Coarse, efficient decision-making

Michael Mandler*

Royal Holloway College, University of London

This version: March 2018

Abstract

To minimize the cost of making choice decisions, an agent should use a set of criteria each of which orders alternatives coarsely. On the domain of movies, for example, one criterion could order movies by genre categories, another by categories of directors, and so on. To arrive at choice decisions, an agent aggregates the criterion orderings, possibly by a weighted vote of the criteria. We show that as criteria become coarser (each criterion has fewer categories) decision-making costs typically fall, though an agent must then use more criteria. The most efficient option is therefore to select the binary criteria with two categories each. This result holds even when the marginal cost of using additional categories diminishes to 0. That coarse criteria are used in practice can therefore be explained as a result of optimization rather than cognitive limitations. Binary criteria also generate choice functions that maximize rational preferences, thus linking decision-making efficiency to rational choice.

JEL codes: D01

Keywords: decision-making efficiency, criteria, rationality.

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

1 Introduction

Suppose an agent wants to determine a preference ordering over a set of movies. One method, which we call *direct evaluation*, considers every pair of movies, deciding for each pair which movie is more desirable. If each pair requires a separate judgment, a preference over n movies would require $\binom{n}{2} = \frac{n(n-1)}{2}$ comparisons, a number that grows quickly as a function of n.[1] Direction evaluation appears to be how economists think of preference formation, but if each comparison is costly then this method will become prohibitively expensive even for reasonably small n.

Consider the alternative of employing a set of *criteria* where each criterion divides the domain of movies into *categories*. One criterion might partition movies into genre categories (dramas, comedies, documentaries, all others), a second might partition movies into types of directors (commercial, arty, all others), a third into actor categories (famous, not famous), and so forth. Each criterion 'orders' its categories – the genre criterion could declare that comedies are preferred to documentaries, and so forth – but the orderings need not be complete or transitive. To arrive at choice decisions, the agent aggregates the criterion orderings, for example by a weighted vote. For each pair of movies, the agent could award a score ω_i to the movie that wins the criterion i comparison and then select the movie with the greatest sum across criteria of its scores against its competitors. Although we will not assume that criterion orderings are rationally ordered, it turns out that if agents choose criteria efficiently then their choices will maximize a rational preference.

A main advantage of criteria over direct evaluation is that a criterion can discriminate within each set of alternatives that other criteria fail to rank, thus expanding the number of choice distinctions. An actor criterion, for example, can distinguish between a pair of movies x and y that are in the same genre and director categories, which will lead to a better decision from $\{x, y\}$ when the criteria are aggregated. The number of preference or choice distinctions that criteria can generate therefore equals the product of the number of criterion categories. Suppose as above that the genre criterion has 4 categories, the director criterion has 3 categories, and the actor criterion has 2 categories. If for each selection of one category from each criterion there is a movie with that combination of features then the three criteria can make $4 \times 3 \times 2 = 24$ choice distinctions.

A second advantage of criteria is equally important: they lower decision-making costs. For each pair of categories in a criterion, the agent must judge which category is superior and, as in direct evaluation, each comparison judgment is costly. But the 24 choice distinctions in our movie example require an agent to compare only a small number of pairs of categories: $\binom{4}{2} = \frac{4\times 3}{2} = 6$

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

comparisons of genre categories, $\binom{3}{2} = 3$ comparisons of director categories, and $\binom{2}{2} = 1$ comparison of actors categories, for a total of 10. In the direct evaluation method, the number of judgments of pairs of movies needed to rank 24 movies equals $\binom{24}{2} = 276$, the number of pairs of 24 movies. If the cost of comparing categories is roughly comparable to the cost of directly comparing alternatives, the gap between these numbers illustrates the cost advantage of criterion-based choice. As we will see, criteria also incur noncomparison costs but the size of the reduction in the numbers of comparisons required – 10 versus 276 – suggests that the benefits of criteria will be difficult to overturn.

We will compare direct evaluation and criterion-based choice as part of a broader analysis of which types of criteria are more efficient. Since an agent's ordering of a criterion's categories will conform with the agent's goals, the availability of an additional criterion or the addition of new category for an existing criterion will benefit the agent: more and finer criteria will allow more finely tuned decisions from more choice sets. For example, it can do no harm if the agent has access to a new scriptwriter criterion or if the drama category of the genre criterion is partitioned into thrillers and nonthrillers, and these extra distinctions may be valuable.

But adding a new criterion and making an existing criterion finer both incur costs, and not just the comparison costs already mentioned. Criteria and their categories have to be identified and criteria must be aggregated. To use a director criterion or make it finer, the agent needs to sort movies into director categories and how to weight the criterion relative to other criteria. These steps require research: checking out past efforts, discovering whose movies show at Cannes and who has won the awards. To introduce a new criterion, the agent has to find a further dimension of the alternatives and decide on its weight. Agents therefore do not and should not try to extract all of the distinctions that a criterion could in principle make or use as many criteria as possible: the benefits must be weighed against the decision-making cost.

Since increases in the number of choice distinctions and reductions in cost are both benefits, an efficient arrangement consists of a set of criteria and corresponding choice function that are undominated with respect to these two goals.

An agent that decides to create more choice distinctions seems to face a trade-off: should the agent use a small list of fine criteria or a large list of coarse criteria? If an agent wants more distinctions among movies and is initially using a director and a genre criteria, should the agent add new categories to these criteria or add a new criterion? Making existing criteria finer makes them more expensive but lets the agent save on the number of criteria.

The answer will be that under comparatively weak conditions the trade-off should be resolved in favor of coarse criteria, even when agents aim for a large number of choice distinctions. In our initial model, optimality is reached in the 'coarseness limit' where agents deploy only binary criteria, which partition alternatives into two categories. Our characterization of the optimality of binary criteria (Theorem 1) shows that binary criteria defeat the finer criteria with more than two categories even when the marginal decision-making cost of using an additional category diminishes to 0: the additional categories of fine criteria could become asymptotically free and still it will be more efficient to use the expensive categories of binary criteria.

The costliest method of all is to use a single criterion with a large number of categories which is in effect the method of direct evaluation: the only judgments the agent makes are preference comparisons between categories, which will then serve as the agent's indifference classes. Our earlier calculation of the numbers of comparison judgments – where a set of three criteria for movies was much cheaper than direct evaluation – assumed implicitly that all judgments are equally costly. The upshot of Theorem 1 is that the poor performance of direct evaluation persists even when the marginal cost of making further direct preference evaluations declines to 0.

In our characterization, the marginal cost of additional categories cannot decline *too* rapidly: if the cost of identifying or weighting a new criterion is sufficiently large then it can be superior to add further categories to existing criteria rather than to introduce a new binary criterion. Coarse criteria will still have the advantage in this case but that advantage need not reach the extreme that only binary criteria are efficient.

Binary criteria lead to choices that maximize a rational preference when criteria are aggregated in standard ways, for example through a weighted vote of criteria or when criteria form a serial dictatorship.[2] The latter result was shown in Mandler et al. (2012) but these authors failed to understand that the link between binary criteria and rational choice holds for a vast range of criterion-based choice procedures. I show here that rational choice functions arise whenever criteria are binary and decisions satisfy axioms that model and generalize a weighted vote of criteria. Since criterion-based choice is a version of multicriterion decision-making,[3] 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.

To incorporate expensive binary criteria and test the robustness of the conclusion that criteria should be coarse even when additional categories are nearly free, I will also present a more general model where criteria can have diverse values and the marginal cost of categories is restricted only in the limit as the number of categories increases. These changes allow criterion-based choice to fit

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

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

with classical utility maximization and permit binary criteria to be arbitrarily expensive. Although optimality will then no longer require criteria to be binary, they must still be coarse. Even a high-value criterion with a marginal cost of categories that descends to 0 should not become too fine: it would be more efficient to use many low-value coarse criteria despite the greater marginal cost of categories an agent would have to pay. It is certainly possible to lay out choice problems well-suited to fine criteria. If for example you care only about acting then you should devote all your research into building a criterion with a fine classification of acting types. But instances where it is optimal to let a criterion become unboundedly fine are more singular than they at first seem: if there are other attributes, even attributes with arbitrarily small values, these cases will disappear.

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 require 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 applies to the psychology of preference discovery and construction and does not have to taken preferences as given.

Coarser is better The advantages of coarse over fine criteria can be summarized concisely. As in the movie example, the maximum number of choice distinctions that can be generated given the number of categories in each criterion equals the product of the number of categories in the criteria deployed. If criterion j uses $e_j > 2$ categories and we replace it with a criterion that uses $e_j - 1$ categories (and the other criteria remain unchanged) then, to avoid a drop in the number of choice distinctions, the agent must add a new criterion. If the added criterion uses the minimum nontrivial number of categories, 2, then the difference between the number of choice distinctions 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 can produce more choice distinctions than the original set.

What is the cost of the new, coarser set of criteria? Since criteria with a single category make no choice distinctions (and are presumably costless), the total number of categories that actually discriminate among alternatives is the same in the above two sets of criteria. So if the marginal cost of using additional discriminating categories is increasing, the shift to the coarser set of criteria will reduce costs. As we will see, the presence of substantial fixed criterion costs can overturn this conclusion.

Coarser criteria thus deliver two distinct benefits: they increase the number of choice distinctions and it is plausible that they reduce costs. The argument given here for the cost reduction assumes that the marginal cost of using additional categories is increasing, but we will see that cost savings can still be achieved when the marginal cost of categories diminishes to 0.

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

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

If the choice function c decides by a simple majority vote of the criteria then choices will cycle on the pairs: $c(\{x,z\}) = \{z\}$, $c(\{y,z\}) = \{y\}$, $c(\{x,y\}) = \{x\}$. Thus c will not be used by an agent with rational preferences. But 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 choice set of alternatives, the option that lies in the greatest number of top categories will now 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.[4] I will generalize considerably in section 5.

Related work on decision-making capacity and criteria 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.[5] 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

[4] 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)).

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

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

Unlike the psychological literature, we 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.

My exclusion of ex ante preferences and emphasis on the costs of decision-making owe a great debt to Herbert Simon (e.g., Simon (1972)). But one conclusion deviates from the Simon program: paying attention to decision-making cost leads agents to rationality. This message complements Mandler (2015), where agents proceed lexicographically through criteria and it is only rational preferences that can always be the outcome of 'quick' sequences of criteria, no matter how the numbers of categories 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 where $x\, C_i\, y$ means that C_i classifies x as superior to y.[6] Criteria need not be rational: they can fail to be transitive for example. A **set of criteria**, $\mathcal{C} = \{C_1, ..., C_N\}$, consists of finitely many criteria (typically N). Criterion indices do not indicate the order in which criteria are consulted.

[6] A C_i is asymmetric if, for all x and y, $x\, C_i\, y$ implies not $y\, C_i\, x$.

To analyze the efficiency of criteria, we need measurement units for both criteria and choices. Alternatives x and y are in the same C_i-category if C_i never treats x and y differently: alternatives C_i-superior to x are also C_i-superior to y and alternatives C_i-inferior to x are also C_i-inferior to y.

Definition 1 *The set $E \subset X$ is a C_i-**category** if it consists of all alternatives that share the same upper contour sets and the same lower contour sets, that is, if for all $x \in E$*

$$y \in E \text{ if and only if } (\{z \in X : z\, C_i\, x\} = \{z \in X : z\, C_i\, y\} \text{ and } \{z \in X : x\, C_i\, z\} = \{z \in X : y\, C_i\, z\})\,.$$ [7]

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.

Criteria will typically divide X into fewer categories than the number of indifference classes of a preference or, in the language we will introduce, the number of choice classes of a choice function. For a variation on the movie example, X could be a set of vacation destinations described by a list of attributes – e.g., climate, amenities available – with each attribute ordered by a criterion. Criteria can be 'incomplete' as well as intransitive: a C_i might not rank every pair of C_i-categories.[8] Even criteria that are complete and transitive need not lead to rationally ordered choices, as seen in the Condorcet example in the introduction.

An agent's construction of criteria is a comparatively easy task when the attributes of alternatives are given exogenously. For each attribute i, there is then a set X_i of possibilities for that attribute and the entire domain of alternatives will be the product of these possibilities, $X = \prod_i X_i$. For example, a domain of newly constructed houses might allow the size and number of rooms, architectural style, heating system, etc., to be set independently. When attributes are prespecified and known, the agent does not have to bear the cost of figuring out what factors are relevant to a decision problem.

Although the categories of a criterion C_i and the criterion's ranking of those categories are formally intertwined, the categories presumably come first in the mind of a decision-maker. As with attributes, criterion categories and detailed information about which alternatives lie in which categories can sometimes be supplied exogenously. Agents still might want to avoid fine categorizations to save on the cost of ordering categories. More frequently, agents must expend effort to determine categories. If for example the attribute is the climate of vacation destinations, the agent must decide whether to distinguish between the hot and sweltering destinations or between

[7] The binary relation I defined by $x\, I\, y$ iff x and y are in the same C_i-category is an equivalence relation on X. When C_i is transitive, see Fishburn (1970) and Mandler (2009) for discussions.

[8] Since counts of the number of categories in an incomplete binary relation can be controversial, it may be helpful to assume that each C_i ranks every pair of its categories; no changes in the paper would be introduced.

destinations liable to rain and liable to pour, and then do research to find out which locations land in which categories. The x_i coordinate of $x \in X$ serves only as an index: information about the attribute value associated with x_i requires effort. For example, x_i could be the name of the director of movie x but provide no substantive information; to group x_i's into director categories the agent must find out who has won the awards, whose films show at the prestige festivals, and so on.[9]

The environment becomes more complex if the agent has to find or construct the attributes which determine the component spaces X_i. If the social understanding of the choice environment is sufficiently rich – if, after a little research, one can discover the factors that others have deemed important in comparable problems – this task need not be onerous. The job becomes harder if agents have to build novel categories on their own.

While we have considered X so far to be a product of attributes, what matters is that for each selection of categories from the criteria in use there are some alternatives that have those attribute values: if E_i is an arbitrary category of the criterion that an agent uses to order attribute i then the intersection $\bigcap_i E_i$ must be nonempty. The stronger assumption that X equals a product $\prod_i X_i$ and even the presence of attributes play no formal role outside of section 6: the agent simply selects criteria that partition X in various ways. But for criteria to be a practical way to make decisions, agents must either order prespecified attributes (the less expensive option) or invent attributes for themselves.

However attributes and criteria are assembled, the agent must convert the criteria that order attributes into choice decisions. Our agents will face a recurrent choice problem that requires decisions for a family of choice sets $\mathcal{F}$, where each $A \in \mathcal{F}$ is a nonempty subset of X. Let c be a choice function defined on $\mathcal{F}$: for every $A \in \mathcal{F}$, $c(A)$ is the agent's nonempty set of selections from A. We assume that $\mathcal{F}$ includes the two-element sets and let $x \in c(A)$ mean both that x is in $c(A)$ and that $A \in \mathcal{F}$. The choice classes of c are defined comparably to criterion categories: two alternatives are in the same choice class if c treats them as interchangeable in every sense.[10]

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

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

[9] Formally, a category is not a subset $I \subset X_i$: the category is $I \times \prod_{j \neq i} X_j$. But we may without confusion identify the category with I.

[10] 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.

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

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

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. One consequence of Definition 2 is that the choice classes form a partition of X (see the Appendix).

When a choice function is determined by criteria, selections must depend only on the distinctions the criteria make. In particular, if alternatives x and y are 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 indistinguishable, that is, in the same choice class.

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

In the movie example with three criteria, two movies that fall into the same genre, director, and acting categories must be in the same choice class when the agent's choice function uses these criteria.

The choice classes permitted by a set of criteria are merely the units of choice decisions: the agent must also adopt an aggregation method that determines a specific choice function. The leading method is to compare alternatives via a weighted vote of the criteria, one version of which is illustrated in Example 1.

Example 1 Given a set of criteria $\mathcal{C}$, for any pair $x, y \in X$, set

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

and let the weight assigned to the criterion C_i be ω_i. The sum of the weighted votes for alternative x, when $x \in A$, is given by $v(x, A) \equiv \sum_{y \in A} \sum_{i \in \{1,\ldots,N\}} \omega_i s_i(x, y)$. Then $v(x, A) = v(y, A)$ if $x, y \in A$ and x and y are contained in the same C_i-category for each $C_i \in \mathcal{C}$. Consequently the choice function c defined by

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

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

We have treated criteria and the aggregation method as fixed as the choice sets vary. They need not be. If for example an agent uses criteria to choose what to eat in restaurants – say using criteria that order meals by their meat content and by their cuisine – then the agent could on each outing vary his ranking of categories or vary the weight assigned to each criterion: one day the agent prefers fish and the next day the agent prefers meat. Although we will not further pursue this modeling possibility, it enjoys the advantage of repeatedly using the same attributes and categories, thus saving on the cost of identifying or building these tools.

3 The optimization problem

If an agent's 'true' preferences are given by the ordering of the choice classes the agent would form if he or she knew the finest possible criteria for all of the attributes then the agent will be better off if a criterion becomes finer or if the agent deploys a criterion for an additional attribute. The choice classes that result will then more closely approximate the indifference classes of the agent's true preferences and the agent will therefore make better decisions from more choice sets: when facing $\{x, y\}$ it is only when some criterion distinguishes x and y that the agent can place the items in different choice classes and choose the better option. Agents therefore seek to increase the number of their choice classes, all else being equal, a goal that we link to classical utility maximization in section 6.1.

As the criteria that order some attribute become finer – e_i becomes larger – the corresponding partitions of X into categories will normally become finer: an agent increases e_i by subdividing existing categories. For example, if initially an agent partitions movie genres into dramas and non-dramas then a finer genre criterion might subdivide dramas into thrillers and non-thrillers. The partition of X into choice classes that results will then also become finer: each choice class that previously fused all dramas will now be subdivided.[11]

Agents have a second goal of decreasing their decision-making costs. 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 section 6, costs will be determined by the number of C_i-categories: for all criteria C_i and $\widehat{C}_i$, $e(C_i) = e(\widehat{C}_i) \Rightarrow \kappa(C_i) = \kappa(\widehat{C}_i)$. We therefore write $\kappa(e)$ when convenient. The **cost of a set of criteria**

[11] We could define an increase in the number of choice classes to be beneficial only if the new partition of choice classes refines the preexisting partition: Proposition 1 and Theorem 1 would continue to hold.

$\mathcal{C} = \{C_1, ..., C_N\}$ is the sum $\kappa[\mathcal{C}] = \sum_{i=1}^{N} \kappa(C_i)$. The costs of criteria could aggregate nonadditively if criterion construction displays economies of scale, a possibility discussed in section 4.

To assess which cost functions are plausible, recall that to form a criterion an agent must find an appropriate attribute, determine the partition of X that defines the criterion's categories, and order these categories. If e is the number of categories in the criterion, the cost of partitioning might be a linear function of the number of categories that actually discriminate, $e-1$, but ordering will require an agent to decide, for any pair of categories, if they are ranked and if so which is superior. There may also be a fixed 'discovery' cost δ of finding a reasonable attribute for a criterion to order and deciding how that attribute should be weighted when the criteria are aggregated. Since $\binom{e}{2} = \frac{e(e-1)}{2}$ is the number of pairs of categories, a reasonable specification is that costs will equal a sum of linear and quadratic functions

$$\kappa(e) = \alpha(e-1) + \beta\frac{e(e-1)}{2} + \delta,$$

where $\alpha, \beta, \delta > 0$. The strict convexity of the function above provides good grounds to conclude that the marginal cost of categories will be strictly increasing in e.

This argument for increasing marginal costs does not apply, however, to binary (two-category) criteria. Criteria with one category do not require partitioning, ordering, or even selection of an attribute and should therefore have a 0 cost, though we will not use this assumption outside of section 4.2. Moreover if one-category criteria were costly no agent would use them: a one-category criterion makes no choice distinctions. The fixed discovery cost δ of identifying and weighting a suitable attribute will therefore first appear in the cost of using a binary criterion to order the attribute. Indeed if we hold the other determinants of costs fixed and increase the discovery cost δ sufficiently then the move from no criteria for an attribute to using a 2-category criterion will be more expensive that the move from 2 to 3 categories: increasing marginal costs will not kick in until we reach the third category.

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

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

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

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

The advantage of criteria is that each criterion can discriminate within every set of alternatives that the other criteria fail to rank, e.g., the genre criterion for movies will discriminate within each director category. The number of choice distinctions can then equal the product of the number of categories in the criteria as long as that product 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.*[12]

If $(\mathcal{C}, c)$ is efficient then $n(c) \geq n(c')$ must hold for any $(\mathcal{C}', c')$ that satisfies the constraints that $\mathcal{C}'$ has the same number of criteria as $\mathcal{C}$ and $e(C_i') = e(C_i)$ for all i (since then $\kappa[\mathcal{C}'] = \kappa[\mathcal{C}]$). To see when $n(c)$ reaches a maximum subject to these constraints, fix some $(\mathcal{C}, c)$. Since alternatives in different choice classes must be distinguished by at least one criterion, $n(c)$ cannot exceed the number of intersections $\bigcap_{i=1}^{N} E_i$, where each E_i is a C_i-category. The number of these intersections is in turn bounded by the product $\prod_{i=1}^{N} e(C_i)$ and criteria can always be chosen to reach this bound. Moreover the bound is necessarily achieved when X is a product of attributes and each C_i orders a distinct attribute. Our examples all enjoy this product feature. Recall that in the movie case, three criteria with 4, 3, and 2 categories can distinguish $24 = 4 \times 3 \times 2$ types of movies: each type equals the intersection of one genre category, one director category, and one actor category. Products of attributes in fact form the prototype of all cases of maximal discrimination: when criteria maximally discriminate, the alternatives can always redescribed as a product of attributes.[13] A choice function does not have to designate each intersection of C_i-categories to be a choice class – an agent might decide to ignore a criterion that, say, categorizes foods by color when the agent cares only about taste – but each intersection will form a distinct choice class when decisions are made by generic weighted-voting choice functions.

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

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

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

4 The efficiency of coarse criteria

Under relatively mild assumptions, it is more efficient to use coarser criteria that distinguish fewer categories. Maximum efficiency is then 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.

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

Example 2 To illustrate, consider choice functions with 9 choice classes. The minimal set of binary criteria that could lead to such a choice function must contain 4 criteria: the maximum number of choice classes that N binary criteria can generate is 2^N (see Proposition 1) and the minimum integer N such that 2^N is greater than or equal to 9 is 4, i.e., $\lceil \log_2 9 \rceil = 4$. The cost of using four binary criteria is therefore $4\kappa(2)$. Ternary criteria with 3 categories each would seem to be a better fit with 9 choice classes given that 9 is an exact multiple of 3. Generating a choice function with 9 choice classes requires 2 ternary criteria, which have a cost of $2\kappa(3)$. A single-category criterion makes no discriminations and should be costless ($\kappa(1) = 0$) and the binary and ternary sets each employ the same number of discriminating categories, namely 4. So if the marginal cost of categories is increasing the binary set will be strictly cheaper. Formally, increasing marginal costs imply $\kappa(3) - \kappa(2) > \kappa(2) - \kappa(1) = \kappa(2)$ and hence $2\kappa(3) > 4\kappa(2)$. Since this inequality is strict it will continue to hold if criteria incur a small discovery cost δ but it could be overturned by a large δ. If the marginal cost of categories is constant, 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 need to generate a choice function with 9 choice classes. Building choice distinctions from a nontrivial set of criteria, whether or not the set is efficient, 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 marginal costs of categories that are at least weakly increasing, the optimality of binary criteria holds even when marginal costs are decreasing.

Let $\mathcal{X}$ denote a *set* of domains, with each $X \in \mathcal{X}$ associated with its own family of choice sets. We say that $\mathcal{C}$ **has a domain** in $\mathcal{X}$ if there is a $X \in \mathcal{X}$ such that each C_i in $\mathcal{C}$ is a binary relation on X. We assume in this section that sets of criteria do not contain single-category criteria (since they make no discriminations).

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:*

1. *any efficient $\mathcal{C}$ that has a domain in $\mathcal{X}$ contains only binary criteria,*

2. $\kappa(e) > \kappa(2) \lceil \log_2 e \rceil$ *for all integers $e > 2$.*

The log cost condition is weak: the marginal cost of additional categories can fall as e increases and even fall to 0. For the reasoning behind half of Theorem 1, suppose the second statement is satisfied: costs will then fall if any criterion C_j with $e > 2$ categories is replaced by $\lceil \log_2 e \rceil$ binary criteria and, since $2^{\lceil \log_2 e \rceil} \geq e$, the binary criteria will contribute at least as much to the product of the e_i as C_j did.

If the marginal cost of criterion categories is in reality increasing, Theorem 1 indicates that the optimality of binary criteria will withstand adjustments to our framework, if not too dramatic. Suppose for example that using a criterion with $e > 1$ categories imposes a fixed discovery cost δ (as in section 3) as well as a cost $\kappa(e)$ that satisfies the log cost condition. Then, consistently with Theorem 1, it will be optimal to use only binary criteria if δ is small. But if δ is sufficiently large the log cost condition in Theorem 1 will fail and it will be efficient to use a nonbinary criterion.[14]

We have assumed that costs are additive across criteria, which is open to question since the costs of forming categories may have spillovers across criteria: an agent's identification of the genre of romantic comedies may make it easier for the agent to recognize cognate types of directors. Condition 2 in Theorem 1 continues to imply condition 1 in such cases if we read $\kappa(e)$ and $\kappa(2)$ as the additional costs of using criteria with e and 2 categories respectively given any array of other criteria in use.[15]

We turn to an application and an extension of Theorem 1.

4.1 The costs of fine criteria and direct preference evaluation

When the marginal cost of categories is increasing, Theorem 1 implies that the penalty for using fine criteria is sizable. If every criterion is constrained to have e categories, the minimum cost of a

[14] Fixing some $e > 2$, $\kappa(e) + \delta \leq (\kappa(2) + \delta) \lceil \log_2 e \rceil$ if δ is sufficiently large.

[15] Condition 1 will imply condition 2 (with the proof unchanged) if $\kappa(e)$ and $\kappa(2)$ are the costs of using a single criterion with e and 2 categories respectively and no other criteria are in use.

set of criteria that generates a choice function with n choice classes is $\lceil \log_e n \rceil \kappa(e)$ (since $\lceil \log_e n \rceil$ is the minimum integer r such that $e^r \geq n$). For approximation purposes, we ignore the difference between $\lceil \log_e n \rceil$ and $\log_e n$. The ratio of the minimum costs of a set of e-ary criteria and a set of binary criteria, when both generate n choice classes, is then

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

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

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

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

Comparisons aside, the cost of using binary criteria, $\kappa(2) \lceil \log_2 n \rceil$, increases slowly as a function of n (as does the cost of using k-ary criteria). The problem introduced at the beginning of the paper, where the cost of preference construction increases on the order of n^2 (or on the order of $n \log n$ for rational preferences) evaporates for criterion-based decision-making: the cost of a set of criteria of fixed coarseness k that makes n choice distinctions increases only on the order of $\log n$.

4.2 Arbitrary increases in coarseness

If the log cost condition holds, first-best efficiency requires criteria to be binary. If we impose the stronger assumption that the marginal cost of categories is increasing then *any* move from finer to coarser criteria brings an efficiency gain.

In a comparison of the coarseness of two sets of criteria, single-category criteria should have no impact. They are presumptively costless and if they do incur a cost they will not be used since they make no discriminations. The economically relevant number of categories of a criterion C_i is therefore given by the number of **discriminating categories** $e_i^* = e_i - 1$ (where as usual $e_i = e(C_i)$). Call the vector of positive integers $(e_1, ..., e_N)$ the **discrimination vector** of $\mathcal{C} = \{C_1, ..., C_N\}$. Following the analogy of first-order stochastic dominance, we consider $\mathcal{C}$ to be coarser than $\mathcal{C}'$ if the proportions of discriminating categories that are smaller than any given level is greater for the discrimination vector of $\mathcal{C}$ than for the discrimination vector of $\mathcal{C}'$. Given a discrimination vector $\mathbf{e} = (e_1, ..., e_N)$ with some $e_i > 1$, and given an integer $k \geq 1$, let $p_k(\mathbf{e})$ denote the proportion of $\sum_{i \in \{1,...,N\}} e_i^*$ that satisfies $e_j^* \leq k$:

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

Formally, the set of criteria $\mathcal{C}$ with the discrimination vector $\mathbf{e}$ is **coarser than** $\mathcal{C}'$ with the discrimination vector $\mathbf{e}'$ if, for each integer $k \geq 1$, $p_k(\mathbf{e}) \geq p_k(\mathbf{e}')$ and strict inequality obtains for some $k \geq 1$.[16]

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

To deal with these points, define marginal costs to be **increasing** if $\kappa(1) = 0$ and $\kappa(e+1) - \kappa(e)$ is increasing in e (on the domain $e \geq 1$) and **strictly increasing** if in addition $\kappa(e+1) - \kappa(e)$ is strictly increasing in e. Significant discovery costs (see section 3) can prevent marginal costs from being increasing. Let $\mathcal{C}$ and $\mathcal{C}'$ have the **same number of discriminating categories** if $\sum_{i=1}^{N} e_i^* = \sum_{i=1}^{N'} e_i'^*$ and **form a tight comparison** if *either* marginal costs are increasing and $\min\left[\prod_{i=1}^{N} e_i, \prod_{i=1}^{N} e_i'\right] < |X|$ *or* marginal costs are strictly increasing.

Theorem 2 ('Coarser is better') *Suppose $\mathcal{C}$ and $\mathcal{C}'$ form a tight comparison and have the same number of discriminating categories. If $\mathcal{C}$ is coarser than $\mathcal{C}'$ then $\mathcal{C}$ is more efficient than $\mathcal{C}'$.*

While Theorem 2 implies that an efficient set of criteria must be all-binary, the condition given in Theorem 1 for all-binary efficiency is much weaker.

[16] If we had defined coarseness using the distribution of the e_i rather than 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.

The heart of the proof of Theorem 2 is simple. With $\mathcal{C}$ and $\mathcal{C}'$ as given in the Theorem, we can append enough single-category criteria to $\mathcal{C}'$ to equalize the number of criteria in $\mathcal{C}$ and $\mathcal{C}'$ without affecting the cost of $\mathcal{C}'$ or the number of choice classes that $\mathcal{C}'$ can generate. Figure 1 illustrates with a finer (solid, blue) set $\mathcal{C}'$ of three criteria and a coarser (dashed, red) set $\mathcal{C}$ of six criteria, with criteria arranged so that the number of categories increases from left to right. The bottom graph adds single-category criteria to $\mathcal{C}'$ to equalize the number of criteria. Now compare the criteria in $\mathcal{C}$ and the amended version of $\mathcal{C}'$ with the greatest number of categories, then compare the criteria with the second greatest number of categories, and so on, i.e., move from right to left in the Figure. The greater coarseness of the criteria in $\mathcal{C}$ and the fact that $\mathcal{C}$ and $\mathcal{C}'$ have the same number of discriminating categories imply that at the first point where $\mathcal{C}$ and $\mathcal{C}'$ differ, it will be the criterion in $\mathcal{C}'$ – call it C'_{more} – that has more categories. Reduce the number of categories in this criterion by 1 and increase by 1 the number of categories in some other criterion C'_k in $\mathcal{C}'$ that has at least two fewer categories than C'_{more} to create a new set $\mathcal{C}''$. In the bottom graph, this change would be a move of a category leftward from a point where the height of a solid column exceeds that of a dashed column. The change reduces costs, and strictly reduces costs if marginal costs are strictly increasing. Moreover, by a calculation similar to the one in the introduction, the product of the e'_i's will increase. Hence, the number of choice classes in some c'' that uses $\mathcal{C}''$ can weakly increase and can strictly increase if $\prod_{i=1}^{N} e'_i < |X|$. Due to the tight-comparison assumption, $\mathcal{C}''$ is strictly more efficient than $\mathcal{C}'$ (though it need not be coarser). Since the number of discriminating categories is the same in $\mathcal{C}$ and $\mathcal{C}'$, there is a sequence of such steps that terminate in a set of criteria with the same cost as $\mathcal{C}'$ with no accompanying drop in the number of choice classes; in the Figure, there is just enough mass where the solids exceed the dashes to fill in the locations where the dashes exceed the solids. Since each step is an efficiency increase, $\mathcal{C}$ must be more efficient than $\mathcal{C}'$.

5 Efficiency leads to rationality

Binary criteria will under mild conditions lead to choice functions that maximize a rational preference. Subject to these conditions, rational choice is therefore a consequence of efficient decision-making. We begin with a direct and intuitive argument that shows that weighted voting, introduced in section 2, will generate a rational choice function when criteria are binary. Binariness is crucial: recall from the introduction that if three criteria rank three alternatives as voters do in the Condorcet paradox, then an equal-weight vote will cycle. Given the possibility results in voting models with dichotomous preferences (see footnote 4), one would expect that binary criteria aggregate well.

Definition 6 *Let $\mathcal{C}$ be a set of criteria and ω the criterion weights $(\omega_1, ..., \omega_N)$. Then c is a*

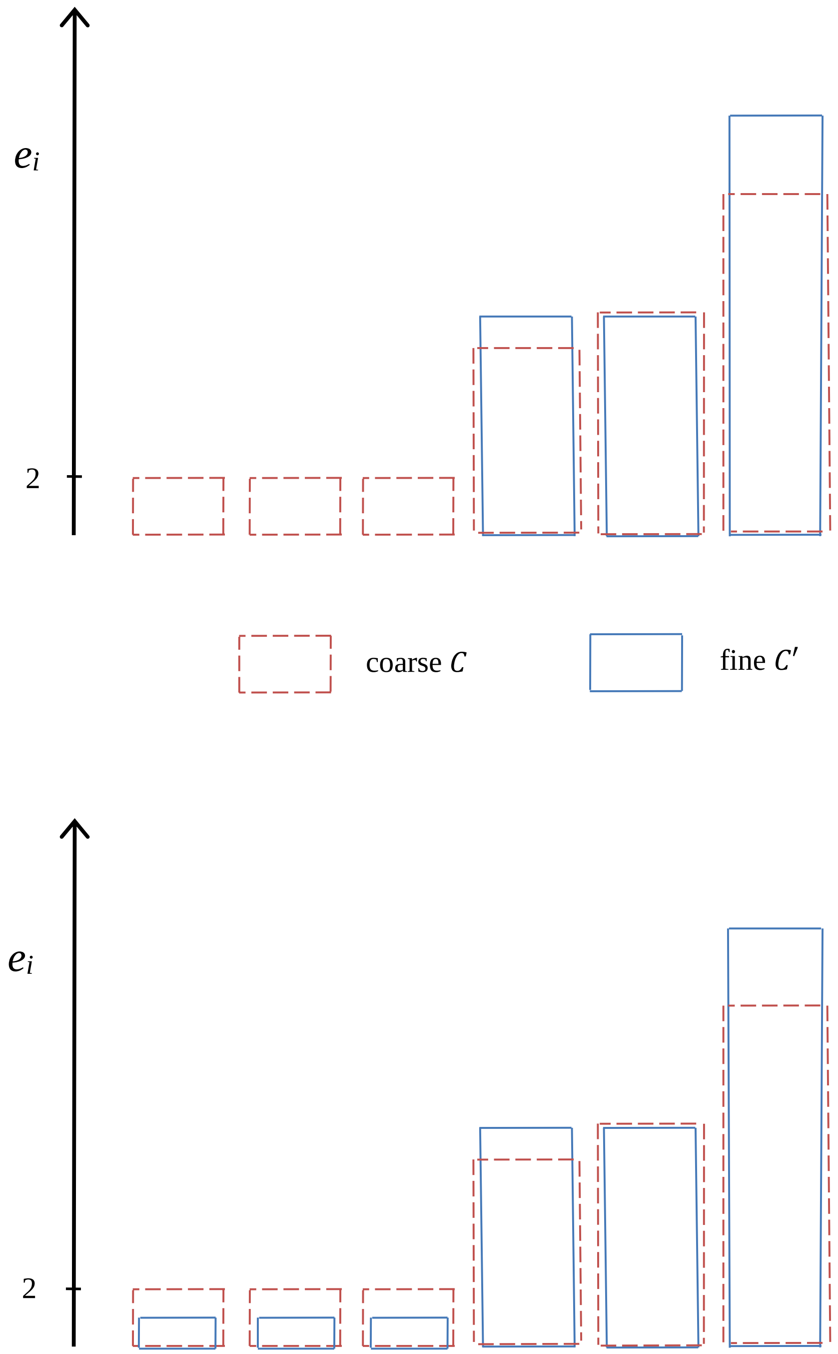

Figure 1: The move from a fine to a coarse distribution of categories

*ω-**weighted-voting choice function that uses** $\mathcal{C}$ if c uses $\mathcal{C}$ and, for all $A \in \mathcal{F}$, $c(A)$ equals*

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

when this set is nonempty.[17]

This definition encompasses Example 1 in section 2 but generalizes by remaining agnostic about what alternative is chosen when no alternative defeats all of its competitors in A in the pairwise weighted votes.

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

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

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

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

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

The reach of Proposition 3 is fairly broad; for example when criteria are binary a seemingly unrelated choice procedure, the lexicographic rule of Manzini and Mariotti (2007), leads to a weighted-voting choice function. The emphasis in Mandler et al. (2012) that lexicographic compositions

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

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

of binary criteria lead to rational choice functions therefore misleads a little: the key ingredient is that criteria are binary, not the lexicography.

For the general result, we continue to assume that any alternative that bests every other element of a choice set A is 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 then $c(A)$ equals the set of Condorcet winners.

Given a set of criteria $\mathcal{C}$, let $U^{x,y} = \{C_i \in \mathcal{C} : x\, C_i\, y\}$ denote the set of criteria that rank x over y. Define the set of criteria U to be **decisive** against the set of criteria V, written $U\, D\, V$, if, for all $a, b \in X$, $U^{a,b} = U$ and $U^{b,a} = V$ imply $a \in c(\{a, b\})$. When U is decisive against V and a is recommended over b by the criteria in U and opposed by the criteria in V then a is chosen from $\{a, b\}$.

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

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

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

It is easy to confirm that a weighted-voting choice function that uses $\mathcal{C}$ satisfies the weighting axioms with respect to $\mathcal{C}$. More complex choice functions also satisfy the voting axioms. For example, suppose $\mathcal{C}$ is partitioned into progressively less powerful 'oligarchies' $\mathcal{C}^1, \mathcal{C}^2, ..., \mathcal{C}^n$ each of which conducts a weighted vote of criteria. Let oligarchy $\mathcal{C}^1$ select all alternatives from the choice set A that $\mathcal{C}^1$ awards its highest score, as calculated in Example 1. Then present these selections to oligarchy $\mathcal{C}^2$ which will further narrow the selections using its highest score, and so on. The advantage of the weighting axioms is that it is easy to confirm that a rule like this satisfies the axioms (if a and b replicate the votes that x and y receive then a will defeat b if x defeats y, a union of victorious criteria becomes more likely to win each oligarchy vote, and a subtraction of criteria will not change the outcome of any oligarchy vote). But there are limits to what the weighting axioms can accomplish. Consider a 'liberalism' rule akin to those in Sen (1970). Suppose each oligarchy $\mathcal{C}^k$ 'owns' the right to decide between some pair of alternatives: the $\mathcal{C}^k$ scores determine the agent's choice from this pair. Then if oligarchy 1 is the determiner for x and y and oligarchy 2 is the determiner for a and b and the two oligarchies have opposite rankings for both pairs the decisiveness axiom above cannot be satisfied.

Theorem 3 *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 the log cost condition – Theorems 1 and 3 together show that efficiency in decision-making implies that choices will be rationally ordered.

6 Diverse criteria

Criteria will now vary by how costly their categories are and by the value of their distinctions. With movies, genre distinctions are presumably cheaper to discover than director distinctions and, depending on the individual, have greater or lesser value. Criteria can also be worth more if they distinguish attribute differences that are more likely to occur in the choice sets an agent encounters. If most menus of movies contain both comedies and thrillers but no documentaries then a criterion that distinguishes comedies from noncomedies is more valuable, ceteris paribus, than a criterion that distinguishes documentaries than nondocumentaries.

When criteria are diverse, it might seem that the efficient way to add choice classes is to use the low-marginal cost categories of high-value criteria and that consequently it would be efficient to let the high-value criteria with categories that enjoy diminishing marginal cost become ever finer. In fact the coarse criteria continue to prevail: even when the marginal cost of categories is diminishing and no matter how many choice distinctions an agent wants to make, the agent should use only criteria that are coarser than some fixed ceiling. This conclusion holds whether or not new criteria incur a large discovery cost, which as we have seen can undermine the efficiency of binary criteria.

The model of diverse criteria will serve as a bridge between the efficiency standard we have used so far and classical utility maximization: both will be special cases of diverse criteria.

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

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 will equal 1 when $e(C_i) = 1$. These values define a **discrimination value function** V on sets of criteria given by

$$V(\mathcal{C}) = \prod\nolimits_{C_i \in \mathcal{C}} v(C_i).$$

This function replaces the number of choice classes as the agent's discrimination objective.[18] The value of a criterion can increase without bound as its number of categories increases, a potential advantage for fine criteria (and a contrast to the utility-maximization model to come which concludes that criteria have bounded value). Coarse criteria in the present model can have negligible value since $v(C_i)$ can be near 1. Despite these biases that can favor fine criteria, the optimal decision is to select coarser criteria.

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

Both our original model and the present generalization determine the value of a set of criteria by multiplying the values of individual criteria. Recall that the advantage of criteria lies in the capacity of each criterion to distinguish within the categories of the remaining criteria. An increase in the distinctions of one criterion C_i therefore allow the *other* criteria to become more powerful since they can distinguish within more C_i-categories. This spillover of benefits can magnify when criteria have diverse values. To illustrate, consider a two-criteria world where C_1 always 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

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

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 4 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.

Our structural assumption that the value of a criterion $v(C_i)$ is bounded above by a linear function of $e(C_i)$ places an upper limit on how much value each additional category can add. If those incremental benefits had no bound then fine criteria could trump coarse criteria – and they may well do so sometimes. But the bound we use serves two goals. First, it identifies the dividing line where coarse criteria can lose their advantage. Second, it is generous enough to permit the benefits that fine criteria enjoy in the model of section 3 and is more than generous enough to encompass utility maximization, which places a ceiling on the value of the criteria that order an attribute regardless of their fineness (section 6.1). Our assumption that the value of a set of criteria V equals the product of the values of $v(C_i)$ also serves to set a common framework: both the number of choice classes and utility can be admitted as agent objectives.

We retain our notation for the cost of criteria but 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.

The set of feasible criteria needs to be well-behaved for our courseness result. Each sequence of criteria $\langle C_i^k \rangle_{k=1}^{\infty}$ for attribute i such that $e(C_i^k) = k$ for each k defines a cost function κ_i where $\kappa_i(k) = \kappa(C_i^k)$. The feasible criteria for i thus define a set of feasible cost functions for i, denoted $\{\kappa_i\}$, one function for each possible $\langle C_i^k \rangle_{k=1}^{\infty}$ and the entire set of cost function is $\bigcup_{i=1}^{\infty} \{\kappa_i\}$. The **set of feasible cost functions is compact** if every sequence of feasible cost functions has a subsequence that converges to a feasible cost function for some attribute i. This assumption ensures that the set of feasible cost functions has no 'holes' but does not rule out any cost functions.[19]

Define κ_i to **dominate fractional power functions** if there exists a $0 < a < 1$ such that, for any $\gamma > 0$, $\kappa_i(e) > \gamma e^a$ for all e sufficiently large. This condition weakens slightly the log cost condition used earlier, but the marginal cost of categories can still descend to 0 as e increases. For example, $\kappa_i(e) = \sqrt{e}$ satisfies the condition (set $a = \frac{1}{4}$).

Theorem 4 *If the feasible cost functions dominate fractional power functions and form a compact set then there is a ceiling b such that any efficient and feasible $\mathcal{C}$ contains only criteria with fewer*

[19] The distance between two cost functions κ and κ' is defined to equal the maximum over $e \geq 1$ of the distance between $\kappa(e) - \kappa'(e)$, that is, $\sup_{e \in \mathbb{N}} |\kappa(e) - \kappa'(e)|$.

than b categories.

The thrust of Theorem 4 is that, even if 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 perpetually finer.

Though binary criteria are not singled out in the assumptions above, the proof of Theorem 4 proceeds by showing that any criterion that is excessively fine can be efficiently replaced by binary criteria. Binary criteria moreover continue to enjoy a special status. If many attributes share the same values and cost functions for categories then to achieve efficiency the criteria for *these* attributes must all be binary, assuming that the log cost condition holds. When there are only a few attributes with shared values and costs the binary criteria need not dominate. But if in addition the assumptions of Theorem 2 hold for these criteria – each pair forms a tight comparison and has the same number of discriminating categories – then it is optimal to smooth the numbers of categories across criteria: the numbers of categories for these criteria should differ by at most one.

6.1 Utility-maximizing criteria

We now derive utility functions for sets of criteria from a more classical decision-theory starting point and show that utility and discrimination value are compatible goals: the utilities will satisfy our assumptions on discrimination value with room to spare.

A criterion implicitly tells an agent how to distinguish among alternatives by category. Without that information, the agent would not know the relationship between the labels of the alternatives and the categories that contain them. To show that discrimination value can order sets of criteria as utility maximizers would, we model this information explicitly.

Let X equal a product of n attributes $X = \prod_{i=1}^{n} X_i$ where n is large enough to accommodate Theorem 4 or counts only those attributes ordered by some criterion in use. Each criterion orders only one attribute and agents again choose at most one criterion per attribute. As before, for each attribute i and $e > 1$ there is a feasible criterion for i that contains e categories.

The agent will select alternatives from finite choice sets $A \subset X$ that lie in a finite family of possible choice sets $\mathcal{A}$. For each $A \in \mathcal{A}$, the attribute possibilities can be chosen independently and thus A is a product $\prod_{i=1}^{n} A_i$ where $A_i \subset X_i$. This assumption, which ensures that the criteria can be evaluated one at a time, fits for example with the 'house' example of section 2 if the agent can choose a house's architecture, heating system, etc., independently.

The agent ex ante does not know the attribute values of an alternative $x = (x_1, ..., x_n)$ or the criterion categories that contain x and therefore does not know the utility x will deliver. This uncertainty is consistent with 'knowing' the labels of the alternatives and attributes. Prior to consulting a criterion for the director attribute, say i, the agent may know the director's name x_i of a movie x but not the category or utility implications of that name. The agent also does not know ex ante which choice set in $\mathcal{A}$ he will face.

Each state s will specify all of the criterion categories that contain each $x \in X$, the utility of each x, and the $A \in \mathcal{A}$ the agent faces. The utility that each x can deliver is the random variable $u(x) = \sum_{i=1}^{n} u_i(x_i)$ which we assume is an integrable function of s. The agent seeks to maximize $\mathbb{E}u(x)$.[20]

The agent when using the set of criteria $\mathcal{C}$ discovers, for each $C_i \in \mathcal{C}$ and each $x \in X$ the agent is considering, the C_i-category that contains x and the conditional expectation of $u_i(x_i)$ given this information, which we write as $u_{C_i}[x_i]$. We assume that $u_{C_i}[x_i]$ is determined by C_i alone: the same random variable $u_{C_i}[x_i]$ obtains if a new $\mathcal{C}'$ is chosen that also contains C_i. This assumption amounts to an independence condition: the criteria in use for other attributes do not affect the distribution of $u_{C_i}[x_i]$. We do not assume however that the draws that make up any of the choice sets are independently (or identically) distributed. Since a criterion C_i with e_i categories partitions X into e_i subsets, $u_{C_i}[x_i]$ can assume at most e_i values. When $e_i = 1$ the criterion C_i makes no discriminations and then $u_{C_i}[x_i](s) = \mathbb{E}[u_i(x_i)]$ for every state s.

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

We assume for all alternatives x and attributes i that there is a nonneglible expected gain to letting a criterion C_i distinguish categories. Formally, the probability that $u_{C_i}[x_i]$ is nontrivially different from $\mathbb{E}[u_i(x_i)]$ must not be arbitrarily near 0 for all C_i with two or more categories: there is a $\varepsilon > 0$ such that $\mathbb{P}\left[\left|u_{C_i}[x_i] - \mathbb{E}[u_i(x_i)]\right| > \varepsilon\right] > \varepsilon$ for all x and all feasible C_i with $e(C_i) \geq 2$. This assumption is mild: it rules out only cases where, say, all of the director criterion's distinctions among director types have an arbitrarily small impact on utility. To satisfy the requirement it is sufficient for there to be two categories in each C_i that contain alternatives x, available in choice sets with nontrivial probability, such that the conditional expected utility of x_i given its category is not arbitrarily near to the ex ante expected utility of x_i.

To account for the agent's uncertainty regarding the choice set, let $A(s) = \{x^1(s), ..., x^{T(s)}(s)\}$ denote the choice set in $\mathcal{A}$ the agent faces at state s. Since the agent will select the available attribute level with the highest conditional expected utility, the expected utility for attribute i

[20] $\mathbb{P}(E)$ will be the probability of an event E and and $\mathbb{E}(Y)$ will be the expectation of a random variable Y.

provided by C_i when facing $A(s)$ is given by

$$U_{C_i}(s) = \mathbb{E}\left[\max\left[u_{C_i}[x_i^1(s)], ..., u_{C_i}[x_i^{T(s)}(s)]\right]\right].^{21}$$

Consequently the ex ante expected utility for attribute i provided by a criterion C_i when A is unknown is $\mathbb{E}U_{C_i}$. Since the attributes can be chosen independently, the expected utility of a set of criteria $\mathcal{C}$ is $U(\mathcal{C}) = \sum_{C_i \in \mathcal{C}} \mathbb{E}U_{C_i}$.

The model incorporates the advantage that a criterion C_i enjoys if it distinguishes between attribute values that are frequently in the same choice set. The expected utility for attribute i delivered by the chosen alternative from A can then differ substantially from the ex ante expected utility for attribute i offered by an arbitrary alternative x in A.[22]

Call a U that can arise when our assumptions on the u_i are satisfied a **utility function for criteria**. Both a utility for criteria and a discrimination value function represent complete and transitive orderings on sets of criteria.[23]

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

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

Example 3 To get a feel for actual numbers, suppose (1) the agent chooses from the choice set $A_1 \times ... \times A_n$ where each A_i consists of two x_i's, (2) for each alternative x and attribute i, $u_i(x_i)$ is uniformly distributed over the interval $[-\frac{1}{2}, \frac{1}{2}]$, and (3) for each C_i and C_i-category, the conditional distribution of $u_i(x_i)$ given the C_i-category that contains x is uniform over a subinterval of $[-\frac{1}{2}, \frac{1}{2}]$ of length $\frac{1}{e(C_i)}$. So the agent believes initially that $u_i(x_i)$ lies in $[-\frac{1}{2}, \frac{1}{2}]$ and after consulting C_i learns that $u_i(x_i)$ lies in one of $e(C_i)$ subintervals of common length. A finer criterion always provides a greater expected benefit than a coarse criterion (costs aside) as it is more likely to distinguish between the utilities of the attributes on offer. Easy calculations show the following relationship

[21] To dispel a possible confusion, the state s is fixed in this expression and serves only to identify the alternatives in the choice set $A(s)$ the agent faces.

[22] That is, if $A = \{x^1, ..., x^T\}$ then $\mathbb{E}\left[\max\left[u_{C_i}[x_i^1], ..., u_{C_i}[x_i^T]\right]\right]$ can differ substantially from $\mathbb{E}[u_i(x_i)]$.

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

between the number of C_i-categories and the expected utility of C_i:

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

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

7 Conclusion

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

If an agent uses monotone attributes – in the case of meals, say, calorie count or cost – then building the needed criteria requires only that the agent choose a cutoff that divides each attribute into two parts with more and less respectively of the attribute. Some attributes that need not be monotone – meat versus vegetarian – may also happen to divide easily into two parts. The upshot of Theorem 1 is that a binary structure such as this, although it seems crude, is the most efficient way to partition meals into a given number of choice classes.

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

A Appendix: Remaining results and proofs

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

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

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

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

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

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

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

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

Let e_i indicate $e(C_i)$, $i = 1, ..., N$, for the remainder of the proof. We now show that $|\mathcal{P}| \geq \min\left[\prod_{i=1}^N e_i, |X|\right]$. If $\prod_{i=1}^N e_i \leq |X|$ there is a partition $\mathcal{Q}$ of X with $\prod_{i=1}^N e_i$ (nonempty) cells and a bijection f from $\mathcal{Q}$ onto $\prod_{i=1}^N \{0, ..., e_i - 1\}$, which we will view as the set of mixed-radix representations (see Knuth (1997)) with bases $e_1, ..., e_N$ of the integers $0, ..., \prod_{i=1}^N e_i - 1$. If $\prod_{i=1}^N e_i > |X|$ let f be a one-to-one map from the partition $\mathcal{Q}$ of singletons of X to $\prod_{i=1}^N \{0, ..., e_i - 1\}$ that contains the points $(0, ..., 0)$, $(\min[e_1 - 1, 1], ..., \min[e_N - 1, 1])$, ..., $(\min[e_1 - 1, \overline{e} - 1], ..., \min[e_N -$

$1, \bar{e}-1]$) in its range where $\bar{e} = \max\{e_1, ..., e_N\}$. Since $e_i \le |X|$ for each i, such an f exists. Whether $\prod_{i=1}^N e_i$ is $\le$ or $>$ than $|X|$, let $f_i(Q)$ denote the $i$the coordinate of $f(Q)$. For $i = 1, ..., N$, define $\widehat{C}_i$ by $x\widehat{C}_i y$ iff $f_i(Q(x)) \ge f_i(Q(y))$, where $Q(z)$ denotes the cell of $\mathcal{Q}$ that contains z. The e_i categories of $\widehat{C}_i$ are then the nonempty sets $E \subset X$ such that, for all $x \in E$, $y \in E$ iff $f_i(Q(x)) = f_i(Q(y))$. The fact that f is onto (when $\prod_{i=1}^N e_i \le |X|$) and our selection of the image of f (when $\prod_{i=1}^N e_i > |X|$) ensures that each $\widehat{C}_i$ has e_i categories. For every $Q, Q' \in \mathcal{Q}$ with $Q \ne Q'$ there is a $i \in \{1, ..., N\}$ such that $f_i(Q) \ne f_i(Q')$. Consequently, if $x, y \in X$ are not in the same cell of $\mathcal{Q}$ then x and y are not contained in the same C_i-category. Since $(\mathcal{C}, c)$ is efficient and $\left\{\widehat{C}_1, ..., \widehat{C}_N\right\}$ has the same cost as $\mathcal{C}$, $|\mathcal{P}| \ge |\mathcal{Q}| = \min\left[\prod_{i=1}^N e_i, |X|\right]$.

Next we show that $|\mathcal{P}| \le \min\left[\prod_{i=1}^N e_i, |X|\right]$. Since $\mathcal{P}$ is a partition of X, $|\mathcal{P}| \le |X|$ and therefore $|X| < \prod_{i=1}^N e_i$ implies $|\mathcal{P}| \le \min\left[\prod_{i=1}^N e_i, |X|\right]$. So assume $\prod_{i=1}^N e_i \le |X|$. To see that $|\mathcal{P}| \le \prod_{i=1}^N e_i$, let $\{C'_1, ..., C'_{N'}\}$ be any set of criteria with $e_1, ..., e_{N'}$ categories. For any $1 \le t \le N'$, apply Definition 8 to $\{C'_1, ..., C'_t\}$ determine a partition $\mathcal{P}'_t$ of X. Then $|\mathcal{P}'_1| \le e_1$. Suppose for $t \in \{1, ..., N-1\}$ that $|\mathcal{P}'_t| \le \prod_{i=1}^t e_i$. Fix some $P_t \in \mathcal{P}_t$. Then ($P_{t+1} \in \mathcal{P}'_{t+1}$ and $P_{t+1} \subset P_t$) if and only if there is a C_{t+1}-category E_{t+1} such that $P_{t+1} = P_t \cap E_{t+1}$. Since there are at most e_{t+1} C_{t+1}-categories, P_t contains at most e_{t+1} cells of $\mathcal{P}'_{t+1}$. Hence $\left|\mathcal{P}'_{t+1}\right| \le e_{t+1}\left(\prod_{i=1}^t e_i\right)$ and we conclude that $\left|\mathcal{P}'_{N'}\right| \le \prod_{i=1}^{N'} e_i$.

Since therefore $|\mathcal{P}| = \min\left[\prod_{i=1}^N e_i, |X|\right]$, we have $n(c) = \min\left[\prod_{i=1}^N e_i, |X|\right]$. ■

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

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

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

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

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

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

Terminology for Lemmas 1 - 4 and Proof of Theorem 2. Given a cost function κ and a N-vector of positive integers $\mathbf{e}$, define $\kappa[\mathbf{e}]$ to equal $\sum_{i=1}^{N} \kappa(e_i)$. If $\mathbf{e}$ and $\mathbf{e}'$ are, respectively, N- and N'-vectors of positive integers, $\mathbf{e}$ is **weakly more efficient** than $\mathbf{e}'$ if $\prod_{i=1}^{N} e_i \geq \prod_{i=1}^{N'} e_i'$ and $\kappa[\mathbf{e}] \leq \kappa[\mathbf{e}']$ for all cost functions κ with increasing marginal costs, and is **more efficient** than $\mathbf{e}'$ if (i) $\kappa[\mathbf{e}] < \kappa[\mathbf{e}']$ for all κ with strictly increasing marginal costs, and (ii) $\prod_{i=1}^{N} e_i > \prod_{i=1}^{N'} e_i'$. The vector $\mathbf{e}$ is **coarser than** $\mathbf{e}'$ if, for some domain X, there exist $\mathcal{C}$ and $\mathcal{C}'$ such that $\mathbf{e}$ is the discrimination vector of $\mathcal{C}$, $\mathbf{e}'$ is the discrimination vector of $\mathcal{C}'$, and $\mathcal{C}$ is coarser than $\mathcal{C}'$. We will follow the convention that, for any N-vector of integers $\mathbf{e}$, coordinate labels are chosen so that e_i increases in i: $e_i \geq e_{i-1}$ for $i = 2, ..., N$.

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

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

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

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

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

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

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

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

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

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

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

To conclude the induction, we show that properties (1) - (4) are satisfied for $k+1$. When $e_{N'-k}^k \geq e_{N-k}$ and therefore $\mathbf{e}^{k+1} = \mathbf{e}^k$, 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), $\sum_{i=1}^{N'-k-1} e_i^{k+1} = \sum_{i=1}^{N'-k-1} e_i^k$ and $\sum_{i=N'-k+m+1}^{N'} e_i^{k+1} = \sum_{i=N'-k+m+1}^{N'} e_i^k$. Therefore we have

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

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

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

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

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

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

we get

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

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

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

Applying Lemma 2,

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

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

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

Lemma 4 If for the N-vector $\mathbf{e}$ and the $\widehat{N}$-vector $\widehat{\mathbf{e}}$ (i) each e_i and $\widehat{e}_i$ is an integer greater than

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

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

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

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

Proof of Theorem 2. Without loss of generality, we may assume that the discrimination vector $\mathbf{e}$ of $\mathcal{C}$ and $\mathbf{e}'$ of $\mathcal{C}'$ contain only integers greater than 1. Due to Lemma 1, $\sum_{i=N'-x+1}^{N'} e_i' \geq \sum_{i=N-x+1}^{N} e_i$ for all $x \in \{1, ..., \min[N, N']\}$ and therefore, by Lemma 3, there exists a vector of positive integers $\widehat{\mathbf{e}}$ with N' coordinates such that $\sum_{i=1}^{N'} (\widehat{e}_i - 1) = \sum_{i=1}^{N'} (e_i' - 1)$, $\widehat{e}_{N'-i+1} \geq e_{N-i+1}$ for $i = 1, ..., \min[N, N']$, and $\widehat{\mathbf{e}}$ is weakly more efficient than $\mathbf{e}$.

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

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

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

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

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

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

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

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

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

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

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

Proof of Theorem 3. Let c satisfy the weighting axioms with respect to a set of binary criteria $\mathcal{C}$ and satisfy the Condorcet rule. Let $x\,R\,y$ mean $x \in c(\{x, y\})$. R is complete. We show first that R is transitive.

Suppose $x\,R\,y\,R\,z$ and, for any $B \subset \{x, y, z\}$, let $\mathcal{C}(B)$ denote the set of criteria C_i such that B is contained in the top C_i-category and $\{x, y, z\}\backslash B$ is contained in the bottom C_i-category.[24] Since criteria are binary, if a criterion C_i does not place x, y, and z in the same C_i-category then C_i must be an element of one of the following six sets: $\mathcal{C}(\{x\})$, $\mathcal{C}(\{x, z\})$, $\mathcal{C}(\{y\})$, $\mathcal{C}(\{x, y\})$, $\mathcal{C}(\{z\})$, and $\mathcal{C}(\{y, z\})$. Therefore

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

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

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

[24] 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. $y\,C_i\,x$).

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

To see that R has finitely many equivalence classes (maximal sets E such that $a,b \in E$ implies $a\,R\,b$ and $b\,R\,a$), suppose to the contrary that $x\,R\,y$ and not $y\,R\,x$ for some x and y in the same C_i-equivalence class for each $C_i \in \mathcal{C}$. Then $\varnothing\, D\, \varnothing$ and hence $y \in c(\{x,y\})$, a contradiction.

Since R is complete and transitive and has finitely many equivalence classes, the set $M(A) = \{x \in A : x\,R\,y \text{ for all } y \in A\}$ is nonempty. By the Condorcet rule, $M(A) = c(A)$. ■

Proof of Theorem 4. We first show the preliminary that there is a a such that $\inf \kappa_j(e) > \sup \kappa_j(2) \lceil e^a \overline{v} \rceil$ for all e sufficiently large, where $\inf \kappa_j(e) = \inf\{\kappa_j(e) : \kappa_j \in \bigcup_{k\in\mathbb{N}}\{\kappa_k\}\}$ and $\sup \kappa_j(e) = \sup\{\kappa_j(e) : \kappa_j \in \bigcup_{k\in\mathbb{N}}\{\kappa_k\}\}$. Since $2x \geq x+1 \geq \lceil x \rceil$ for all $x \geq 1$ and since $\overline{v} > 1$, $2e^a\overline{v} \geq \lceil e^a \overline{v} \rceil$ for $a > 0$ and $e \geq 1$. Hence $2 \sup \kappa_j(2) e^a \overline{v} \geq \sup \kappa_j(2) \lceil e^a \overline{v} \rceil$ for $a > 0$ and $e \geq 1$. For $\kappa_i \in \bigcup_{k\in\mathbb{N}}\{\kappa_k\}$ there is, by assumption, an $0 < a < 1$ such that $\kappa_i(e) > 2\sup\kappa_j(2)\overline{v}e^a$ for all e sufficiently large. Hence $\kappa_i(e) > \sup \kappa_j(2) \lceil e^a \overline{v} \rceil$ for all e sufficiently large. To conclude this step, suppose to the contrary that for each $n \in \mathbb{N}$ there is an increasing sequence of natural numbers $\langle e_l^n \rangle$ that satisfies $\inf \kappa_j(e_l^n) \leq \sup \kappa_j(2) \left\lceil (e_l^n)^{\frac{1}{n}}\, \overline{v} \right\rceil$. The compactness assumption implies that for each n and e_l^n there is a $\kappa^{n,e_l^n} \in \bigcup_{k\in\mathbb{N}}\{\kappa_k\}$ such that $\kappa^{n,e_l^n}(e_l^n) \leq \sup \kappa_j(2) \left\lceil (e_l^n)^{\frac{1}{n}}\, \overline{v} \right\rceil$. Hence for each $n \in \mathbb{N}$ there is a $\widehat{e}^{\,n} \in \mathbb{N}$ and $\kappa^n \in \bigcup_{k\in\mathbb{N}}\{\kappa_k\}$ such that (1) $(\widehat{e}^{\,n})^{\frac{1}{n}}\, \overline{v} \to \infty$ (and therefore $\widehat{e}^{\,n} \to \infty$) and (2) $\kappa^n(\widehat{e}^{\,n}) \leq \sup \kappa_j(2) \left\lceil (\widehat{e}^{\,n})^{\frac{1}{n}}\, \overline{v} \right\rceil$. Due to compactness, there is a subsequence of natural numbers $\langle n_k \rangle$ and a $\overline{\kappa} \in \bigcup_{k\in\mathbb{N}}\{\kappa_k\}$ such that $\kappa^{n_k} \to \overline{\kappa}$. Since, for some $\overline{a} > 0$, $\overline{\kappa}(e) > 2\sup\kappa_j(2) \lceil e^{\overline{a}} \overline{v} \rceil$ for all e sufficiently large and since $\frac{1}{n_k} < \overline{a}$ for all n_k sufficiently large, there exist $\overline{e} > 0$ and $\overline{n} > 0$ such that $\overline{\kappa}(e) > 2\sup\kappa_j(2) \left\lceil e^{\frac{1}{n_k}} \overline{v} \right\rceil$ for all $e > \overline{e}$ and all $n_k > \overline{n}$. But due to (2), for each n_k,

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

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

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

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

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

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

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

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

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

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

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

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

Proof of Proposition 2. Since U is additively separable, we may add a constant to each u_i without changing the ordering represented. Set these constants so that, for $i = 1, ..., n$,

$$\min\left\{\mathbb{E}\left[u_i(x_i)\right] : x_i = x_i^k(s) \text{ for some state } s \text{ and some } k \in \{1, ..., T(s)\}\right\} = 0.$$

Define $V(\mathcal{C}) = \exp(U(\mathcal{C}))$ for any set of feasible criteria $\mathcal{C}$ and $v(C_i) = \exp(U_{C_i})$ for any feasible C_i. If $e(C_i) > 1$ then $U_{C_i} > 0$ and $v(C_i) > 1$. Setting $\underline{v} = \inf\{v(C_i) : C_i \text{ is feasible and } i \in \{1, ..., n\}\}$, we therefore have $\underline{v} \geq 1$. Given that $U_{C_i}(s)$ is bounded away from 0 for any state s and any feasible C_i with $e(C_i) > 1$, $v(C_i)$ is bounded away from 1 for the same C_i and hence $\underline{v} > 1$. Fix s and let $\mathbb{E}\left[u_i(x_i)|C_i\right]$ denote the conditional expectation of $u_i(x_i)$ given the C_i-category that contains x_i. Since

$$U_{C_i}(s) \leq \mathbb{E}\left[\sum_{k=1}^{T(s)} |u_{C_i}[x_i^k(s)]|\right] \leq \sum_{k=1}^{T(s)} \mathbb{E}\left[\mathbb{E}|u_i(x_i^k(s))|\,\Big|\,C_i\right] = \sum_{k=1}^{T(s)} \mathbb{E}\left[|u_i(x_i^k(s))|\right],$$

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[28] Sen, A., 1970, 'The impossibility of a Paretian liberal' *Journal of Political Economy* 78: 152-157.

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

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

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

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

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