strictly more efficient than $C'$ (though it need not be coarser).

Since the number of costly categories is the same in $C$ and $C'$, there is a sequence of such steps that terminate in a set of criteria with the same cost as $C'$ 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, $C$ must be more efficient than $C'$.

In section 4.2, 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 now 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: it will be suboptimal for one of these criteria to be finer than the rest.

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, and each criterion has equal
Figure 1: The move from a fine to a coarse distribution of categories
weight, then revealed preferences will cycle on the pairs of alternatives. Given the ‘possibility’ results in voting models with dichotomous preferences (see footnote 6), one would expect that binary criteria aggregate well.

**Definition 6** Let $C$ be a set of criteria and $w$ the criterion weights $(w_1, ..., w_N)$. Then $c$ is a \textit{w-weighted-voting choice function that uses} $C$ if $c$ uses $C$ and, for all $A \in F$, $c(A)$ equals

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

when this set is nonempty.\(^{21}\)

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

**Proposition 2** Given any criterion weights $w$, if each criterion in the set $C$ is binary then the $w$-weighted-voting choice function that uses $C$ is rational.

**Proof.** For $C = \{C_1, ..., C_N\}$, 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 $x, y \in X$,

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

Due to the finiteness of $A$, $\sum_{i=1}^{N} u_i(x)$ 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 $\succeq$ on $X$ defined by $x \succeq y$ 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. \(\blacksquare\)

The reach of Proposition 2 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.

\(^{21}\)Recall from section 2 that, for any pair $x, y$,

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

- \( x \in c(x, y) \Rightarrow U^{x,y} \cup U^{y,x} \) (decisiveness),
- \( UDV, U'DV', and U \cap U' = \emptyset \Rightarrow (U \cup U')D(V \cup V') \) (union),
- \( UDV \) and \( W \subset (U \cap V) \Rightarrow (U \setminus W)D(V \setminus W) \) (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 divide each set of meals not discriminated by the other criteria into two parts. 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, cost, spiciness – 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 3 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 $B \subseteq X$ and $a \in X$, let $B – a$ denote $B \setminus \{a\}$ 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 \( B \cap \{x, z\} = \emptyset \) 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 \( B \cap \{x, z\} = \emptyset \) and \( x \in c(B + x) \) imply \( z \in c(B + z) \).

Finally assume \( B \cap \{x, z\} = \emptyset, w \in B, \) 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 \( B \cap \{x, z\} = \emptyset, w \in B, \) and \( w \in c(B + x) \) imply \( w \in c(B + z) \). \( \blacksquare \)

**Proof of Proposition 1.** Suppose \((C, c)\) is efficient. Since \( c \) uses \( C \), \( n(c) \leq |P| \). Moreover, given \( C \) and hence \( \mathcal{P} \), there exists a \( \hat{c} \) that uses \( C \) such that \( n(\hat{c}) = |\mathcal{P}| \). 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 \( \hat{c} \) select from any choice set \( A \) only those alternatives \( x \in A \) with the largest \( R(x) \): \( \hat{c}(A) = \{x \in A : R(x) \geq R(y) \text{ for all } y \in A\} \). It is easy to confirm that \( \hat{c} \) uses \( C \), 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 \( \{x, y\} \in \mathcal{F} \), \( \hat{c}(\{x, y\}) = \{x, y\} \). Therefore \( R(x) = R(y) \) and hence \( x \) and \( y \) are in the same cell of \( \mathcal{P} \). Since therefore \( n(\hat{c}) = |\mathcal{P}| \), 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 \( C \) such that \( e(C_i) = e_i \) for each \( i \), assume first that \( \prod_{i=1}^{N} e_i \leq |X| \). Set \( \hat{C}_1 \) to have \( e_1 \) categories with cardinalities that, if \( X \) is finite, equal \( \left[ \frac{|X|}{e_1} \right] \) or \( \left[ \frac{|X|}{e_1} \right] + 1 \) and otherwise are all infinite. Proceeding by induction, suppose that \( \{\hat{C}_1, ..., \hat{C}_t\} \) 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 \( \hat{C}_{t+1} \) 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 \( [e_{t+1}, \left\lfloor \frac{m}{e_{t+1}} \right\rfloor + 1] \), 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 \( \hat{C}_{t+1} \) by \( x \hat{C}_{t+1} y \iff k_x \geq k_y \). Then \( \hat{C}_{t+1} \) 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 \([\cdot]\), either \( \left\lfloor \frac{|X|}{\prod_{i=1}^{t+1} e_i} \right\rfloor \) 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 \( \{\hat{C}_1, ..., \hat{C}_N\} \) 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 \(\{C'_1, ..., C'_{N'}\}\) determines a partition \(\mathcal{P}'_t\) of \(X\) and \(|\mathcal{P}'_t| \leq e_1\). 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 \((P_{t+1} \in \mathcal{P}'_{t+1} \text{ and } P_{t+1} \subset P_t)\) 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 \(|\mathcal{P}'_{t+1}| \leq e_{t+1} (\prod_{i=1}^{t} e_i)\) and we conclude that \(|\mathcal{P}'_{N'}| \leq \prod_{i=1}^{N'} e_i\).

When \(\prod_{i=1}^{N} e_i > |X|\), terminate the recursion used to construct the \(\hat{C}_i\) at the largest \(t\) such that \(\prod_{i=1}^{t} e_i \leq |X|\). To define \(\hat{C}_{t+1}\), let each partition \(\mathcal{P}(P_t)\) have \(|P_t|\) rather than \(e_{t+1}\) cells. The remainder of the construction of \(\hat{C}_{t+1}\) is unchanged, leading to a \(\hat{C}_{t+1}\) with \(\max\{|P_t| : P_t \in \mathcal{P}_t\}\) categories, and a partition \(\mathcal{P}_{t+1}\) with \(|X|\) cells. The criteria \(\hat{C}_{t+2}, ..., \hat{C}_N\) 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 \((\hat{C}_1, ..., \hat{C}_N)\) such that \((e(\hat{C}_1), ..., e(\hat{C}_N)) = (e_1, ..., e_N)\)) 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 \(\kappa(e) > \kappa(2) [\log_2 e]\) 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 \(N - 1 + [\log_2 e]\) criteria such that, for \(j \in \{1, ..., N\} \setminus \{i\}\), \(e(C'_j) = e(C_j)\) and where the remaining \([\log_2 e]\) 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 \(|\mathcal{P}'| = \min [\prod_{j=1}^{N-1+\lfloor \log_2 e \rfloor} e'_j, |X|]\) and \(n(c') = |\mathcal{P}'|\). Since \(\kappa(\mathcal{C}) = \sum_{j=1}^{N} \kappa(C_j)\) and \(\kappa(\mathcal{C}') = \sum_{j \in \{1, ..., N\} \setminus \{i\}} \kappa(C_j) + \kappa(2) [\log_2 e]\),

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

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

\[
\left( \prod_{j=1}^{N-1+\lfloor \log_2 e \rfloor} e'_j \right) - \left( \prod_{j=1}^{N} e_j \right) = \left( \prod_{j \in \{1, ..., N\} \setminus \{i\}} e_j \right) \left( 2^{\lfloor \log_2 e \rfloor} - e \right) \geq 0,
\]

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

**Proof of Theorem 2.** We record that the topology on $\bigcup_{i \in \mathbb{N}} \{k_i\}$ is a metric space with $\delta(k_i, k_j) = \min \{1, \sup \bigcup_{e \in \mathbb{N}} \{|k_i(e) - k_j(e)|\}\}$ serving as the metric.

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

$$\bar{\kappa}(\tilde{c}^{n_k}) \leq \sup k_j(2) \lceil (\tilde{c}^{n_k})^{\frac{1}{n_k}} \bar{v} \rceil + (\bar{\kappa}(\tilde{c}^{n_k}) - \kappa^n(\tilde{c}^{n_k})),$$

and hence, since the metric $\delta$ implies $\kappa^n(\tilde{c}^{n_k}) - \bar{\kappa}(\tilde{c}^{n_k}) \to 0$ and since $(\tilde{c}^{n_k})^{\frac{1}{n_k}} \bar{v} \to \infty$, we conclude that

$$\bar{\kappa}(\tilde{c}^{n_k}) \leq \sup k_j(2) \lceil (\tilde{c}^{n_k})^{\frac{1}{n_k}} \bar{v} \rceil + (\bar{\kappa}(\tilde{c}^{n_k}) - \kappa^n(\tilde{c}^{n_k})) \leq 2 \sup k_j(2) \lceil (\tilde{c}^{n_k})^{\frac{1}{n_k}} \bar{v} \rceil$$

for all $n_k$ sufficiently large, a contradiction. Hence there exist $\bar{v} > 0$ and an integer $\bar{b} > 0$ such that $\inf k_i(e) > \sup k_j(2) \lceil e^\bar{v} \rceil$ for all $e > \bar{b}$.
To prove the Theorem, suppose the set of \( N' \) criteria \( C' \) contains a \( C_k' \) with \( e \) categories. Define \( \mathcal{C} \) to coincide with \( C' \) except that \( C_k' \) is replaced by \( \lfloor e \pi \rfloor \) binary criteria with indices \( N' + 1, \ldots, N' + \lfloor e \pi \rfloor \). Due to the previous paragraph, \( \kappa(\mathcal{C}) < \kappa(C') \) if \( e > b \). As for value, 

\[
V(C') = \prod_{j \in \{1, \ldots, N'\}} v(C'_j) \quad \text{and} \quad V(\mathcal{C}) = \prod_{j \in \{1, \ldots, N'\}} v(C'_j) \quad \text{and} \quad V(\mathcal{C}) \geq \frac{1}{v(e \pi)} \left( \prod_{j \in \{1, \ldots, N'\}} v(C'_j) \right).
\]

Hence

\[
V(\mathcal{C}) - V(C') \geq \frac{1}{v(e \pi)} \left( \prod_{j \in \{1, \ldots, N'\}} v(C'_j) \right) - \prod_{j \in \{1, \ldots, N'\}} v(C'_j)
= \left( \prod_{j \in \{1, \ldots, N'\}} v(C'_j) \right) \left( \frac{1}{v(e \pi)} - 1 \right).
\]

Since \( \frac{1}{v(e \pi)} \geq \frac{1}{v(e \pi)} \) and \( v(e \pi) \geq v(C_k') \), if \( \frac{1}{v(e \pi)} > v(e \pi) \) for all \( e \) sufficiently large then there is an integer \( b' \) such that \( V(\mathcal{C}) - V(C') > 0 \) when \( e > b' \). To conclude that \( \frac{1}{v(e \pi)} > v(e \pi) \) for all \( e \) sufficiently large, it is sufficient that any of the following equivalent conditions

\[
\frac{e \pi}{\ln e} \rightarrow \infty \iff \ln \frac{e \pi}{\ln e} \rightarrow \infty \iff e \pi \ln e - \ln \pi e \rightarrow \infty
\]

obtains. The last condition follows from \( \frac{\ln e}{e} \rightarrow 0 \) and the implications

\[
\frac{\ln e}{e} \rightarrow 0 \Rightarrow \frac{\frac{1}{\ln e}}{e} \rightarrow 0 \Rightarrow \frac{\ln e}{e \ln e} + \frac{\frac{1}{e \ln e}}{e} \rightarrow 0 \Rightarrow \ln \pi + \ln e \rightarrow 0.
\]

So set \( b \) in the Theorem equal to \( \max\{\bar{b}, b'\} \). \( \blacksquare \)

**Proof of Theorem 3.** Assume that \( \kappa(k) > \kappa(2) \lfloor \log_2 k \rfloor \) for all \( k > 2 \) and suppose, for the positive integer \( n \), that \( (k', N') \) is feasible for \( n \) and \( k' \neq 2 \). Since \( 2^{\lfloor \log_2 n \rfloor} \geq n \), \( (2, \lfloor \log_2 n \rfloor) \) is feasible for \( n \).

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

\[
\kappa(2)(a + \lfloor \log_2 k \rfloor) + \kappa(k)(b - 1) < \kappa(2)a + \kappa(k)b.
\]

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 \( \kappa(k) > \kappa(2) \lfloor \log_2 k \rfloor \) and the \((a + b)\)-vector \( w^1 \) consists of \( a \geq 0 \) 2's and \( b > 0 \) \( k \)'s, then the \((a + \lfloor \log_2 k \rfloor + b - 1)\)-vector \( w^2 \) that consists of \((a + \lfloor \log_2 k \rfloor) \) 2's and \( b - 1 \) \( k \)'s satisfies

\[
\prod_{j=1}^{a+\lfloor \log_2 k \rfloor+b-1} w_j^2 \geq \prod_{j=1}^{a+b} w_j^1.
\]
To prove the Lemma, note that since $2^{[\log_2 k]} \geq 2^{\log_2 k} = k$,

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

Beginning with $\nu^1$, 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, $N \geq [\log_2 n]$. Since $\kappa(2)N < \kappa(k)N'$, $\kappa(2) [\log_2 n] < \kappa(k)N'$.

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

**Terminology for Lemmas 1 - 4 and Proof of Theorem 4.** Given a cost function $\kappa$ and a $N$-vector of positive integers $\mathbf{e}$, we 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 defined to be **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 $\kappa$ with increasing marginal costs, and to be **more efficient** than the $\mathbf{e}'$ if (i) $\kappa[\mathbf{e}] < \kappa[\mathbf{e}']$ for all $\kappa$ 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 discrimination vector with $N$ (resp. $N^+$) coordinates. If $\mathbf{e}$ is coarser than $\mathbf{e}^+$ and $\sum_{i=1}^{N^+} e_i^{+\ast} = \sum_{i=1}^N e_i^\ast$ then, for all integers $1 \leq x \leq \min[N, N^+]$,

$$\sum_{i=N^-x+1}^N e_i^+ = \sum_{i=1}^{N^+} e_i^\ast \geq \sum_{i=N^-x+1}^N e_i^\ast$$

and, for some integer $1 \leq x \leq \min[N, N^+]$, $\sum_{i=N^-x+1}^{N^+} e_i^+ > \sum_{i=N^-x+1}^N e_i^\ast$.

**Proof.** Suppose to the contrary that there is a smallest integer $1 \leq x \leq \min[N, N^+]$ such that

$$\sum_{i=N^-x+1}^N e_i^+ < \sum_{i=N^-x+1}^N e_i^\ast.$$  

Since $x$ is smallest, $e_{N^-x+1} > e_{N^-x+1}^\ast$. Since $\sum_{i=N^-x+1}^{N^+} e_i^+ = \sum_{i=N^-x+1}^N e_i^\ast$ and both are sums of $x$ numbers, $\sum_{i=N^-x+1}^{N^+} e_i^+ < \sum_{i=N^-x+1}^N e_i^\ast$. Since $\sum_{i=1}^{N^+} e_i^\ast \geq \sum_{i=1}^N e_i^\ast$, $\sum_{i=1}^{N^+} e_i^\ast < \sum_{i=1}^N e_i^\ast$. Hence $e_{N^-x+1}^\ast > e_{N^-x+1}^\ast$, contradicting the fact that $\mathbf{e}$ is coarser than $\mathbf{e}^+$.

For the final claim note that if, for all integers $1 \leq x \leq \min[N, N^+]$, $\sum_{i=N^-x+1}^{N^+} e_i^\ast \geq \sum_{i=N^-x+1}^N e_i^\ast$ then, since $\sum_{i=1}^{N^+} e_i^\ast = \sum_{i=1}^N e_i^\ast$, we would have $\mathbf{e}^+ = \mathbf{e}$ which implies that $\mathbf{e}$ could not be coarser than $\mathbf{e}^+$. \[\square\]
Lemma 2  Given the discrimination vector \( \bar{e} = (\bar{e}_1, ..., \bar{e}_N) \), let \( \tilde{e} = (\tilde{e}_1, ..., \tilde{e}_N) \) be defined by \( \tilde{e}_i = \bar{e}_i - 1 \), \( \tilde{e}_j = \bar{e}_j + 1 \), and \( \tilde{e}_k = \bar{e}_k \) for \( k \neq i, j \). If \( \bar{e}_i \geq \bar{e}_j + 2 \) then \( \prod_{k=1}^{N} \tilde{e}_k > \prod_{k=1}^{N} \bar{e}_k \).

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

\[
\prod_{l=1}^{N} \tilde{e}_l = \left( \prod_{l \neq i, j} e_l \right) (\bar{e}_i - 1)(\bar{e}_j + 1) = \left( \prod_{l=1}^{N} \bar{e}_l \right) + \left( \prod_{l \neq i, j} e_l \right) (\bar{e}_i - e_j - 1) > \prod_{l=1}^{N} \bar{e}_l. \quad \square
\]

**Lemma 3**  Let the discrimination vector \( e \) (resp. \( e' \)) have \( N \) (resp. \( N' \)) coordinates. If \( \sum_{i=1}^{N'} e_i' \geq \sum_{i=1}^{N} e_i \) 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 discrimination vector \( \bar{e} \) with \( N' \) coordinates such that \( \sum_{i=1}^{N'} \bar{e}_i = \sum_{i=1}^{N} e_i \), \( \bar{e}_{N'+1} \geq e_{N'+1} \) for \( i \in \{1, ..., \min[N, N']\} \), and \( \bar{e} \) is weakly more efficient than \( e' \).

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

If \( e_{N'-k}^k \geq e_{N-k} \) then set \( e^{k+1} = e^k \). If \( e_{N'-k}^k < e_{N-k} \), 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) \( \sum_{i=N'-k}^{N'-k+m} e_i \geq \sum_{i=N-k}^{N-k+m} e_i \). There must be such a \( m \) since we can set \( m = k + 1 \). Then set (A) \( e_{N'-i}^k = e_{N-i}^{k+1} \) for \( i = k-m+1, ..., k \), (B) \( e_{N'-k-m}^k = \sum_{i=N'-k}^{N'-k+m} e_i^k - \sum_{i=N-k}^{N-k+m} e_i \) (or equivalently \( e_{N'-k-m}^{k+1} = e_{N-k-m}^k + \sum_{i=N'-k}^{N'-k+m} e_i^k \)) and (C) \( e_{i+1}^{k+1} = e_i^k \) for \( i = 1, ..., N' - k - 1 \) and \( i = N' - k + m + 1, ..., N' \). In both cases, \( e^{k+1} \) is a \( N' \)-vector.

To conclude the induction, we show that properties (1) - (4) are satisfied for \( k+1 \). When \( e_{N'-k}^k \geq e_{N-k} \) and therefore \( e^{k+1} = e^k \), this is immediate. So assume \( e_{N'-k}^k < e_{N-k} \). 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), \( \sum_{i=1}^{N'-k} e_i^{k+1} = \sum_{i=1}^{N'-k} e_i^k \) 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 \( e_{N'-i+1}^{k+1} \geq e_{N-i}^{k+1} \) 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), \( e_{N'-k+m}^{k+1} \geq e_{N-k+m} \). 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
Given our assumption that \( e_{N'-k} > e_{N'-k} \), and hence weakly more efficient than \( e^k \) and therefore weakly more efficient than \( e'd' \), concluding the argument for Property 4.

With the induction complete, we conclude by setting \( \bar{e} = e^\min[N, N'] \).

Lemma 4 If for the \( N \)-vector \( e \) and the \( \tilde{N} \)-vector \( \tilde{e} \), (i) each \( e_i \) and \( \tilde{e}_i \) is an integer greater than 1, (ii) \( N > \tilde{N} \), (iii) \( \tilde{e}_{N-i+1} \geq e_{N-i+1} \) for \( i = 1, \ldots, \tilde{N} \), and (iv) \( \sum_{i=1}^{N} e_i^* = \sum_{i=1}^{N} e_i^* \), then \( e \) is more efficient than \( \bar{e} \).
Proof. Define $\tilde{e} = (1, \ldots, 1, \tilde{e}_1, \ldots, \tilde{e}_N)$, where the number of 1’s equals $N - \tilde{N}$. Note that $\sum_{i=1}^{N} \tilde{e}_i^* = \sum_{i=1}^{\tilde{N}} \tilde{e}_i^* = \sum_{i=1}^{N} e_i^*$ and $\tilde{e}_{N-i+1} \geq e_{N-i+1}$ for $i = 1, \ldots, \tilde{N}$. We can therefore build a sequence of $N$-vectors $(e(j))$ with $e(1) = \tilde{e}$ and terminal vector $e(t) = e$ such that, for $j = 1, \ldots, t-1$, $e_k(j + 1) = e_k(j) + 1 \leq e_k$ for some $k = 1, \ldots, N - \tilde{N}$, $e_{k'}(j + 1) = e_{k'}(j) - 1 \geq e_{k'}$ for some $k' = N - \tilde{N} + 1, \ldots, N$, and $e_l(j + 1) = e_l(j)$ for $l \in \{1, \ldots, N\} \setminus \{k, k'\}$. Suppose $\kappa$ has strictly increasing marginal costs. Then, since

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

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

Proof of Theorem 4. Without loss of generality, we may assume that the discrimination vector $e$ of $C$ and $e'$ of $C'$ 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 $x \in \{1, \ldots, \min[N, N']\}$ and therefore, by Lemma 3, there exists a discrimination vector $\tilde{e}$ with $N'$ coordinates such that $\sum_{i=1}^{N'} (\tilde{e}_i - 1) = \sum_{i=1}^{N'} (e_i' - 1)$, $\tilde{e}_{N'-i+1} \geq e_{N-i+1}$ for $i = 1, \ldots, \min[N, N']$, and $\tilde{e}$ is weakly more efficient than $e$.

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

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

Next we show that for all $y \in \{1, \ldots, N\}$, $\sum_{i=1}^{y} e_i' \geq \sum_{i=1}^{y} e_i$. If to the contrary there is a minimal $y \in \{1, \ldots, N\}$ 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

$$\frac{\sum_{i=1}^{y} e_i'^* \leq e_y'^*} {\sum_{i=1}^{N} e_i'^*} \leq \frac{\sum_{i=1}^{y-1} e_i'^*} {\sum_{i=1}^{N} e_i'^*} < \frac{\sum_{i=1}^{y-1} e_i'^*} {\sum_{i=1}^{N} e_i'^*},$$

contradicting $e$ being coarser than $e'$.

Using this fact, we conclude that $\sum_{i=1}^{\tilde{x}} e_i' \geq \sum_{i=1}^{\tilde{x}} e_i$, which when combined with $\sum_{i=N'-\tilde{x}+1}^{N'} e_i' > \sum_{i=N-\tilde{x}+1}^{N} e_i$, 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 $e$ is more efficient than $\tilde{e}$ and hence more efficient than $e'$. 

31
Let \( c \) use \( C \) and maximally discriminate (the proof of Proposition 1 shows such a \( c \) exists) and let \( c' \) use \( C' \). 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 \( e \) is more efficient than \( e' \), \( \prod_{i=1}^{N} e_i > \prod_{i=1}^{N'} e_i' \) 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 \( C \) and \( C' \) 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 \( e \) is more efficient than \( e' \), \( \kappa[e] < \kappa[e'] \) for all \( \kappa \) with strictly increasing marginal costs. If \( \kappa \) fails to have strictly increasing marginal costs then \( (\kappa(1), ..., \kappa(N)) \) can be approximated by a \( \hat{\kappa} \) with strictly increasing marginal costs. Hence \( \kappa[e] \leq \kappa[e'] \) in all cases and so \( (C, c) \) is more efficient than \( (C', c') \). Alternatively, suppose marginal costs are strictly increasing. Then, since \( e \) is more efficient than \( e' \), \( \kappa[e] < \kappa[e'] \) and, since \( n(c) \geq n(c') \), \( (C, c) \) is again more efficient than \( (C', c') \). \( \blacksquare \)

**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 \( C(B) \) denote the set of criteria \( C_i \) such that \( B \) is contained in the top \( C_i \)-category and \( \{x, y, z\} \not\subset B \) is contained in the bottom \( C_i \)-category.\(^{22}\) 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: \( C(\{x\}), C(\{x, z\}), C(\{y\}), C(\{x, y\}), C(\{z\}), \) and \( C(\{y, z\}) \). Therefore

\[
\begin{align*}
U^{xy} &= C(\{x\}) \cup C(\{x, z\}), U^{yx} = C(\{y\}) \cup C(\{y, z\}), \\
U^{yz} &= C(\{y\}) \cup C(\{x, y\}), U^{zy} = C(\{z\}) \cup C(\{x, z\}), \\
U^{xz} &= C(\{x\}) \cup C(\{x, y\}), U^{zx} = C(\{z\}) \cup C(\{y, z\}).
\end{align*}
\]

Since \( x R y R z \), the union assumption implies

\[
(\{x\}) \cup (\{x, z\}) \cup (\{y\}) \cup (\{x, y\}) \cup (\{y, z\}) \cup (\{z\}) \cup (\{x, z\}) \cup (\{y, x\}) \cup (\{z\}) \cup (\{x, y\}) \cup (\{y, z\}) \\cup (\{z\}) \cup (\{x, z\}).
\]

So, by the subtraction assumption, \( (\{x\}) \cup (\{x, y\}) \cup (\{y\}) \cup (\{x, y\}) \cup (\{y, z\}) \cup (\{z\}) \cup (\{x, z\}) \) and hence \( U^{xz} = U^{zx} \). Therefore \( x R z \).

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

\(^{22}\) 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. \( y C_i x \)).
References


