COARSE, EFFICIENT DECISION-MAKING

Michael Mandler
Royal Holloway College,
University of London

Abstract
To minimize the cost of making decisions, an agent should use criteria to sort alternatives and each criterion should sort coarsely. To decide on a movie, for example, an agent could use one criterion that orders movies by genre categories, another by director categories, and so on, with a small number of categories in each case. The agent then needs to aggregate the criterion orderings, possibly by a weighted vote, to arrive at choices. As criteria become coarser (each criterion has fewer categories) decision-making costs fall, even though an agent must then use more criteria. The most efficient option is consequently to select the binary criteria with two categories each. This result holds even when the marginal cost of using additional categories diminishes to 0. The extensive use of coarse criteria in practice may therefore be a result of optimization rather than cognitive limitations. Binary criteria also generate choice functions that maximize rational preferences: decision-making efficiency implies rational choice. (JEL: D01)

1. Introduction

Suppose an agent wants to determine a preference ordering over a set of movies. One method, which I call *direct evaluation*, considers every pair of movies, deciding for each pair which movie is more desirable. If each pair needs 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] Direct evaluation is how economists usually think

The editor in charge of this paper was Juuso Välimäki.

Acknowledgments: 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, Juuso Välimäki, Peter Wakker, and two thoughtful referees for helpful comments and suggestions.

E-mail: m.mandler@rhul.ac.uk

1. If an agent knows that these judgments will form a rational preference, the number of judgments required, $n \log_2 n$, will still increase at a rate greater than n. Although the cost of constructing a preference or choice function is distinct from the complexity of representing one, Apesteguia and Ballester (2010) provide consonant results.

Journal of the European Economic Association 2020 00(0):1–39 DOI: 10.1093/jeea/jvaa002

of preference formation, but if each comparison is costly 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 superior to documentaries, and so forth—but these orderings do not have to be rational, they might, for example, be intransitive. To arrive at choice decisions, the agent needs to aggregate 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 these scores. Although criteria need not be rational, it turns out that if agents choose criteria efficiently then their choices will maximize a rational preference.

The 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 potential number of choice distinctions. An actor criterion, for example, can distinguish between a pair of movies that are in the same genre and director categories. The number of preference or choice distinctions that criteria can generate will therefore equal the product of the number of criterion categories. In our movie example, 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.

The discriminatory power of criteria will 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 with criteria each judgment is more productive. The 24 choice distinctions in the movie example require an agent to compare only a small number of pairs of categories: $\binom{4}{2} = \frac{4\times 3}{2} = 6$ pairs of genre categories, $\binom{3}{2} = 3$ pairs of director categories, and $\binom{2}{2} = 1$ pair of actors categories, for a total of 10. With direct evaluation, the number of judgments needed to order 24 types of movies equals $\binom{24}{2} = 276$, the number of pairs of 24 types. If the cost of making comparisons is roughly comparable in the two methods, the gap between these numbers illustrates the savings that criteria lead to. Criteria will incur additional costs, which I discuss presently, but the size of the reduction in the numbers of comparisons required—10 versus 276—suggests that the advantages of criteria will be difficult to overturn.

Adding a new criterion or a new category to an existing criterion will bring the benefit that an agent can make more precisely tuned decisions from more choice sets. A new scriptwriter criterion or the partitioning of the drama genre into thrillers and nonthrillers will let an agent discriminate among movies that previously were indistinguishable. At the same time, adding a new criterion or making an existing criterion finer incurs 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 do the research that sorts movies into director categories: checking out previous works, 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 aim for criteria that maximally discriminate 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 criterion and a genre criterion, should the agent add new categories to these criteria or add a new criterion? Making existing criteria finer raises their cost but lets the agent save on the number of criteria.

The answer will be that the trade-off should be resolved in favor of coarse criteria, even when agents aim for a large number of choice distinctions. In the paper's initial model, optimality is reached in the "coarseness limit" where agents deploy only binary criteria, which partition alternatives into two categories (Theorem 1). 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 as the number of categories increases: 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 amounts to direct evaluation. The only judgments the agent then makes are preference comparisons between categories, which become the agent's indifference classes. Our earlier calculation comparing a set of three criteria for movies with direct evaluation assumed implicitly that all judgments are equally costly. One upshot of Theorem 1 is that the poor performance of direct evaluation will persist even when the marginal cost of making further direct preference comparisons declines to 0.

The marginal cost of additional categories cannot decline too rapidly however. When the expense of identifying or weighting a new criterion is sufficiently large, the cost of binary criteria—the coarsest criteria that actually make discriminations—will spike. Adding further categories to existing criteria can then provide a better way to generate more choice distinctions. Coarse criteria will still have the advantage in these cases 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, Manzini,

2. In a serial dictatorship, 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.

and Mariotti (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 will show here that rational choice functions arise whenever criteria are binary and decisions satisfy axioms that generalize both weighted votes of criteria and serial dictatorships. 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 recedes.

To incorporate expensive binary criteria and test the robustness of the conclusion that criteria should be coarse even when additional categories are nearly free, an extension of the paper's initial model will let criteria have diverse values and require only that the marginal cost of categories does not fall too quickly as the number of categories increases. Classical utility maximization then qualifies as a special case of criterion-based choice. Optimality will no longer imply that criteria must be binary in this broader setting, but 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 even if they require an agent to pay a greater marginal cost for categories. It is of course possible to lay out a choice problem that is well-suited to fine criteria. If you care only about acting then you should devote all your research into building a criterion with a fine classification of actors. But instances where it is optimal to let a criterion become unboundedly fine are more singular than they at first appear: if there are other valuable attributes, even attributes with arbitrarily small utility, these cases disappear.

The efficiency of coarse criteria fits with the psychological research showing 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. But 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 learn or evolve 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 take preferences as given.

1.1. Coarser is Better

As the movie example illustrated, the maximum number of choice distinctions that can be generated given the number of categories in each criterion will equal 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

3. See Figueira, Greco, and Ehrgott (2005) and Bouyssou et al. (2006) for overviews.

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

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

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 discriminating or costly categories in the new set of criteria is the same as in the original set. So if the marginal cost of this class of categories is increasing, the shift to the coarser set of criteria will reduce costs. The presence of substantial fixed criterion costs, on the other hand, works in the opposite direction and discourages the use of large numbers of criteria.

Coarser criteria can 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 coarse criteria still deliver cost savings when the marginal cost of categories is diminishing.

1.2. The Binariness-Rationality Connection

The Condorcet paradox provides a familiar example of how rational criteria can lead to irrational choices. Each of the following three criteria judges a higher option to be superior to a lower option:

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

If the choice function c decides by a simple majority vote of the criteria then choices will cycle on the pairs: $c(\{x, y\}) = \{x\}$, $c(\{x, z\}) = \{z\}$, $c(\{y, z\}) = \{y\}$. Thus c cannot represent the decisions of 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.

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).

1.3. 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 small. In our setting, an agent who deploys a criterion has to hold in mind the category comparisons that the criterion prescribes. 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, Treisman and Gibbs (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 $e(e-1)/2$, a bound of four on the number of comparisons would place 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, I will stress the efficiency advantage of coarse criteria. Since decision-making becomes more efficient as the number of categories per criterion shrinks, the cognitive constraints that limit the number of categories in decision-making might be the outcome of optimization or adaptation. The binary criteria that use two categories are especially prevalent in everyday decision-making, and their efficiency may help to explain this fact. This conclusion aligns closely with 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 the costs of decision-making 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 per criterion 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, Manzini, and Mariotti (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

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

the present paper. Some of this research 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.

The advantages of coarse criteria in decision-making parallel the benefits of sorting information coarsely, which in turn helps explain why devices that store information use bits (Mandler 2019).

2. Choice Via Criteria

Let X be a domain of alternatives 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, for example, they can fail to be transitive. A *set of criteria* $\mathcal{C} = \{C_1, \ldots, C_N\}$ will have finitely many criteria, typically N. Criterion indices do not indicate the order in which criteria are consulted.

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 that are 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: for all $x \in E$,

$$y \in E \text{ if and only if}$$

$$\big(\{z \in X : z\, C_i\, x\} = \{z \in X : z\, C_i\, y\} \text{ and } \{z \in X : x\, C_i\, z\} = \{z \in X : y\, C_i\, z\}\big).^7$$

I 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, I require each e_i to be finite.

Criteria will typically divide X into fewer categories than the number of indifference classes of a preference or, in the language I 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—for example, 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.

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

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, readers are free to assume that each C_i ranks every pair of its categories; no changes in the paper would be introduced.

An agent's construction of criteria will be an easier task when alternatives can be described by exogenously given attributes. For each attribute i, let X_i be the set of possibilities for that attribute and let the entire domain of alternatives 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, and so forth, to be specified independently. Assuming the attributes are known, the agent does not have to bear the cost of figuring out what factors are relevant to a decision problem. The domain X need not literally equal a product of attributes to deliver this benefit: what matters is that there are alternatives in X that match every combination of criterion categories that an agent might employ.

Although the categories of a criterion C_i and the criterion's ranking of those categories are formally intertwined, the categories would normally come first in the mind of a decision-maker. As with attributes, criterion categories and knowledge of how alternatives are sorted into categories can sometimes be provided exogenously—though even in this case agents might still want to avoid fine criteria to save on the cost of ordering categories. More frequently, agents must expend effort to determine categories. If for example the climate of vacation destinations is an attribute, the agent must decide whether to distinguish between hot and sweltering destinations or between destinations liable to rain and liable to pour, and then do research to find out which locations land in which categories. For an attribute i, a $x_i \in X_i$ serves only as an index: information about the meaning of x_i can require effort. For example, x_i could be the name of the director of movie x but provide no substantive information; to partition X_i 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 that 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 attributes on their own.

The presence of attributes will play no formal role outside of Section 6: the agent simply selects criteria that partition X in various ways. For criteria to be a practical way to make decisions, however, agents must either order prespecified attributes (the less expensive option) or invent attributes for themselves.

However criteria are assembled, an agent must apply the criteria to form choice decisions. The agent will face 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. I 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 analogously to criterion categories: two

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

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

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

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. The choice classes always 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: 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 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 given in Example 1.

EXAMPLE 1. *Given a set of criteria $\mathcal{C} = (C_1, \ldots, C_N)$, 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 $v(x, A) \equiv \sum_{y \in A} \sum_{i \in \{1, \ldots, N\}} \omega_i s_i(x, y)$. If

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.

$x, y \in A$ and x and y are contained in the same C_i-category for each $C_i \in \mathcal{C}$ then $v(x, A) = v(y, A)$. 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 and have larger weights. ■

I have treated criteria and the aggregation method as fixed as the choice sets vary. They need not be. If an agent uses criteria to decide what to eat in restaurants—say using criteria that order meals by their meat content and cuisine—then the agent could on each outing vary his ranking of categories: one day the agent opts for fish and the next day the agent opts for meat. An agent could also vary the weights assigned to criteria over time. Although we will not further pursue these modeling possibilities, they enjoy 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

When an agent adopts a finer criterion or a criterion for an additional attribute, he or she can then make more precise choice distinctions. If an agent has "true" underlying preferences then a choice function with more choice classes can more closely reflect those preferences: choice classes can better approximate the agent's indifference classes and the agent can make better decisions from more choice sets.[11] When for example an agent faces $\{x, y\}$, some criterion must distinguish x and y in order for the agent to place the items in different choice classes and select the better option. Agents therefore seek to increase the number of their choice classes, all else being equal, a goal that I will 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 comedies and noncomedies then a finer genre criterion might subdivide comedies into slapstick comedies and the remainder. The partition of X into choice classes that results can then also become finer: each choice class that previously fused all comedies can now be subdivided.[12]

Agents have a second goal of decreasing their decision-making costs. Let $\kappa(C_i)$ denote the cost of criterion C_i. I assume throughout that, for any criterion C_i, $\kappa(C_i)$

11. We could define an agent's true preferences by the ordering the agent would form if he or she had access to every feasible criterion.

12. If an increase in the number of choice classes qualifies as beneficial only if the partition of X into choice classes becomes finer, results comparable to Proposition 1 and Theorem 1 would continue to hold.

≥ 0. Until Section 6, costs will be determined by the number of C_i-categories: for all criteria C_i and $\widehat{C}_i$, $e(C_i) = e(\widehat{C}_i) \Rightarrow \kappa(C_i) = \kappa(\widehat{C}_i)$. I therefore write $\kappa(e)$, defined for integers $e \geq 1$, when convenient. The *cost of a set of criteria* $\mathcal{C} = \{C_1, \dots, C_N\}$ 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 I will discuss in Section 4.

To assess which cost functions are plausible, recall that to form a criterion an agent must normally find an appropriate attribute, determine a partition that defines the criterion's categories, and order these categories. Criteria with one category do not require any of these steps. So if e is the number of categories in the criterion, the partitioning cost might be a linear function of the number of categories that actually discriminate, $e - 1$. Ordering will require an agent to decide, for any pair of categories, if they are ranked and if so, which is superior: $\binom{e}{2} = \frac{e(e-1)}{2}$ decisions must be made. 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. Putting these factors together, one reasonable specification is for the cost of $e > 1$ categories to equal

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

where $\alpha, \beta, \delta > 0$. The strict convexity of this function suggests that the marginal cost of categories will be strictly increasing in e. This function serves only as an example. Our results will not impose parametric forms.

The case for increasing marginal costs does not apply, however, to the additional costs incurred by binary (two-category) criteria. Since single-category criteria do not require partitioning, ordering, or even selection of an attribute, they should have a 0 cost (though we will not impose this assumption outside of Section 4.2). And if single-category criteria were costly no agent would use them: they make no choice distinctions. The fixed discovery cost of identifying and weighting a suitable attribute therefore forms part of the cost of using a binary criterion. A large discovery cost (δ in the previous paragraph) can therefore make a move from no criterion for an attribute to a 2-category criterion more expensive than a move from 2 to 3 categories. Increasing marginal costs would then 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 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 the sets of alternatives that the other criteria fail to rank, for example, the genre criterion for movies will discriminate within each director category. The number of choice distinctions can therefore 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 $\min[\prod_{i=1}^{N} e(C_i), |X|]$.

PROPOSITION 1. *If $(\mathcal{C}, c)$ is efficient then $(\mathcal{C}, c)$ maximally discriminates.*[13]

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 $\prod_{i=1}^{N} e(C_i)$. Criteria moreover can always be chosen to reach this bound, and 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.[14] 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.

4. The Efficiency of Coarse Criteria

Under relatively mild assumptions, efficiency will be enhanced by letting coarse criteria—criteria with fewer categories—replace fine criteria. Maximum efficiency is achieved by the binary criteria that have just two categories each, the minimum nontrivial number, and for this result the needed assumptions are milder still.

Increasing the number of categories e in a criterion seems to present a trade-off. Although the affected criterion presumably becomes more expensive to form,

13. Proofs omitted from the text are in the Appendix.

14. 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.

the creation of a given number of choice classes will require fewer criteria. Under conditions I 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 consist of* 4 *criteria: the maximum number of choice classes that N binary criteria can generate is* 2^N (*Proposition* 1) *and the minimum integer N such that 2^N is greater than or equal to* 9 *is* 4*, that is,* $\lceil \log_2 9 \rceil = 4$. *The cost of using four binary criteria is therefore* $4\kappa(2)$. *Ternary criteria with* 3 *categories each would seem to be a better fit with* 9 *choice classes given that* 9 *is an exact multiple of* 3. *Generating a choice function with* 9 *choice classes requires* 2 *ternary criteria, which have a cost of* $2\kappa(3)$. *Since a single-category criterion makes no discriminations and should be costless* ($\kappa(1) = 0$), *the binary and ternary sets each employ the same number of discriminating categories, namely* 4. *So if the marginal cost of discriminating categories is increasing the binary set will be 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. So under mild assumptions binary criteria will enjoy both a cost and a number-of-choice-classes advantage over ternary criteria.*

Both the binary and the ternary sets of criteria we have defined 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 the direct evaluation method of making a separate decision for each pair of choice classes.* ■

Although Example 2 might seem to suggest that the advantage of binary criteria relies on marginal costs of categories that are at least weakly increasing, the scope of binary optimality extends much further.

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. Since single-category criteria make no discriminations, we assume in this section that they are excluded from sets of criteria.

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 (2) is a weak assumption: 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 (2) 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 (the potential number of choice classes) as C_j did.

If the marginal cost of criterion categories is in reality increasing, the optimality of binary criteria will withstand sizable adjustments to our framework. Suppose 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. As δ increases, a point will come where total criterion costs fail to satisfy the log cost condition and it will be efficient to use a nonbinary criterion.[15] But when $\kappa(e)$ displays increasing marginal costs that point will come later: discovery costs can vary more widely without threatening binary optimality.

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.[16]

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

4.1. The Costs of Fine Criteria and Direct Preference Evaluation

Theorem 1 implies that the penalty for using fine criteria can be formidable. If every criterion is constrained to have e categories, the minimum cost of a set of criteria that generates a choice function with n choice classes is $\lceil \log_e n \rceil \kappa(e)$ (since $\lceil \log_e n \rceil$ is the minimum integer N such that $e^N \geq n$). 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 are plausible, suppose κ is linear or superlinear in e. Then, due to the slow-increasing $\log_2 e$ term in the denominator, the stated ratio will grow rapidly as a function of e: the losses incurred by using fine criteria become substantial as fineness increases. The costliest method of all lies at the

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

16. 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.

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 a contest between coarse and fine criteria. 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 $\kappa(e)/\kappa(k)\log_k e$ 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. Coarser is Better

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 make no discriminations. The economically relevant number of categories of a criterion C_i is given by the number of *discriminating categories* $e_i^* \equiv e_i - 1$, where as usual $e_i = e(C_i)$. Call the vector of positive integers $(e_1, \ldots, e_N)$ the *discrimination vector* of $\mathcal{C} = \{C_1, \ldots C_N\}$. 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 $\mathbf{e} = (e_1, \ldots, e_N)$ with some $e_i > 1$, and given an integer $k \geq 1$, let $p_k(\mathbf{e})$ denote the proportion of $\sum_{i \in \{1,\ldots,N\}} e_i^*$ consisting of terms that satisfy $e_j^* \leq k$:

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

The set of criteria $\mathcal{C}$ with the discrimination vector $\mathbf{e}$ is *coarser than* the set $\mathcal{C}'$ with the discrimination vector $\mathbf{e}'$ if, for each integer $k \geq 1$, $p_k(\mathbf{e}) \geq p_k(\mathbf{e}')$ and strict inequality obtains for some $k \geq 1$.[17]

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 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 (*strictly*) *increasing* if $\kappa(1) = 0$ and $\kappa(e+1) - \kappa(e)$ is (strictly) increasing in e. Significant discovery costs (see Section 3) can prevent marginal costs from being increasing. Let $\mathcal{C}$ and $\mathcal{C}'$ form a *tight comparison* if $\sum_{i=1}^{N} e_i^* = \sum_{i=1}^{N'} e_i'^*$ and *either* marginal costs are increasing and $\min[\prod_{i=1}^{N} e_i, \prod_{i=1}^{N} e_i'] < |X|$ *or* marginal costs are strictly increasing.

THEOREM 2 ("Coarser is better"). *If $\mathcal{C}$ and $\mathcal{C}'$ form a tight comparison and $\mathcal{C}$ is coarser than $\mathcal{C}'$ then $\mathcal{C}$ is more efficient than $\mathcal{C}'$.*

Although Theorem 2 implicitly provides sufficient conditions for an efficient set of criteria to be all-binary, the log cost condition of Theorem 1 is a much weaker requirement.

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, that is, 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 the rightmost point where the height of a solid column exceeds that of a

17. If coarseness were defined using the distribution of the e_i rather than the e_i^*, one could take any discrimination vector, append a large number of 1s to it, and thereby make it appear to be highly coarse even though its cost and discriminatory power would not have changed.

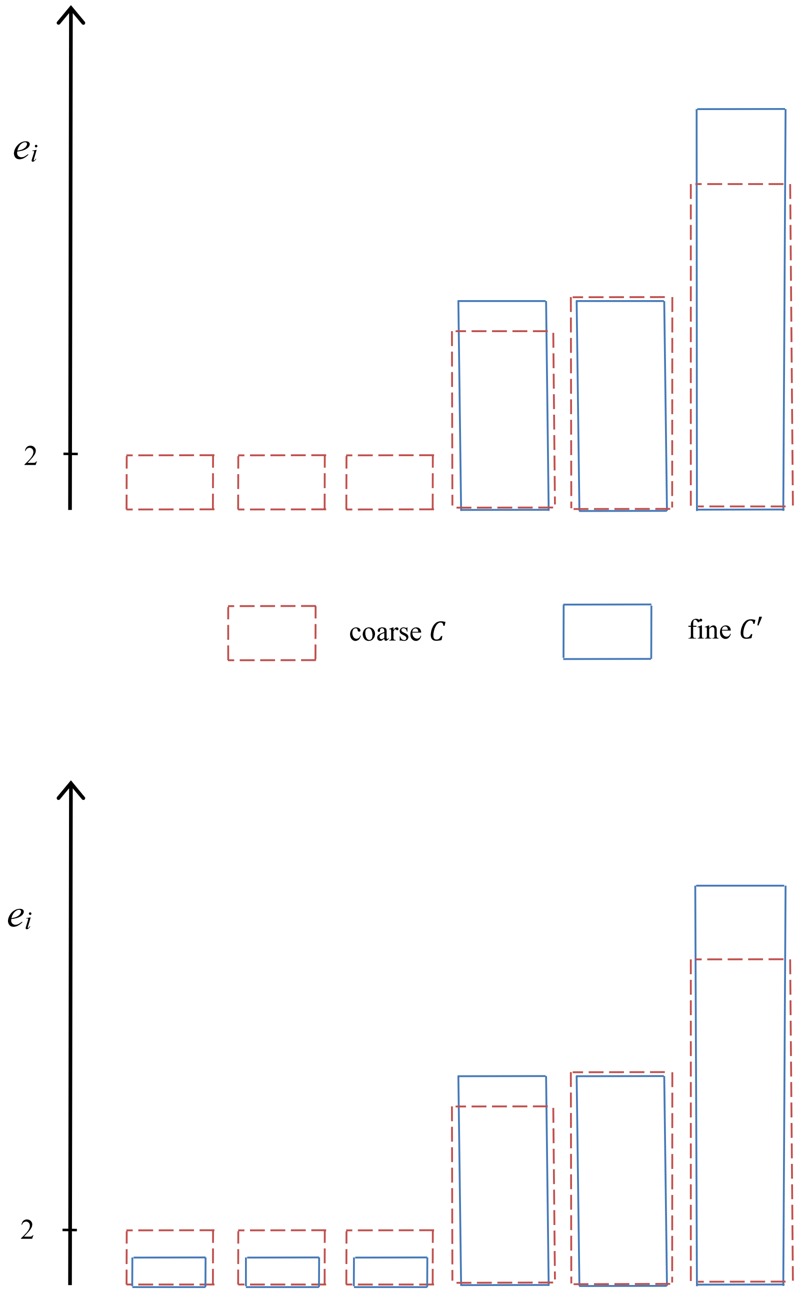

FIGURE 1. The move from a fine to a coarse distribution of categories.

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 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 the set of criteria $\mathcal{C}$; 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. I 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 binary criteria to aggregate well.

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

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

when this set is nonempty.[18]

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 pairwise weighted votes.

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

18. Recall from Section 2 that, for any pair x, y,

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

PROPOSITION 2. *For any criterion weights ω, if each criterion in the set $\mathcal{C}$ is binary then the ω-weighted-voting choice c function that uses $\mathcal{C}$ is rational.*

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

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

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

For the general result, we continue to assume that 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 entire set of these winners.

We generalize weighted voting by requiring there to be a binary relation on sets of criteria that represents which sets are more powerful or "decisive" than others. Let $U^{x,y} = \{C_i \in \mathcal{C} : x\, C_i\, y\}$ denote the set of criteria that rank x over y.

DEFINITION 7. A choice function c satisfies the *weighting axioms* with respect to the set of criteria $\mathcal{C}$ if there is a binary relation D on subsets of criteria such that, for all x, $y \in X$ with $x \neq y$ and all subsets of criteria U, V, U', V', and W in $\mathcal{C}$,

- $x \in c(\{x, y\}) \Leftrightarrow U^{x,y}\, D\, U^{y,x}$ (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 $W \subset (U \cap V) \Rightarrow (U \setminus W)\, D\, (V \setminus W)$ (subtraction).

Decisiveness implies that when one alternative is chosen over another then the criteria that back the first alternative are decisive against the criteria that back the second: the victorious set of criteria has greater "weight" than the defeated set.

Accordingly when the agent faces a different pair of alternatives backed by the same sets of criteria the alternative backed by the victorious set should again win. Union states that if two sets of disjoint criteria defeat other sets of criteria separately 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 can also satisfy the voting axioms. For example, suppose $\mathcal{C}$ is partitioned into progressively less powerful "oligarchies" $\mathcal{C}^1, \mathcal{C}^2, \ldots, \mathcal{C}^n$ 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 will win at least as many oligarchy votes, and a subtraction of criteria will not change the outcome of any oligarchy vote). There are limits of course 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 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 than a criterion that distinguishes documentaries than nondocumentaries.

When criteria are diverse, it might seem that an agent should add choice classes by refining the high-value criteria that display diminishing marginal costs for categories. 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 with fewer categories 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 goal that agents have pursued so far, maximizing the number of choice classes, and utility maximization: both will be special cases of diverse criteria.

Each criterion index i will now identify a fixed attribute of the domain, for example, genre or the type of director. 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, I 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, I assume there is an attribute i for each integer $i > 1$.

The *value of a criterion* $v(C_i)$ will incorporate both the importance of attribute i and the benefits of the number of categories in C_i. Our earlier model took the value $v(C_i)$ of a category C_i to equal its number of categories $e(C_i)$ and measured the productivity of a set of criteria by the number of choice classes the set could generate, namely, the product $\prod_{i=1}^{N} e(C_i)$. The new model retains the multiplication of the $v(C_i)$ but lets those values be modified by the significance of the distinctions that a C_i can make. Criterion values are multiplied to incorporate their interactive benefits: an increase in the distinctions of one criterion C_i can allow the other criteria to become more productive since they can distinguish within more C_i-categories. Setting the value of a set of criteria to equal the product of the $v(C_i)$ also sets a common framework: both the number of choice classes and utility can be admitted as agent objectives. The optimality-of-coarse-criteria result, Theorem 4, will therefore imply that utility maximization requires criteria to be coarse.

Formally, I assume that the values of criteria satisfy the following *admissibility* requirement: each $v(C_i)$ lies in the interval $[\underline{v}, \bar{v}e(C_i)]$ when $e(C_i) > 1$, where $1 < \underline{v} < \bar{v}$, and $v(C_i) = 1$ when $e(C_i) = 1$. Though I will not need to impose a formal assumption, for fixed i the value $v(C_i)$ would normally increase as $e(C_i)$ grows.

The $v(C_i)$ define a *discrimination value function* V on sets of criteria given by

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

This function replaces the number of choice classes as the agent's discrimination objective. The original model is the special case where $v(C_i) = e(C_i)$ for all C_i. Admissibility allows the value of a criterion to grow 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 will conclude 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 will be to select coarse criteria.

An individual criterion's value $v(C_i)$ mixes together the value per category of C_i and the number of categories in C_i. Some special cases disentangle these two effects. Suppose, for example, 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 $(1/2) < \underline{w} < \overline{w}$ 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.

The spillover of benefits, where the distinctions of one C_i make other criteria more productive, can magnify when criteria have diverse values. To illustrate, consider a two-criteria world where C_1 always has greater value per category than C_2, which in the special case of the previous paragraph would mean $w_1(e) > w_2(e')$ for all integers e, $e' > 1$. An expansion of e_2 would allow C_1 to distinguish more finely: each of the larger set of C_2-categories can be partitioned by C_1 into $e(C_1)$ distinct subsets. If, say, C_1 and C_2 are both initially binary an expansion of e_2 from 2 to 3 would allow C_1 to make its more valuable, binary distinction within 3 rather than 2 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. Theorem 4 accordingly 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 advantageous each additional category can be. If those incremental benefits had no bound then fine criteria could trump coarse criteria—and they may well do so sometimes. The bound will serve 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 (see Section 6.1).

I 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 strict inequality.

The set of feasible criteria needs to be well-behaved for our coarseness 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 on the natural numbers by setting $\kappa_i(k) = \kappa(C_i^k)$. 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 feasible cost functions 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.[19] This assumption ensures that the set of feasible cost functions has no "holes" but does not rule out any cost functions.

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 somewhat 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 $a = 1/4$).

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 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 of Theorem 4, the proof 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 assumptions comparable to those imposed by Theorem 2 hold then it will be 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

I 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. Theorem 4 therefore applies to utility-maximizing agents.

A criterion implicitly tells an agent how to distinguish among alternatives by the categories to which they belong. Without that information, the agent would not know which categories contain which alternatives. To show that discrimination value can order sets of criteria as utility maximizers would, I 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

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

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 $A \subset X$ drawn from a finite family of possible choice sets $\mathcal{A}$. For each $A \in \mathcal{A}$, the attribute possibilities can be chosen independently and therefore A will be a product $\prod_{i=1}^{n} A_i$ where $A_i \subset X_i$ for each i. This assumption ensures that the criteria that order a specific attribute can be evaluated separately from the criteria for other attributes; it will fit the house example of Section 2 if a house's architecture, heating system, and so forth, can be chosen independently.

The agent ex ante does not know the attribute levels of an alternative $x = (x_1, \ldots, x_n)$ 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 i that categorizes by director, the agent may know the director's name x_i of a movie x but not the category or utility implications of that name. An agent also might not know ex ante which choice set in $\mathcal{A}$ he or she 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 I assume is an integrable function of s. The agent seeks to maximize $\mathbb{E}u(x)$.[20] Let $A(s) = \{x^1(s), \ldots, x^{T(s)}(s)\}$ denote the choice set in $\mathcal{A}$ the agent faces at state s where $T(s) > 1$.

The agent when using the set of criteria $\mathcal{C}$ discovers, for each $C_i \in \mathcal{C}$ and $x \in X$, the C_i-category that contains x and the conditional expectation of $u_i(x_i)$ given this information, which is a random variable that I denote by $u_{C_i}[x_i]$. I 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 $u_{C_i}[x_i]$. I 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. Since in this case the C_i-category that contains x equals X, the agent does not revise his ex ante expectation of the attribute i utility of x: $u_{C_i}[x_i](s) = \mathbb{E}[u_i(x_i)]$ for every state s.

To give fine criteria an edge, I do *not* suppose, as one normally would in expected utility theory, that the values that each u_i can attain are bounded. The model thus permits fine criteria that with positive probability can inform an agent that $u_i(x_i)$ has surpassed any given threshold, no matter how large.

I assume that there is a nonnegligible expected gain to letting a criterion C_i distinguish categories. For any attribute i let there be a probability bounded away from 0 that, for each available alternative, a criterion with more than one category will report that some other available alternative lies in a nontrivially superior attribute i

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

category. Formally, we require there to be a $\varepsilon > 0$ and, for each attribute i, $A \in \mathcal{A}$, $x^k \in A$, and C_i with $e(C_i) > 1$, a $x^j \in A$ such that

$$\mathbb{P}\left[u_{C_i}\left[x_i^j\right] - u_{C_i}\left[x_i^k\right] > \varepsilon\right] > \varepsilon.$$

This assumption, which is mild, rules out only cases where a criterion will nearly always recommend the same alternative from some choice set (which would render the criterion almost useless).

Since the agent will select the available attribute levels with the highest conditional expected utility, the expected utility for attribute i provided by C_i when facing the choice set $A(s)$ is given by

$$U_{C_i}(s) = \mathbb{E}\left[\max\left[u_{C_i}\left[x_i^1(s)\right], \ldots, u_{C_i}\left[x_i^{T(s)}(s)\right]\right]\right].$$ [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 levels 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 levels that are frequently in the same choice set. The expected utility for attribute i delivered by the alternative chosen from A can then differ substantially from the ex ante expected utility for attribute i offered by an arbitrary alternative in A: $\mathbb{E}U_{C_i}$ can be significantly greater than $\mathbb{E}[u_i(x_i)]$.

A *utility maximization model* specifies, for each C_i, a $\mathbb{E}U_{C_i}$ that can obtain when our assumptions on the u_i are satisfied and thus a function U. A utility maximization model *qualifies as a discrimination value model* if there are values for criteria that satisfy the admissibility requirement such that (1) for each attribute i, $\mathbb{E}U_{C_i}$ and $v(C_i)$, seen as functions of C_i, represent the same ordering and (2) U and the discrimination value function that results from the values for criteria represent the same ordering over sets of criteria.[22]

PROPOSITION 3. *Any utility maximization model qualifies as a discrimination value model.*

Although the diverse criterion model allows $v(C_i)$ to grow without bound as $e(C_i)$ increases, the proof of Proposition 3 shows that the values for 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 $A_1 \times \cdots \times A_n$ where each A_i consists of two attribute levels* (2) *for each alternative x and attribute i, $u_i(x_i)$ is uniformly distributed over the interval*

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. Representation has its standard meaning: a function $w : Y \to \mathbb{R}$ represents the binary relation $\succsim$ if $w(y) \geq w(y') \Leftrightarrow y \succsim y'$.

$[-1/2, 1/2]$, *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* $[-1/2, 1/2]$ *of length* $1/e(C_i)$*. So the agent believes initially that* $u_i(x_i)$ *lies in* $[-1/2, 1/2]$ *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 attribute levels on offer. Easy calculations show the following relationship between the number of* C_i*-categories and the expected utility of* C_i:

$e(C_i)$	$\mathbb{E}U_{C_i}$
1	0
2	0.125
3	0.148
4	0.156
∞	0.167

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 goal of discriminating sufficiently among meals and making the fewest decisions. Theorem 1 instructs you to use binary criteria and, to satisfy maximal discrimination, you should set the binary distinction of each criterion to "cut across" the distinctions made by the other criteria. To achieve these goals, you should view the set of meals as a product of attributes, with one attribute for each criterion, and let each criterion partition its attribute levels 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 the domain of alternatives 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 alternatives 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.

Appendix: Remaining Results and Proofs

PROPOSITION A.1. *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 $xRyRz$. I show that xRz.

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

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

Next assume $B \cap \{x, z\} = \varnothing$ and $x \in c(B + x)$. Suppose $y \in B$. Then, since xRy, $y \in c(B + x)$ and so, since yRz and letting $B + x - y$ 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 suppose $y \notin B$. 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 $B - y + z$ 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 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 A.1. 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 $\hat{c}$ that uses C such that $n(\hat{c}) = |\mathcal{P}|$. For example, assign a distinct number $r(P)$ to each $P \in \mathcal{P}$, set $R(x) = r(P)$ where $P \in \mathcal{P}$ satisfies $x \in P$, and let $\hat{c}$ select from any choice set A only those alternatives $x \in A$ with the largest $R(x)$: $\hat{c}(A) = \{x \in A : R(x) \geq R(y) \text{ for all } y \in A\}$. It is easy to confirm that $\hat{c}$ uses $\mathcal{C}$, that is, if x and y are elements of the same cell of $\mathcal{P}$ then x and y are elements of the same choice class. Conversely if x and y are elements of the same choice class then, since $\{x, y\} \in \mathcal{F}$, $\hat{c}(\{x, y\}) = \{x, y\}$. Therefore $R(x) = R(y)$ and hence x and y are in the same cell of $\mathcal{P}$. Since therefore $n(\hat{c}) = |\mathcal{P}|$, we must have $n(c) = |\mathcal{P}|$.

Let e_i indicate $e(C_i)$, $i = 1, \ldots, N$, for the remainder of the proof. We now show that $|\mathcal{P}| \geq \min[\prod_{i=1}^N e_i, |X|]$. 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 $\prod_{i=1}^N \{0, \ldots, e_i - 1\}$, which we will view as the set of mixed-radix representations (see Knuth (1997)) with bases $e_1, \ldots, e_N$ of the integers $0, \ldots, \prod_{i=1}^N e_i - 1$. 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 $\prod_{i=1}^N \{0, \ldots, e_i - 1\}$ that contains in its range the points

$$(0, \ldots, 0), (\min[e_1 - 1, 1], \ldots, \min[e_N - 1, 1]), \ldots,$$
$$(\min[e_1 - 1, \bar{e} - 1], \ldots, \min[e_N - 1, \bar{e} - 1]),$$

where $\bar{e} = \max\{e_1, \ldots, e_N\}$. Since $\bar{e} \le |X|$, such a f exists. Whether $\prod_{i=1}^N e_i$ is $\le$ or $>$ than $|X|$, let $f_i(Q)$ denote the ith coordinate of $f(Q)$. For $i = 1, \ldots, N$, 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 $f_i(Q(x)) = f_i(Q(y))$. Due to the fact that f is onto (when $\prod_{i=1}^N e_i \le |X|$) and our selection of the image of f (when $\prod_{i=1}^N e_i > |X|$), each $\widehat{C}_i$ has e_i categories. For every $Q, Q' \in \mathcal{Q}$ with $Q \neq Q'$ there is a $i \in \{1, \ldots, N\}$ such that $f_i(Q) \neq f_i(Q')$. Consequently, $x, y \in X$ are in the same cell of $\mathcal{Q}$ iff x and y are contained in the same $\widehat{C}_i$-category for all i. Hence, as in the preceding paragraph, there is a $\hat{c}$ that uses $\widehat{\mathcal{C}} = \{\widehat{C}_1, \ldots, \widehat{C}_N\}$ such that $n(\hat{c}) = |\mathcal{Q}|$. Since $(\mathcal{C}, c)$ is efficient, $n(c) = |\mathcal{P}|$, and $\widehat{\mathcal{C}}$ has the same cost as $\mathcal{C}$, $|\mathcal{P}| \ge |\mathcal{Q}| = \min[\prod_{i=1}^N e_i, |X|]$.

Next we show that $|\mathcal{P}| \le \min[\prod_{i=1}^N e_i, |X|]$. Since $\mathcal{P}$ is a partition of X, $|\mathcal{P}| \le |X|$ and therefore $|X| < \prod_{i=1}^N e_i$ implies $|\mathcal{P}| \le \min[\prod_{i=1}^N e_i, |X|]$. So assume $\prod_{i=1}^N e_i \le |X|$. To show that $|\mathcal{P}| \le \prod_{i=1}^N e_i$, for any $1 \le t \le N$, apply Definition A.1 to $\{C_1, \ldots, C_t\}$ to determine a partition $\mathcal{P}_t$ of X. Then $|\mathcal{P}_1| \le e_1$. Suppose for $t \in \{1, \ldots, N-1\}$ 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 $|\mathcal{P}_{t+1}| \le e_{t+1}(\prod_{i=1}^t e_i)$ and we conclude that $|\mathcal{P}| = |\mathcal{P}_N| \le \prod_{i=1}^N e_i$.

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

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. Let $\mathcal{C}'$ be a set of $N - 1 + \lceil \log_2 e \rceil$ criteria such that, for $j \in \{1, \ldots, N\} \setminus \{i\}$, $e(C'_j) = e(C_j)$ 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 $\kappa[\mathcal{C}'] = \sum_{j \in \{1, \ldots, N\} \setminus \{i\}} \kappa(C_j) + \kappa(2)\lceil \log_2 e \rceil$,

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

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

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

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

Conversely, assume that any efficient $\mathcal{C}$ on a domain in $\mathcal{X}$ contains only binary criteria and suppose that, for some $e > 2$, $\kappa(e) \leq \kappa(2)\lceil \log_2 e \rceil$. Set $X \in \mathcal{X}$ so that $|X| = e$.

We show first that there exists an efficient $\widehat{\mathcal{C}}$. Define the family $\mathbb{S}$ of sets of criteria by $\bar{\mathcal{C}} \in \mathbb{S}$ iff (i) there is a $\bar{c}$ that uses $\bar{\mathcal{C}}$ such that $n(\bar{c}) = e$ and (ii) for each $2 \leq h \leq e$, $\bar{\mathcal{C}}$ contains at most $\lceil \log_h e \rceil$ criteria with h categories. Since $|X| = e$, each criterion C_i must satisfy $e(C_i) \leq e$ and hence $\mathbb{S}$ is finite. If $\mathcal{C}''$ is a set of criteria outside of $\mathbb{S}$ then, for some $2 \leq h' \leq e$, $\mathcal{C}''$ contains more than $\lceil \log_{h'} e \rceil$ criteria with h' categories. Letting $\mathcal{C}'''$ have $\lceil \log_{h'} e \rceil$ criteria, each with h' categories, we have $\kappa[\mathcal{C}'''] \leq \kappa[\mathcal{C}'']$. Since $\prod_{j=1}^{N} e'''_j \geq e$, the proof of Proposition 1 implies that there is a $(\tilde{\mathcal{C}}, \tilde{c})$ such that $\tilde{\mathcal{C}}$ has $\lceil \log_{h'} e \rceil$ criteria, each with h' categories, $\kappa[\tilde{\mathcal{C}}] = \kappa[\mathcal{C}''']$, and $n(\tilde{c}) = e$. Thus, for each $\mathcal{C}''$ not in $\mathbb{S}$, there is a $\tilde{\mathcal{C}} \in \mathbb{S}$ and a $\tilde{c}$ that uses $\tilde{\mathcal{C}}$ such that $\kappa[\tilde{\mathcal{C}}] \leq \kappa[\mathcal{C}'']$ and $n(\tilde{c}) = e$. Since $\mathbb{S}$ is finite, there exists an efficient $\widehat{\mathcal{C}}$ in $\mathbb{S}$.

By assumption, $\widehat{\mathcal{C}}$ contains only binary criteria and, given the proof of Proposition 1, $|\widehat{\mathcal{C}}| = \lceil \log_2 e \rceil$. So $\kappa[\widehat{\mathcal{C}}] = \kappa(2)\lceil \log_2 e \rceil$. But any $(\mathcal{C}', c')$ such that $\mathcal{C}'$ consists of a single criterion with e categories and $n(c') = e$ satisfies $\kappa[\mathcal{C}'] = \kappa(e) \leq \kappa(2)\lceil \log_2 e \rceil$. The set $\mathcal{C}'$ is therefore efficient, giving a contradiction.

Terminology for Lemmas A.1–A.4 and Proof of Theorem 2

Given a cost function κ and a vector of positive integers $\mathbf{e} = (e_1, \ldots, e_N)$, 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 each integer $k \geq 1$, $p_k(\mathbf{e}) \geq p_k(\mathbf{e}')$ and strict inequality obtains for some $k \geq 1$. 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: $e_{i+1} \geq e_i$ for $i = 1, \ldots, N-1$.

LEMMA A.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, contrary to the first claim, 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 $\sum_{i=1}^{N^{+}} e_i^{+*} = \sum_{i=1}^{N} e_i^{*}$,

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

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 $1 \leq x \leq \min[N, N^{+}]$, $\sum_{i=N^{+}-x+1}^{N^{+}} e_i^{+} = \sum_{i=N-x+1}^{N} e_i$ then, since $\sum_{i=1}^{N^{+}} e_i^{+*} = \sum_{i=1}^{N} e_i^{*}$, we would have $\mathbf{e}^{+} = \mathbf{e}$, which implies that $\mathbf{e}$ could not be coarser than $\mathbf{e}^{+}$. □

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

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

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

□

LEMMA A.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* $x \in \{1, \ldots, \min[N, N']\}$ *and* $\sum_{i=1}^{N'} e_i'^{*} = \sum_{i=1}^{N} e_i^{*}$, *then there exists a vector of positive integers* $\hat{\mathbf{e}}$ *with* N' *coordinates such that* $\sum_{i=1}^{N'} \hat{e}_i^{*} = \sum_{i=1}^{N'} e_i'^{*}$, $\hat{e}_{N'-i+1} \geq e_{N-i+1}$ *for* $i \in \{1, \ldots, \min[N, N']\}$, *and* $\hat{\mathbf{e}}$ *is weakly more efficient than* $\mathbf{e}'$.

Proof. To proceed by induction, set $\mathbf{e}^1 = \mathbf{e}'$. Given some $k \in \{1, \ldots, \min[N, N'] - 1\}$, 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 $i = 1, \ldots, k$,

(3) $\sum_{i=N'-x+1}^{N'} e_i^k \geq \sum_{i=N-x+1}^{N} e_i$ for all $x \in \{1, \ldots, \min[N, N']\}$, 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 $\mathbf{e}^{k+1} = \mathbf{e}^k$.

Assuming now that $e_{N'-k}^k \leq e_{N-k}$, 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) $\sum_{i=N'-k}^{N'-k+m} e_i^k \geq \sum_{i=N-k}^{N-k+m} e_i$. To see that there must be such a m, observe that (ii) holds when $m = k$ (set $x = k + 1$) and, since (i) holds when $m = 1$, if we progressively reduce m from k there must be a first point where (i) holds. Now set

(A) $e_{N'-i}^{k+1} = e_{N-i}$ for $i = k - m + 1, \ldots, k$,

(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 $i = 1, \ldots, N' - k - 1$ and $i = N' - k + m + 1, \ldots, N'$.

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 henceforth that $e_{N'-k}^k \leq e_{N-k}$.

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),

$$\sum_{i=1}^{N'-k-1} e_i^{k+1} = \sum_{i=1}^{N'-k-1} e_i^k \quad \text{and} \quad \sum_{i=N'-k+m+1}^{N'} e_i^{k+1} = \sum_{i=N'-k+m+1}^{N'} e_i^k.$$

Therefore,

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

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 $i = k - m + 2, \ldots, k + 1$ by (A), for $i = k - m + 1$ by (ii) and (B), and for $i = 1, \ldots, k - m$ by (2) and (C).

Property 3. Due to (C) and (3),

$$\sum_{i=N'-x+1}^{N'} e_i^{k+1} \geq \sum_{i=N-x+1}^{N} e_i \tag{A.2}$$

for $x = 1, \ldots, k - m$. Due to (B) and (ii), $e^{k+1}_{N'-k+m} \geq e_{N-k+m}$. So, given that (A.2) holds for $x = 1, \ldots, k - m$, (A.2) holds for $x = k - m + 1$. Similarly, due to (A) and given that (A.2) holds for $x = 1, \ldots, k - m + 1$, (A.2) holds for $x = k - m + 2, \ldots, k + 1$. Finally, due to (C) and (A.1), $\sum_{i=j}^{N'} e^{k+1}_i = \sum_{i=j}^{N'} e^k_i$ holds for $j = 1, \ldots, N' - k - 1$. Hence $\sum_{i=N'-x+1}^{N'} e^{k+1}_i = \sum_{i=N'-x+1}^{N'} e^k_i$ for $x > k + 1$ and therefore (3) implies that (A.2) 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 $N' - k, \ldots, N' - k + m$, and beginning with $\mathbf{e}(0) = (e^k_{N'-k}, \ldots, e^k_{N'-k+m})$. If $e_{N'-k}(j) \leq e_{N-k}$ and there exists a coordinate $l \in \{N' - k + 1, \ldots, N' - k + m\}$ with $e_l(j) > e^{k+1}_l$ 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 $\mathbf{e}(j)$. Given our assumption that $e_{N-k} > e^k_{N'-k}$ and (A),

$$e^{k+1}_{N'-k} > e^k_{N'-k}.$$

Due to (2) and (A),

$$e^{k+1}_i \leq e^k_i \text{ for } i = N' - k + 1, \ldots, N' - k + m - 1.$$

Combining (A) and (B) gives

$$\sum_{i=N'-k}^{N'-k+m} e^{k+1}_i = \sum_{i=N'-k}^{N'-k+m} e^k_i,$$

whereas combining (A) and (i) gives $\sum_{i=N'-k}^{N'-k+m-1} e^k_i < \sum_{i=N'-k}^{N'-k+m-1} e^{k+1}_i$. Hence

$$e^{k+1}_{N'-k+m} \leq e^k_{N'-k+m}.$$

The preceding four indented conditions imply that, for some positive integer t, $\mathbf{e}(t) = (e^{k+1}_{N'-k}, \ldots, e^{k+1}_{N'-k+m})$, at which point $\langle \mathbf{e}(j) \rangle$ terminates.

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

Applying Lemma 2,

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

for $j = 0, \ldots, t-1$. 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 $\hat{\mathbf{e}} = \mathbf{e}^{\min[N,N']}$. □

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

Proof. Define $\tilde{\mathbf{e}} = (1, \ldots, 1, \hat{e}_1, \ldots, \hat{e}_{\widehat{N}})$, where the number of 1s equals $N - \widehat{N}$. Note that $\sum_{i=1}^{N} \tilde{e}_i^* = \sum_{i=1}^{\widehat{N}} \hat{e}_i^* = \sum_{i=1}^{N} e_i^*$ and $\tilde{e}_{N-i+1} \geq e_{N-i+1}$ for $i = 1, \ldots, \widehat{N}$. We can therefore build a sequence of N-vectors $\langle \mathbf{e}(j) \rangle$ with $\mathbf{e}(1) = \tilde{\mathbf{e}}$ and terminal vector $\mathbf{e}(t) = \mathbf{e}$ such that, for $j = 1, \ldots, t-1$, $e_k(j+1) = e_k(j) + 1 \leq e_k$ for some $k \in \{1, \ldots, N - \widehat{N}\}$, $e_{k'}(j+1) = e_{k'}(j) - 1 \geq e_{k'}$ for some $k' \in \{N - \widehat{N} + 1, \ldots, N\}$, and $e_l(j+1) = e_l(j)$ for $l \in \{1, \ldots, N\} \setminus \{k, k'\}$. 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 $\tilde{\mathbf{e}}$ equal the vector $\hat{\mathbf{e}}$ and the remaining coordinates equal 1, $\mathbf{e}(1)$ is weakly more efficient than $\hat{\mathbf{e}}$. Hence $\mathbf{e}$ is more efficient than $\hat{\mathbf{e}}$. □

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 $x \in \{1, \ldots, \min[N, N']\}$ and therefore, by Lemma 3, there exists a vector of positive integers $\hat{\mathbf{e}}$ with N' coordinates such that

$$\sum_{i=1}^{N'} (\hat{e}_i - 1) = \sum_{i=1}^{N'} (e_i' - 1),$$

$\hat{e}_{N'-i+1} \geq e_{N-i+1}$ for $i = 1, \ldots, \min[N, N']$, and $\hat{\mathbf{e}}$ is weakly more efficient than $\mathbf{e}$.

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

$$\sum_{i=1}^{N'} (\hat{e}_i - 1) = \sum_{i=1}^{N'} (e_i' - 1),$$

$\sum_{i=1}^{N'} (e_i' - 1) > \sum_{i=1}^{N} (e_i - 1)$, 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 $x \in \{1, \ldots, N\}$ implies $\mathbf{e} = \mathbf{e}'$, Lemma 1 implies there is a $\hat{x} \in \{1, \ldots, N\}$ such that $\sum_{i=N'-\hat{x}+1}^{N'} e'_i > \sum_{i=N-\hat{x}+1}^{N} e_i$.

Next we show that for all $y \in \{1, \ldots, N\}$, $\sum_{i=1}^{y} e'_i \geq \sum_{i=1}^{y} e_i$. If to the contrary there is a minimal $y \in \{1, \ldots, N\}$ such that $\sum_{i=1}^{y} e'_i < \sum_{i=1}^{y} e_i$ then $\sum_{i=1}^{y-1} e'_i \geq \sum_{i=1}^{y-1} e_i$ and $e'_y \leq e_y$. 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 $\sum_{i=1}^{\hat{x}} e'_i \geq \sum_{i=1}^{\hat{x}} e_i$, which when combined with $\sum_{i=N'-\hat{x}+1}^{N'} e'_i > \sum_{i=N-\hat{x}+1}^{N} e_i$, 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 $\hat{\mathbf{e}}$ 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 $n(c) = \min[\prod_{i=1}^{N} e_i, |X|]$ and $n(c') \leq \min[\prod_{i=1}^{N'} e'_i, |X|]$. Given that $\mathbf{e}$ is more efficient than $\mathbf{e}'$, $\prod_{i=1}^{N} e_i > \prod_{i=1}^{N'} e'_i$ and therefore $\min[\prod_{i=1}^{N} e_i, |X|] \geq \min[\prod_{i=1}^{N'} e'_i, |X|]$. Hence $n(c) \geq n(c')$.

Since $\mathcal{C}$ and $\mathcal{C}'$ form a tight comparison, either $\min[\prod_{i=1}^{N} e_i, \prod_{i=1}^{N} e'_i] < |X|$ 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 $\langle \hat{\kappa}_n \rangle$ where each $\hat{\kappa}_n$ has strictly increasing marginal costs and $(\hat{\kappa}_n(1), \ldots, \hat{\kappa}_n(N)) \to (\kappa(1), \ldots, \kappa(N))$. Hence $\kappa[\mathbf{e}] \leq \kappa[\mathbf{e}']$ whether marginal costs increase strictly or weakly and so $(\mathcal{C}, c)$ is more efficient than $(\mathcal{C}', c')$. In the second case, where marginal costs are strictly increasing, $\kappa[\mathbf{e}] < \kappa[\mathbf{e}']$ since $\mathbf{e}$ is more efficient than $\mathbf{e}'$. Given that $n(c) \geq n(c')$, $(\mathcal{C}, c)$ 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. Define the binary relation R on X by $x R y$ iff $x \in c(\{x, y\})$. Since $\mathcal{F}$ contains the two-element subsets of X, R is complete.

To show 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

$\{x, y, z\} \setminus B$ is contained in the bottom C_i-category.[23] 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 $\mathcal{C}(\{y, z\})$. Then

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

Since $x\,R\,y\,R\,z$, the decisiveness assumption implies $U^{xy}\,D\,U^{yx}$ and $U^{yz}\,D\,U^{zy}$. So, by the union assumption,

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

Hence, by the subtraction assumption, $(\mathcal{C}(\{x\}) \cup \mathcal{C}(\{x, y\}))D\,(\mathcal{C}(\{y, z\}) \cup \mathcal{C}(\{z\}))$, that is, $U^{x,z}\,D\,U^{z,x}$. Therefore, $x\,R\,z$.

Next I show that R generates finitely many equivalence classes which in the present setting we define to be maximal sets $E \subset X$ such that $a, b \in E$ and $a \neq b$ imply $a\,R\,b$ and $b\,R\,a$. Let $x \neq y$ be in the same C_i-category for each $C_i \in \mathcal{C}$. Since $\{x, y\} \in \mathcal{F}$, we may suppose without loss of generality that $x \in c(\{x, y\})$ and therefore $x\,R\,y$. Then by decisiveness $\varnothing\,D\,\varnothing$ and hence $y \in c(\{x, y\})$ and $y\,R\,x$. Consequently any $\bigcap_{i=1}^{n} E_i$ such that each E_i is a C_i-category is contained in a R equivalence class. Since the $\bigcap_{i=1}^{n} E_i$ are finite in number and partition X and the R equivalence classes also partition X, there can be only finitely many R equivalence classes.

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

Proof of Theorem 4

We first show the preliminary that there is a $a > 0$ such that $\inf \kappa(e) > \sup \kappa(2)\lceil e^a \bar{v}\rceil$ for all e sufficiently large, where $\inf \kappa(e) = \inf\{\kappa_j(e) : \kappa_j \in \bigcup_{k\in\mathbb{N}}\{\kappa_k\}\}$ and $\sup \kappa(e) = \sup\{\kappa_j(e) : \kappa_j \in \bigcup_{k\in\mathbb{N}}\{\kappa_k\}\}$. Since $2x \geq x + 1 \geq \lceil x\rceil$ for all $x \geq 1$ and since $\bar{v} > 1, 2e^a\bar{v} \geq \lceil e^a\bar{v}\rceil$ for $a > 0$ and $e \geq 1$. Hence $2\sup\kappa(2)e^a\bar{v} \geq \sup\kappa(2)\lceil e^a\bar{v}\rceil$ 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 $\kappa_i(e) > 2\sup\kappa(2)\bar{v}e^a$ for all e sufficiently large. Hence $\kappa_i(e) > \sup\kappa(2)\lceil e^a\bar{v}\rceil$ for all e sufficiently large. (Similarly, since for $\kappa_i \in \bigcup_{k\in\mathbb{N}}\{\kappa_k\}$ there is a $0 < \bar{a} < 1$ such that $\kappa_i(e) > 4\sup\kappa(2)\bar{v}e^{\bar{a}}$ for all large e, $\kappa_i(e) > 2\sup\kappa(2)\lceil e^{\bar{a}}\bar{v}\rceil$ for all large e, an observation we use later.) 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

23. 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. $y\,C_i\,x$).

$\inf \kappa(e_l^n) \le \sup \kappa(2)\lceil (e_l^n)^{\frac{1}{n}} \bar{v}\rceil$. 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 $\kappa^{n,e_l^n}(e_l^n) \le \sup \kappa(2)\lceil (e_l^n)^{\frac{1}{n}} \bar{v}\rceil$. Hence there are sequences $\langle \hat{e}^n\rangle$ and $\langle \kappa^n\rangle$, where each $\hat{e}^{\,n} \in \mathbb{N}$ and $\kappa^n \in \bigcup_{k\in\mathbb{N}}\{\kappa_k\}$, such that (1) $(\hat{e}^{\,n})^{\frac{1}{n}}\bar{v} \to \infty$ (and therefore $\hat{e}^{\,n} \to \infty$) and (2) $\kappa^n(\hat{e}^{\,n}) \le \sup \kappa(2)\lceil (\hat{e}^{\,n})^{\frac{1}{n}} \bar{v}\rceil$ for each $n \in \mathbb{N}$. Due to compactness, there is a subsequence of natural numbers $\langle n_h\rangle$ and a $\bar{\kappa} \in \bigcup_{k\in\mathbb{N}}\{\kappa_k\}$ such that $\kappa^{n_h} \to \bar{\kappa}$. Given the parenthetical observation earlier in the proof that, for some $\bar{a} > 0$, $\bar{\kappa}(e) > 2\sup \kappa(2)\lceil e^{\bar{a}}\bar{v}\rceil$ for all e sufficiently large and since $1/n_h < \bar{a}$ for all n_h sufficiently large, there exist $\bar{e} > 0$ and $\bar{n} > 0$ such that $\bar{\kappa}(e) > 2\sup \kappa(2)\lceil e^{\frac{1}{n_h}}\bar{v}\rceil$ for all $e > \bar{e}$ and all $n_h > \bar{n}$. But due to (2), for each n_h,

$$\bar{\kappa}(\hat{e}^{\,n_h}) \le \sup \kappa(2)\left\lceil (\hat{e}^{\,n_h})^{\frac{1}{n_h}}\, \bar{v}\right\rceil + (\bar{\kappa}(\hat{e}^{\,n_h}) - \kappa^{n_h}(\hat{e}^{\,n_h})),$$

and hence, since $\kappa^{n_h}(\hat{e}^{\,n_h}) - \bar{\kappa}(\hat{e}^{\,n_h}) \to 0$ and $(\hat{e}^{\,n_h})^{\frac{1}{n_h}}\bar{v} \to \infty$, we conclude that

$$\bar{\kappa}(\hat{e}^{\,n_h}) \le \sup \kappa(2)\left\lceil (\hat{e}^{\,n_h})^{\frac{1}{n_h}}\, \bar{v}\right\rceil + (\bar{\kappa}(\hat{e}^{\,n_h}) - \kappa^{n_h}(\hat{e}^{\,n_h})) \le 2\sup \kappa(2)\left\lceil (\hat{e}^{\,n_h})^{\frac{1}{n_h}}\, \bar{v}\right\rceil$$

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

To prove the Theorem, suppose the set of N' criteria $\mathcal{C}'$ contains a C_k' with e categories. Define $\bar{\mathcal{C}}$ to coincide with $\mathcal{C}'$ except that C_k' is replaced by $\lceil e^{\bar{a}}\bar{v}\rceil$ binary criteria with indices $N'+1, \ldots, N' + \lceil e^{\bar{a}}\bar{v}\rceil$. Due to the previous paragraph, $\kappa[\bar{\mathcal{C}}] < \kappa[\mathcal{C}']$ if $e > \bar{b}$. As for value, $V(\mathcal{C}') = \prod_{j\in\{1,\ldots,N'\}} v(C_j')$ and $V(\bar{\mathcal{C}}) \ge \underline{v}^{\lceil e^{\bar{a}}\bar{v}\rceil}(\prod_{j=\{1,\ldots,N'\}\setminus\{k\}} v(C_j'))$. Hence

$$\begin{aligned} V(\bar{\mathcal{C}}) - V(\mathcal{C}') &\ge \underline{v}^{\lceil e^{\bar{a}}\bar{v}\rceil}\left(\prod_{j=\{1,\ldots,N'\}\setminus\{k\}} v(C_j')\right) - \prod_{j\in\{1,\ldots,N'\}} v(C_j') \\ &= \left(\prod_{j=\{1,\ldots,N'\}\setminus\{k\}} v(C_j')\right)\left(\underline{v}^{\lceil e^{\bar{a}}\bar{v}\rceil} - v(C_k')\right). \end{aligned}$$

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

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

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

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

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

Proof of Proposition 3

Since U is additively separable, we may add a constant to each u_i without changing the orderings that $\mathbb{E}U_{C_i}$ and U represent. Set these constants so that, for $i = 1, \ldots, n$,

$$\mathbb{E}\Big[\max\Big[\mathbb{E}\left[u_i\left(x_i^1(s)\right)\right], \ldots, \mathbb{E}\Big[u_i\Big(x_i^{T(s)}(s)\Big)\Big]\Big]\Big] = 0,$$

where s is fixed in each inner expectation and the outer expectation integrates over s. This condition implies that $\mathbb{E}U_{C_i} = 0$ when $e(C_i) = 1$. Define $v(C_i) = \exp(\mathbb{E}U_{C_i})$ for any feasible C_i and $V(\mathcal{C}) = \exp(U(\mathcal{C}))$ for any set of feasible criteria $\mathcal{C}$.

Since $\exp(\cdot)$ is an increasing transformation and $\exp(U(\mathcal{C})) = \prod_{C_i \in \mathcal{C}} \exp(\mathbb{E}U_{C_i})$, it suffices to show that, for each C_i, $v(C_i)$ is an admissible value for criteria. When $e(C_i) = 1$, $\mathbb{E}U_{C_i} = 0$ and so $v(C_i) = 1$. Set $\underline{v} = \inf\{v(C_i) : i \in \{1, \ldots, n\}, C_i$ is feasible, and $e(C_i) > 1\}$ and define

$$k'(s) = \arg \max_{k \in \{1,\ldots,T(s)\}} \mathbb{E}\Big[u_i\left(x_i^k(s)\right)\Big] \text{ for each } s.$$

By assumption, there is an $\varepsilon > 0$ such that for all C_i with $e(C_i) > 1$ and all s there is a $k''(s)$ with $\mathbb{P}[u_{C_i}[x_i^{k''(s)}(s)] - u_{C_i}[x_i^{k'(s)}(s)] > \varepsilon] > \varepsilon$. Therefore, for any C_i with $e(C_i) > 1$ and each s, $U_{C_i}(s) = \mathbb{E}[\max[u_{C_i}[x_i^1(s)], \ldots, u_{C_i}[x_i^{T(s)}(s)]]]$ is greater than and bounded away from $\max[\mathbb{E}[u_i(x_i^1(s))], \ldots, \mathbb{E}[u_i(x_i^{T(s)}(s))]]$ and so $\mathbb{E}U_{C_i}$ is greater than and bounded away from $\mathbb{E}[\max[\mathbb{E}[u_i(x_i^1(s))], \ldots, \mathbb{E}[u_i(x_i^{T(s)}(s))]]]$. Hence $\underline{v} > 1$.

Fix s and, given a function $f : \mathbb{R} \to \mathbb{R}$, let $\mathbb{E}[f(x_i)|C_i]$ denote the random variable equal to the conditional expectation of $f(x_i)$ given the C_i-category that contains x_i. For each C_i,

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

where, since $u_i(x_i)$ is integrable, for each i the constant $a_i(s) = \sum_{k=1}^{T(s)} \mathbb{E}[|u_i(x_i^k(s))|]$ is well-defined. Since $\mathcal{A}$ is finite, $\mathbb{E}U_{C_i} \leq \mathbb{E}a_i$ and so $v(C_i) \leq \exp(\mathbb{E}a_i)$ for all feasible C_i. With the $v(C_i)$ and $\underline{v}$ already defined, by setting $\bar{v} = \max[\exp(\mathbb{E}a_1), \ldots, \exp(\mathbb{E}a_n)]$ we conclude that the $v(C_i)$ are admissible.

References

Apesteguia, J. and M. Ballester (2010). "The Computational Complexity of Rationalizing Behavior." *Journal of Mathematical Economics*, 46, 356–363.

Apesteguia, J. and M. Ballester (2013). "Choice by Sequential Procedures." *Games and Economic Behavior*, 77, 90–99.

Arrow, K. and H. Raynaud (1986). *Social Choice and Multicriterion Decision-Making*. MIT Press, Cambridge.

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.

Chipman, J. (1960). "The Foundations of Utility." *Econometrica*, 28, 193–224.

Chipman, J. (1971). "On the Lexicographic Representation of Preference Orderings." In *Preference, Utility, and Demand*, edited by J. Chipman, L. Hurwicz, M. Richter, and H. Sonnenschein. Harcourt, New York.

Cowan, N. (2000). "The Magical Number 4 in Short-Term Memory: A Reconsideration of Mental Storage Capacity." *Behavioral and Brain Sciences*, 24, 87–185.

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

Fishburn, P. (1970). *Utility Theory for Decision Making*. Wiley, New York.

Fishburn, P. (1974). "Lexicographic Orders, Utilities and Decision Rules: A Survey." *Management Science*, 20 1442–1471.

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

Houy, N. and K. Tadenuma (2009). "Lexicographic Compositions of Multiple Criteria for Decision Making." *Journal of Economic Theory*, 144, 1770–1782.

Inada, K. (1964). "A Note on the Simple Majority Decision Rule." *Econometrica*, 32, 525–531.

Ju, B. (2011). "Collectively Rational Voting Rules for Simple Preferences." *Journal of Mathematical Economics*, 47, 143–149.

Kahneman, D., A. Treisman, and B. Gibbs (1992). "The Reviewing of Object Files: Object Specific Integration of Information." *Cognitive Psychology*, 24, 175–219.

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

Luck, S. and E. Vogel (1997). "The Capacity of Visual Working Memory for Features and Conjunctions." *Nature*, 390, 279–281.

Mandler, M., P. Manzini, and M. Mariotti (2012). "A Million Answers to Twenty Questions: Choosing by Checklist." *Journal of Economic Theory*, 147, 71–92.

Mandler, M. (2009). "Indifference and Incompleteness Distinguished by Rational Trade." *Games and Economic Behavior*, 67, 300–314.

Mandler, M. (2015). "Rational Agents are the Quickest." *Journal of Economic Theory*, 155, 206–233.

Mandler, M. (2019). "Increasing Returns and the Efficient Organization of Information." Working paper.

Mandler, M. (forthcoming). "The Lexicographic Method in Preference Theory." *Economic Theory*.

Maniquet, F. and P. Mongin (2015). "Approval Voting and Arrow's Impossibility Theorem." *Social Choice and Welfare*, 44, 519–532.

Manzini, P. and M. Mariotti (2007). "Sequentially Rationalizable Choice." *American Economic Review*, 97(5), 1824–1839.

Manzini, P. and M. Mariotti (2012). "Choice by Lexicographic Semiorders." *Theoretical Economics*, 7, 1–23.

Miller, G. (1956). "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information." *Psychological Review*, 63, 81–97.

Rubinstein, A. (1996). "Why are Certain Properties of Binary Relations Relatively More Common in Natural Language?" *Econometrica*, 64, 343–355.

Salant, Y. (2009). "Procedural Analysis of Choice Rules with Applications to Bounded Rationality." *American Economic Review*, 101(2), 724–748.

Sen, A. (1970). “The Impossibility of A Paretian Liberal.” *Journal of Political Economy*, 78, 152–157.
Simon, H. (1972). “Theories of Bounded Rationality.” In *Decision and Organization*, edited by C., McGuire and R. Radner. North-Holland, Amsterdam.
Simon, H. (1974). “How Big is A Chunk?” *Science*, 183, 482–488.
Treisman, A. (2006). “Object Tokens, Binding and Visual Memory.” In *Handbook of Binding and Memory: Perspectives from Cognitive Neuroscience*, edited by H. Zimmer, A. Mecklinger, and U. Lindenberger. Oxford University Press, New York.
Tversky, T. and I. Simonson (1993). “Context-Dependent Preferences.” *Management Science*, 39, 1179–1189.
Vorsatz, M. (2007). “Approval Voting on Dichotomous Preferences.” *Social Choice and Welfare*, 28, 127–141.