Weighted Sets of Probabilities and Minimax Weighted Expected Regret: New Approaches for Representing Uncertainty and Making Decisions*

Joseph Y. Halpern Samantha Leung
Computer Science Department
Cornell University
Ithaca, NY 14853
halpern | samlyy@cs.cornell.edu

April 14, 2015

Abstract

We consider a setting where an decision maker's uncertainty is represented by a set of probability measures, rather than a single measure. Measure-by-measure updating of such a set of measures upon acquiring new information is well-known to suffer from problems. To deal with these problems, we propose using *weighted sets of probabilities*: a representation where each measure is associated with a *weight*, which denotes its significance. We describe a natural approach to updating in such a situation and a natural approach to determining the weights. We then show how this representation can be used in decision-making, by modifying a standard approach to decision making—minimizing expected regret—to obtain *minimax weighted expected regret* (MWER). We provide an axiomatization that characterizes preferences induced by MWER both in the static and dynamic case.

Keywords: Decision theory, ambiguity aversion, minimax regret.
JEL codes: D010, D810

*The authors thank Joerg Stoye for useful comments. Work supported in part by NSF grants IIS-0812045, IIS-0911036, and CCF-1214844, by AFOSR grants FA9550-08-1-0438 and FA9550-09-1-0266, by the Multidisciplinary University Research Initiative (MURI) program administered by the AFOSR under grant FA9550-12-1-0040, and by ARO grants W911NF-09-1-0281 and W(INF-14-1-0017.

	1 defect	10 defect
deliver	10,000	-10,000
cancel	0	0
check	5,001	-4,999

Table 1: Payoffs for the quality-control problem. Acts are in the leftmost column. The remaining two columns describe the outcome for the two sets of states that matter.

1 Introduction

From deciding between crispy fries and a bland salad, to forming an investment portfolio, to military planning, our decisions can significantly impact our lives and those of others. These problems can often be abstracted as decision problems with uncertainty. For decisions based on the outcome of the toss of a fair coin, the uncertainty can be well characterized by probability. However, what is the probability of you gaining weight if you eat fries at every meal? What if you have salads instead? Even experts would not agree on a single probability.

Representing uncertainty by a single probability measure and making decisions by maximizing expected utility leads to further problems. Consider the following stylized quality control-problem, which serves as a running example in this paper. A business owner (the decision maker) is contracted to produce $1,000$ items, and will be rewarded when she delivers the items, but punished if she delivers a batch with too many defective items. She has recently switched a raw material supplier, and does not know whether the supplier is reliable, and provides good quality materials, or unreliable, and provides bad quality materials. For simplicity, assume that using good quality raw materials results in one defective item in the batch of $1,000$, while using bad quality raw materials results in ten defective items.[1]

The owner's choices and their consequences are summarized in Table 1. Decision theorists typically model decision problems with states, acts, and outcomes: the world is in one of many possible states, and the decision maker chooses an *act*, a function mapping states to outcomes. For now, we use the simplest possible state space for this problem: $\{\text{one defect}, \text{ten defects}\}$. These two possible states are sufficient to capture the owner's uncertainty in the quality-control problem. (However, we later consider a more refined

[1] It is more natural to assume that the quality of the raw materials affects the distribution of the number of defects, rather than directly affecting the final number; we make the latter assumption here for simplicity.

state space.)

The owner can choose among three acts: *deliver*: deliver the products; *cancel*: cancel the contract; or *check*: inspect enough items to determine the number of defects, then decide to deliver or cancel the contract. The client will tolerate at most one defect in the lot of $1,000$. Therefore, if the owner chooses *deliver*, and if there is only one defect, then the client is happy, and the owner obtains a utility of $10,000$; if there are ten defects, then the outcome then the client will penalize the owner, resulting in a utility of $-10,000$. If the owner chooses to *cancel*, then the contract is canceled, and the owner gets a utility of 0. Finally, checking the items costs $4,999$ units of utility but is reliable, so if the owner chooses *check*, and if there is one defect, then the products will be delivered after the check, and the client will be happy; the owner nets a utility of $5,001$. On the other hand, if there are ten defects, then after the check, the contract will be canceled, and the owner nets a utility of $-4,999$.

To maximize expected utility, we must assume some probability over states. What measure should be used? There are two hypotheses that the owner entertains: (1) the raw material is of high quality and (2) the raw material is of low quality. Each of these hypotheses places a different probability on states.

If the raw material is of high quality, then with probability 1 there will only be one defect; if the raw material is of low quality, then with probability 1 there will be ten defects. One way to model the owner's uncertainty about the quality of the material is to take each hypothesis to be equally likely. However, not having any idea about which hypothesis holds is very different from believing that all hypotheses are equally likely. It is easy to check that taking each hypothesis to be equally likely makes *check* the act that maximizes utility, but taking the probability that the raw material has low quality .51 makes *cancel* the act that maximizes expected utility, and taking the probability that the raw material has high quality to be .51 makes *deliver* the act that maximizes expected utility. What makes any of these choices the "right" choice?

It is easy to construct many other examples where a single probability measure does not capture uncertainty, and does not result in what seem to be reasonable decisions, when combined with expected utility maximization. A natural alternative, which has often been considered in the literature, is to represent the decision maker's uncertainty by a *set* of probability measures. For example, in the quality-control problem, the owner's beliefs could be represented by two probability measures, $\Pr_1$ and $\Pr_{10}$, one for each hypothesis. Thus, $\Pr_1$ assigns uniform probability to all states with exactly one defective items, and $\Pr_{10}$ assigns uniform probability to all states with exactly ten defective items.

But this representation also has problems. Consider the quality-control problem again. Why should the owner be sure that there is exactly either one or ten defective items? Of course, we can replace these two hypotheses by hypotheses that say that the probability of an item being defective is either .001 or .01, but this doesn't solve the problem. Why should the decision maker be sure that the probability is either exactly .001 or exactly .01? Couldn't it also be .0999? Representing uncertainty by a set of measures still places a sharp boundary on what measures are considered possible and impossible.

A second problem involves updating beliefs. How should beliefs be updated if they are represented by a set of probability measures? The standard approach for updating a single measure is by conditioning. The natural extension of conditioning to sets of measure is measure-by-measure updating: conditioning each measure on the information (and also removing measures that give the information probability 0).

However, measure-by-measure updating can produce some rather counterintuitive outcomes. In the quality-control problem, suppose that the owner knows that the first 100 items that came off the assembly line are good. We denote this piece of information by E. Intuition tells us that there is now more reason to believe that there is only one defective item. The simple two-state state space we used is not sufficient to capture this new information, so we now expand the state space from $\{\text{one defect}, \text{ten defects}\}$ to $\{\text{good}, \text{defective}\}^{1000}$. That is, each item that gets produced is numbered from 1 to $1,000$, and each one can be either defective, or good. We can adapt the two hypotheses in the obvious way to this new state space, and the hypotheses can be conditioned on the new information.

However, $\Pr_1 \mid E$ places uniform probability on all states where the first 100 items are good, and there is exactly one defective item among the last 900 items. Similarly, $\Pr_{10} \mid E$ places uniform probability on all states where the first 100 items are good, and there are exactly ten defective items among the last 900. $\Pr_1 \mid E$ still places probability 1 on there being one defective product, just like $\Pr_1$, and $\Pr_{10} \mid E$ still places probability 1 on there being ten defective products. There is no way to capture the fact that the owner now views the hypothesis $\Pr_{10}$ as less likely, even if the owner was told that the first 990 items are all good!

Of course, both of these problems would be alleviated if we placed a probability on hypotheses, but, as we have already observed, simply maximizing expected utility with respect to this second-order probability distribution has its problems. In this paper, we propose an intermediate approach: representing uncertainty using *weighted sets of probabilities.* That is, each probability measure is associated with a weight. These weights can be viewed as probabilities; indeed, if the set of probabilities is finite, we can normalize them so that they are effectively probabilities. Moreover, in one important setting, we update them in the same way that we would update probabilities, using likelihood (see below). On the other hand, these weights do not act like probabilities if the set of probabilities is infinite. For example, if we had a countable set of hypotheses, we could assign them all weight 1 (so that, intuitively, they are all viewed as equally likely), but there is no uniform measure on a countable set. To avoid complications about measurability, we think of our representation as a weighted set of probabilities. However, one can equally well think of the representation as a probability on probabilities.

More importantly, when it comes to decision making, we use the weights quite differently from how we would use second-order probabilities on probabilities. Second-order probabilities would let us define a probability on events (by taking expectation) and maximize expected utility, in the usual way. Using the weights, we instead define a novel decision rule, *minimax weighted expected regret (MWER)*, that has some rather nice properties. If all the weights are 1, then MWER is just the standard *minimax expected regret (MER)* rule (described below). If the set of probabilities is a singleton, then MWER agrees with (subjective) expected utility maximization (SEU). More interestingly perhaps, if, through updating, the weights converge to a single hypothesis/probability measure (which happens in one important special case, discussed below), MWER converges to SEU. Thus, the weights give us a smooth, natural way of interpolating between MER and SEU.

In summary, weighted sets of probabilities allow us to represent ambiguity (uncertainty about the correct probability distribution). Real individuals are sensitive to this ambiguity when making decisions, and the MWER decision rule takes this into account. Updating the weighted sets of probabilities using likelihood allows the initial ambiguity to be resolved as more information about the true distribution is obtained.

We now briefly explain MWER, by first discussing MER. MER is a probabilistic variant of the minimax regret decision rule proposed by Niehans [21] and Savage [25].[2] Most likely, at some point, we've second-guessed

[2]Note that our definition of regret minimization, while standard, differs from that used by Loomes and Sugden [18], where probabilities are given, and where the decision maker not only feels regret but also "rejoice" if the chosen alternative is better than the unchosen ones.

ourselves and thought "had I known this, I would have done that instead". That is, in hindsight, we regret not choosing the act that turned out to be optimal for the realized state, called the *ex post* optimal act. The *regret* of an act a in a state s is the difference (in utility) between the ex post optimal act in s and a. Of course, typically one does not know the true state at the time of decision. Therefore the regret of an act is the worst-case regret, taken over all states. The *minimax regret* rule orders acts by their regret.

The definition of regret can be used if there is no probability on states. If an decision maker's uncertainty is represented by a single probability measure, then we can compute the *expected regret* of an act a: just multiply the regret of an act a at a state s by the probability of s, and then sum. It is well known that the order on acts induced by minimizing expected regret is identical to that induced by maximizing expected utility (see [13] for a proof). If an decision maker's uncertainty is represented by a set $\mathcal{P}$ of probabilities, then we can compute the expected regret of an act a with respect to each probability measure $\Pr \in \mathcal{P}$, and then take the worst-case expected regret. The MER (Minimax Expected Regret) rule orders acts according to their worst-case expected regret, preferring the act that minimizes the worst-case regret. If the set of measures is the set of *all* probability measures on states, then it is not hard to show that MER induces the same order on acts as (probability-free) minimax regret. Thus, MER generalizes both minimax regret (if $\mathcal{P}$ consists of all measures) and expected utility maximization (if $\mathcal{P}$ consists of a single measure).

MWER further generalizes MER. If we start with a *weighted* set of measures, then we can compute the weighted expected regret for each one (just multiply the expected regret with respect to Pr by the weight of Pr) and compare acts by their worst-case weighted expected regret.

Sarver [24] also proves a representation theorem that involves putting a multiplicative weight on a regret quantity. However, his representation is fundamentally different from MWER. In his representation, regret is a factor only when comparing two *sets* of acts; the ranking of individual acts is given by expected utility maximization. By way of contrast, we do not compare sets of acts.

It is standard in decision theory to axiomatize a decision rule by means of a representation theorem. For example, Savage [26] showed that if an decision maker's preferences $\succeq$ satisfied several axioms, such as completeness and transitivity, then the decision maker is behaving as if she is maximizing expected utility with respect to some utility function and probabilistic belief.

If uncertainty is represented by a set of probability measures, then we can generalize expected utility maximization to *maxmin expected utility*

(MMEU). MMEU compares acts by their worst-case expected utility, taken over all measures. MMEU has been axiomatized by Gilboa and Schmeidler [11]. MER was axiomatized by Hayashi [13] and Stoye [28]. We provide an axiomatization of MWER. We make use of ideas introduced by Stoye [28] in his axiomatization of MER, but the extension seems quite nontrivial.

We also consider a dynamic setting, where beliefs are updated by new information. If observations are generated according to a probability measure that is stable over time, then, as we suggested above, there is a natural way of updating the weights given observations, using ideas of likelihood. The idea is straightforward. After receiving some information E, we update each probability $\Pr \in \mathcal{P}$ to $\Pr \mid E$, and take its weight to be $\alpha_{\Pr} = \Pr(E)/\sup_{\Pr' \in \mathcal{P}} \Pr'(E)$. If more than one $\Pr \in \mathcal{P}$ gets updated to the same $\Pr \mid E$, the sup of all such weights is used. Thus, the weight of Pr after observing E is modified by taking into account the likelihood of observing E assuming that Pr is the true probability. We refer to this method of updating weights as *likelihood updating*.

If observations are generated by a stable measure (e.g., we observe the outcomes of repeated flips of a biased coin) then, as the decision maker makes more and more observations, the weighted set of probabilities of the decision maker will, almost surely, look more and more like a single measure. The weight of the measures in $\mathcal{P}$ closest to the measure generating the observations converges to 1, and the weight of all other measures converges to 0. This would not be the case if uncertainty were represented by a set of probability measures and we did measure-by-measure updating, as is standard. As we mentioned above, this means that MWER converges to SEU.

We provide an axiomatization for dynamic MWER with likelihood updating. We remark that a dynamic version of MMEU with measure-by-measure updating has been axiomatized by Jaffray [15], Pires [22], and Siniscalchi [27]. Likelihood updating is somewhat similar in spirit to an updating method implicitly proposed by Epstein and Schneider [8]. They also represented uncertainty by using (unweighted) sets of probability measures. They choose a threshold α with $0 < \alpha < 1$, update by conditioning, and eliminate all measures whose relative likelihood does not exceed the threshold. This approach also has the property that, over time, all that is left in $\mathcal{P}$ are the measures closest to the measure generating the observations; all other measures are eliminated. However, it has the drawback that it introduces a new, somewhat arbitrary, parameter α.

Chateauneuf and Faro [4] also consider weighted sets of probabilities (they model the weights using what they call *confidence functions*). They then define and provide a representation of a generalization of MMEU using weighted sets of probabilities that parallels our generalization of MER. Chateauneuf and Faro do not discuss the dynamic situation; specifically, they do not consider how weights should be updated in light of new information. We discuss the relationship of our work that of Chateauneuf and Faro in more detail in Section B.

Klibanoff et al. [16] propose a model of decision making that associates weights with probability measures, but makes decisions based on a "weighted" expected utility function. Maccheroni et al. [19] study a model of decision making where additive, instead of multiplicative, weights are associated with probability measures. Hayashi [13] considers a model of expected-regret-minimization where regrets computed with respect to each state are taken to a positive power before expectations are taken. Others have also proposed and studied approaches of representing uncertainty that are similar to weighted probabilities (see, e.g. [1, 6, 20, 31]).

The rest of this paper is organized as follows. Section 2 introduces the weighted sets of probabilities representation, and Section 3 introduces the MWER decision rule. Axiomatic characterizations of static and dynamic MWER are provided in Sections 4 and 5, respectively. We conclude in Section 6.

2 Weighted Sets of Probabilities

A set $\mathcal{P}^+$ of *weighted probability measures* on a set S consists of pairs $(\Pr, \alpha_{\Pr})$, where $\alpha_{\Pr} \in [0,1]$ and $\Pr$ is a probability measure on S.[3] Let $\mathcal{P} = \{\Pr : \exists \alpha ((\Pr, \alpha) \in \mathcal{P}^+)\}$. We assume that, for each $\Pr \in \mathcal{P}$, there is exactly one α such that $(\Pr, \alpha) \in \mathcal{P}^+$. We denote this number by $\alpha_{\Pr}$, and view it as the *weight of* $\Pr$. We further assume for convenience that weights have been normalized so that there is at least one measure $\Pr \in \mathcal{P}$ such that $\alpha_{\Pr} = 1$.[4]

As we observed in the introduction, one way of updating weighted sets of probabilities is by using likelihood updating. We use $\mathcal{P}^+ \mid E$ to denote the result of applying likelihood updating to $\mathcal{P}^+$. Define $\overline{\mathcal{P}}^+(E) = \sup\{\alpha_{\Pr} \Pr(E) :$

[3]In this paper, for ease of exposition, we take the state space S to be finite, and assume that all sets are measurable. We can easily generalize to arbitrary measure spaces.

[4]While we could take weights to be probabilities, and normalize them so that they sum to 1, if $\mathcal{P}$ is finite, this runs into difficulties if we have an infinite number of measures in $\mathcal{P}$. For example, if we are tossing a coin, and $\mathcal{P}$ includes all probabilities on heads from 1/3 to 2/3, using a uniform probability, we would be forced to assign each individual probability measure a weight of 0, which would not work well in the definition of MWER.

$\Pr \in \mathcal{P}\}$; if $\overline{\mathcal{P}}^+(E) > 0$, set $\alpha_{\Pr,E} = \sup_{\{\Pr' \in \mathcal{P} : \Pr' | E = \Pr | E\}} \frac{\alpha_{\Pr'} \Pr'(E)}{\overline{\mathcal{P}}^+(E)}$. Note that given a measure $\Pr \in \mathcal{P}$, there may be several distinct measures $\Pr'$ in $\mathcal{P}$ such that $\Pr' \mid E = \Pr \mid E$. Thus, we take the weight of $\Pr \mid E$ to be the sup of the possible candidate values of $\alpha_{\Pr,E}$. By dividing by $\overline{\mathcal{P}}^+(E)$, we guarantee that $\alpha_{\Pr,E} \in [0,1]$, and that there is some measure Pr such that $\alpha_{\Pr,E} = 1$, as long as there is some pair $(\alpha_{\Pr}, \Pr) \in \mathcal{P}$ such that $\alpha_{\Pr} \Pr(E) = \overline{\mathcal{P}}^+(E)$. If $\overline{\mathcal{P}}^+(E) > 0$, we take $\mathcal{P}^+ \mid E$ to be

$$\{(\Pr \mid E, \alpha_{\Pr,E}) : \Pr \in \mathcal{P}\}.$$

If $\overline{\mathcal{P}}^+(E) = 0$, then $\mathcal{P}^+ \mid E$ is undefined.

In computing $\mathcal{P}^+ \mid E$, we update not just the probability measures in $\mathcal{P}$, but also their weights. The new weight combines the old weight with the likelihood. Clearly, if all measures in $\mathcal{P}$ assign the same probability to the event E, then likelihood updating and measure-by-measure updating coincide. This is not surprising, since such an observation E does not give us information about the relative likelihood of measures. We stress that using likelihood updating is appropriate only if the measure generating the observations is assumed to be stable. For example, if observations of heads and tails are generated by coin tosses, and a coin of possibly different bias is tossed in each round, then likelihood updating would not be appropriate.

It is well known that, when conditioning on a single probability measure, the order that information is acquired is irrelevant; the same observation easily extends to sets of probability measures. As we now show, it can be further extended to weighted sets of probability measures.

Proposition 1. *Likelihood updating is consistent in the sense that for all* $E_1, E_2 \subseteq S$, $(\mathcal{P}^+ \mid E_1) \mid E_2 = (\mathcal{P}^+ \mid E_2) \mid E_1 = \mathcal{P}^+ \mid (E_1 \cap E_2)$, *provided that* $\mathcal{P}^+ \mid (E_1 \cap E_2)$ *is defined.*

Proof. By standard results, $(\Pr \mid E_1) \mid E_2 = (\Pr \mid E_2) \mid E_1 = \Pr \mid (E_1 \cap E_2)$. Since the weight of the measure $\Pr \mid E_1$ is proportional to $\alpha_{\Pr} \Pr(E_1)$, the weight of $(\Pr \mid E_1) \mid E_2$ is proportional to $\alpha_{\Pr} \Pr(E_1) \Pr(E_2 \mid E_1) = \alpha_{\Pr} \Pr(E_1 \cap E_2)$. Likewise, the weight of $(\Pr \mid E_2) \mid E_1$ is proportional to $\alpha_{\Pr} \Pr(E_2) \Pr(E_1 \mid E_2) = \alpha_{\Pr} \Pr(E_1 \cap E_2)$. Since, in all these cases, the sup of the weights is normalized to 1, the weights of corresonding measures in $\mathcal{P}^+ \mid (E_1 \cap E_2)$, $(\mathcal{P}^+ \mid E_1) \mid E_2$ and $(\mathcal{P}^+ \mid E_2) \mid E_1$ must be equal. □

3 MWER

We now define MWER formally. Given a set S of states and a set X of outcomes, an *act* f (over S and X) is a function mapping S to X. For simplicity in this paper, we take S to be finite. Associated with each outcome $x \in X$ is its utility $u(x)$. We call a tuple (S, X, u) a *(non-probabilistic) decision problem*. To define regret, we need to assume that we are also given a set $M \subseteq X^S$ of feasible acts, called the *menu*. The reason for the menu is that, as is well known (and we will demonstrate by example shortly), regret can depend on the menu. Moreover, we assume that every menu M has utilities bounded from above. That is, we assume that for all menus M, $\sup_{g \in M} u(g(s))$ is finite. This ensures that the regret of each act is well defined.[5] For a menu M and act $f \in M$, the regret of f with respect to M and decision problem (S, X, u) in state s is

$$reg_M(f, s) = \left(\sup_{g \in M} u(g(s)) \right) - u(f(s)).$$

That is, the regret of f in state s (relative to menu M) is the difference between $u(f(s))$ and the highest utility possible in state s (among all the acts in M). The regret of f with respect to M and decision problem (S, X, u) is the worst-case regret over all states:

$$\max_{s \in S} reg_M(f, s).$$

We denote this as $reg_M^{(S,X,u)}(f)$, and usually omit the superscript (S, X, u) if it is clear from context. If there is a probability measure Pr over the states, then we can consider the *probabilistic decision problem* $(S, X, u, \Pr)$. The *expected regret* of f with respect to M is

$$reg_{M,\Pr}(f) = \sum_{s \in S} \Pr(s) reg_M(f, s).$$

If there is a set $\mathcal{P}$ of probability measures over the states, then we consider the $\mathcal{P}$-decision problem $(S, X, u, \mathcal{P})$. The maximum expected regret of $f \in$

[5] Stoye [29] assumes that, for each menu M, there is a finite set A_M of acts such that M consists of all the convex combinations of the acts in A_M. We clearly allow a larger set of menus than Stoye. We return to the issue of what menus to consider after we discuss the representation theorem in Section B, and again when we discuss choice functions in Section 4.

M with respect to M and $(S, X, u, \mathcal{P})$ is

$$reg_{M,\mathcal{P}}(f) = \sup_{\Pr \in \mathcal{P}} \left(\sum_{s \in S} \Pr(s) reg_M(f, s) \right).$$

Finally, if beliefs are modeled by weighted probabilities $\mathcal{P}^+$, then we consider the $\mathcal{P}^+$-decision problem $(S, X, u, \mathcal{P}^+)$. The maximum weighted expected regret of $f \in M$ with respect to M and $(S, X, u, \mathcal{P}^+)$ is

$$reg_{M,\mathcal{P}^+}(f) = \sup_{\Pr \in \mathcal{P}} \left(\alpha_{\Pr} \sum_{s \in S} \Pr(s) reg_M(f, s) \right).$$

The MER decision rule is thus defined for all $f, g \in X^S$ as

$$f \succeq^{S,X,u}_{M,\mathcal{P}} g \text{ iff } reg^{(S,X,u)}_{M,\mathcal{P}}(f) \leq reg^{(S,X,u)}_{M,\mathcal{P}}(g).$$

That is, f is preferred to g if the maximum expected regret of f is less than that of g. We can similarly define $\succeq_{M,reg}$, $\succeq^{S,X,u}_{M,\Pr}$, and $\succeq^{S,X,u}_{M,\mathcal{P}^+}$ by replacing $reg^{(S,X,u)}_{M,\mathcal{P}}$ by $reg^{(S,X,u)}_{M}$, $reg^{(S,X,u)}_{M,\Pr}$, and $reg^{(S,X,u)}_{M,\mathcal{P}^+}$, respectively. Again, we usually omit the superscript (S, X, u) and subscript Pr or $\mathcal{P}^+$, and just write $\succeq_M$, if it is clear from context.

To see how these definitions work, consider the quality-control problem from the introduction, there are $1,000$ states with one defective item, and $C(1000, 10)$ states with ten defective items, where $C(m, n)$ is the number of combinations of n items from a collection of 1000 unique items. The regret of each action in a state depends only on the number of defective items, and is given in Table 2. It is easy to see that the action that minimizes regret is *check*, with *deliver* and *cancel* having equal regret. If we represent uncertainty using the two probability measures $\Pr_1$ and $\Pr_{10}$, the expected regret of each of the acts with respect to $\Pr_1$ (resp., $\Pr_{10}$) is just its regret with respect to states with one (resp. ten) defective items. Thus, the action that minimizes maximum expected regret is again *check*.

As we said above, the ranking of acts based on MER or MWER can change if the menu of possible choices changes. For example, suppose that we introduce a new choice in the quality-control problem, whose gains and losses are twice those of *deliver*, resulting in the payoffs and regrets described in Table 3. In this new setting, *deliver* has a lower maximum expected regret $(10,000)$ than *check* $(14,999)$, so MER prefers *deliver* over *check*. Thus, the introduction of a new choice can affect the relative order of acts according to MER (and MWER), even though other acts are preferred to the new

	1 defect		10 defects	
	Payoff	Regret	Payoff	Regret
deliver	10,000	0	-10,000	10,000
cancel	0	10,000	0	0
check	5,001	4,999	-4,999	4,999

Table 2: Payoffs and regrets for the quality-control problem.

	1 defect		10 defects	
	Payoff	Regret	Payoff	Regret
deliver	10,000	10,000	-10,000	10,000
cancel	0	20,000	0	0
check	5,001	14,999	-4,999	4,999
new	20,000	0	-20,000	20,000

Table 3: Payoffs and regrets for the quality-control problem with a new choice added.

choice. By way of contrast, the decision rules MMEU and SEU are *menu-independent*; the relative order of acts according to MMEU and SEU is not affected by the addition of new acts.

We next consider a dynamic situation, where the decision maker acquires information about a stable randomizing mechanism (i.e., a stationary probability distribution). Specifically, in the context of the quality-control problem,

suppose that The owner learns E—the first 100 items are good. Initially, suppose that The owner has no reason to believe that one hypothesis is more likely than the other, so assigns both hypotheses weight 1. Note that $P_1(E) = 0.9$ and $\Pr_{10}(E) = C(900,10)/C(1000,10) \approx 0.35$. Thus,

$$\mathcal{P}^+ \mid E = \{(\Pr_1 \mid E, 1), (\Pr_{10} \mid E, C(900,10)/(.9C(1000,10))\}.$$

We can also see from this example that MWER interpolates between MER and expected utility maximization. Suppose that a passer-by tells The owner that the first N cupcakes are good. If $N = 0$, MWER with initial weights 1 is the same as MER. On the other hand, if $N \geq 991$, then the likelihood of $\Pr_{10}$ is 0, and the only measure that has effect is $\Pr_1$, which means minimizing maximum weighted expected regret is just maximizing expected utility with respect to $\Pr_1$. If $0 < N < 991$, then the likelihoods (hence weights) of $\Pr_1$ and $\Pr_{10}$ are 1 and $\frac{C(1000-N,10)}{C(1000,10)} \times \frac{1000}{1000-N} <$

$((999 - N)/999)^9$. Thus, as N increases, the weight of $\Pr_{10}$ goes to 0, while the weight of $\Pr_1$ stays at 1.

4 A Characterization of MWER

We now provide a representation theorem for MWER. That is, we provide a collection of properties (i.e., axioms) that hold of MWER such that a preference order on acts that satisfies these properties can be viewed as arising from MWER. To get such an axiomatic characterization, we restrict to what is known in the literature as the *Anscombe-Aumann* (AA) framework [2], where outcomes are lotteries on prizes. This framework is standard in the decision theory literature; axiomatic characterizations of SEU [2], MMEU [11], and MER [13, 28] have already been obtained in the AA framework. We draw on these results to obtain our axiomatization.

In this section, we provide a characterization of MWER using choice functions as the primitives. In Appendix B, we provide a characterization of MWER with menu-indexed preference orders as the primitives, which allows us to compare our axioms to axioms that have been used to characterize other decision rules.

A choice function maps every finite set M of acts to a subset M' of M. Intuitively, the set M' consists of the "best" acts in M. Thus, a choice function gives less information than a preference order; it gives only the top elements of the preference order. Stoye [29] provides a representation theorem for MER where the axioms are described in terms of choice functions.

As usual, a *choice function* C is a function from menus to menus, where $C(M) \subseteq M$.[6] We start by taking the domain of a choice to be the set $\mathcal{M}_B$ of all bounded menus. As we now show, we can get a representation theorem for MWER by using all the axioms given by Stoye [28], except for the "betweenness" axiom that restricts the representation to consist of probability distributions (of weight one). For completeness, we reproduce the axioms below.

Given a set Y (which we view as consisting of *prizes*), a *lottery* over Y is just a probability with finite support on Y. Let $\Delta(Y)$ consist of all finite probabilities over Y. In the AA framework, the set of outcomes has the form $\Delta(Y)$. So now acts are functions from S to $\Delta(Y)$. (Such acts are sometimes called *Anscombe-Aumann acts*.) We can think of a lottery as modeling objective uncertainty, while a probability on states models subjective uncertainty; thus, in the AA framework we have both objective and

[6] We use $\subseteq$ to denote subset, and $\subset$ to denote strict subset.

subjective uncertainty. The technical advantage of considering such a set of outcomes is that we can consider convex combinations of acts. If f and g are acts, define the act $\alpha f + (1-\alpha)g$ to be the act that maps a state s to the lottery $\alpha f(s) + (1-\alpha)g(s)$.

In this setting, we assume that there is a utility function U on prizes in Y. The utility of a lottery l is just the expected utility of the prizes obtained, that is,

$$u(l) = \sum_{\{y \in Y\,:\, l(y)>0\}} l(y)U(y).$$

This makes sense since $l(y)$ is the probability of getting prize y if lottery l is played. The expected utility of an act f with respect to a probability Pr is then just $u(f) = \sum_{s\in S} \Pr(s)u(f(s))$, as usual. We also assume that there are at least two prizes y_1 and y_2 in Y, with different utilities $U(y_1)$ and $U(y_2)$.

Given a set Y of prizes, a utility U on prizes, a state space S, and a set $\mathcal{P}^+$ of weighted probabilities on S, we can define a family $\succeq^{S,\Delta(Y),u}_{M,\mathcal{P}^+}$ of preference orders on Anscombe-Aumann acts determined by weighted regret, one per menu M, as discussed above, where u is the utility function on lotteries determined by U. For ease of exposition, we usually write $\succeq^{S,Y,U}_{M,\mathcal{P}^+}$ rather than $\succeq^{S,\Delta(Y),u}_{M,\mathcal{P}^+}$.

Axiom 1 (Nontriviality). *$C(M) \subset M$ for some menu M.*

Let $g(s)^*$ denote the constant act that maps all states to the outcome $g(s)$. Given a state s, define the choice function C_s by taking $f \in C_s(M)$ if and only if $f(s)^* \in C(\{g(s)^* : g \in M\})$. Thus, C_s is a choice function that is concerned only with state s.

Axiom 2 (Monotonicity). *If $f \in M$ and $f \in C_s(M)$ for all s, then $f \in C(M)$.*

Intuitively, this axioms says that if, for each state s, f is a best choice in M when restricted to s, then f is a best choice in M overall.

Given a menu M and an act f, let $\lambda M + (1-\lambda)f$ be the menu $\{\lambda h + (1-\lambda)f : h \in M\}$.

Axiom 3 (Independence). *$C(\lambda M + (1-\lambda)f) = \lambda C(M) + (1-\lambda)f$.*

Axiom 4 (Independence of Irrelevant Alternatives (IIA) for Constant Acts). *If M and N consist of constant acts, then*

$$C(M \cup N) \cap M \in \{C(M), \emptyset\}.$$

Axiom 5 (Independence of Never Strictly Optimal Alternatives (INA))**.** *If* $C_s(M \cup N) \cap M \neq \emptyset$ *for all* s*, then*

$$C(M \cup N) \cap M \in \{C(M), \emptyset\}.$$

Axiom 6 (Mixture Continuity)**.** *If* $f \notin M$ *and* $C(M \cup \{f\}) = \{f\}, g \in M$*, and* h *is an act, then there exists* $\lambda \in (0,1)$ *such that* $C(M \cup \{\lambda f + (1-\lambda)h\}) = \{\lambda f + (1-\lambda)h\}$*, and* $\lambda g + (1-\lambda)h \notin C(M \cup \{f, \lambda g + (1-\lambda)h\})$*.*

Axiom 7 (Ambiguity Aversion)**.** *If* $\lambda \in [0,1]$*, and* $M \supseteq \{f, g, \lambda f + (1-\lambda)g\}$*, and* $\{f,g\} \subseteq C(M)$*, then* $\lambda f + (1-\lambda)g \in C(M)$*.*

The MWER choice function is defined as

$$C^{S,Y,U}_{\mathcal{P}^+}(M) = \operatorname*{argmin}_{f \in M} reg^{(S,X,u)}_{M,\mathcal{P}^+}(f).$$

We now state and prove a representation theorem for MWER. Roughly, the representation theorem states that a choice function satisfies Axioms 1–5 if and only if it has a MWER representation with respect to some utility function and weighted probabilities. In the representation theorem for SEU [2], not only is the utility function unique (up to affine transformations, so that we can replace U by $aU + b$, where $a > 0$ and b are constants), but the probability is unique as well. Similarly, in the MMEU representation theorem of Gilboa and Schmeidler [11], the utility function is unique, and the set of probabilities is also unique, as long as one assume that the set is convex and closed.

To get uniqueness in the representation theorem for MWER, we need to consider a different representation of weighted probabilities. Define a *sub-probability measure* $\mathbf{p}$ on S to be like a probability measure (i.e., a function mapping measurable subsets of S to $[0,1]$ such that $\mathbf{p}(T \cup T') = \mathbf{p}(T) + \mathbf{p}(T')$ for disjoint sets T and T'), without the requirement that $\mathbf{p} = 1$. We can identify a weighted probability distribution $(\Pr, \alpha)$ with the sub-probability measure $\alpha \Pr$. (Note that given a sub-probability measure $\mathbf{p}$, there is a unique pair $(\alpha, \Pr)$ such that $\mathbf{P} = \alpha \Pr$: we simply take $\alpha = \mathbf{p}(S)$ and $\Pr = \mathbf{p}/\alpha$.) A set B of sub-probability measures is *downward-closed* if, whenever $\mathbf{p} \in B$ and $\mathbf{q} \leq \mathbf{p}$, then $\mathbf{q} \in B$. We get a unique set of sub-probability measures in our representation theorem if we restrict to sets that are convex, downward-closed, closed, and contain at least one (proper) probability measure. (The latter requirement corresponds to having $\alpha_{\Pr} = 1$ for some $\Pr \in \mathcal{P}^+$.) For convenience, we will call a set *regular* if it is convex, downward-closed, and closed.

We identify each set of weighted probabilities $\mathcal{P}^+$ with the set of sub-probability measures

$$C(\mathcal{P}^+) = \{\alpha \Pr : (\Pr, \alpha_{\Pr}) \in \mathcal{P}^+, 0 \leq \alpha \leq \alpha_{\Pr}\}.$$

Note that if $(\alpha, \Pr) \in \mathcal{P}^+$, then $C(\mathcal{P}^+)$ includes all the sub-probability measures between the all-zero measure and $\alpha_{\Pr} \Pr$.

We need to restrict to closed and convex sets of sub-probability measures to get uniqueness in the representation of MWER for much the same reason that we need to restrict to closed and convex sets to get uniqueness in the representation of MMEU. Convexity is needed because a set B of sub-probability measures always induce the same MWER preferences as its convex hull. For example, consider the quality-control problem and the expected regrets in Table 2, and the distribution $a \Pr_1 + (1-a) \Pr_{10}$, for some $a \in (0,1)$. The weighted expected regret of any act with respect to $a \Pr_1 + (1-a) \Pr_{10}$ is bounded above by the maximum weighted expected regret of that act with respect to $\Pr_1$ and $\Pr_{10}$. Therefore, adding $a \Pr_1 + (1-a) \Pr_{10}$ to $\mathcal{P}^+$ for some weight $a \in (0,1)$ does not change the resulting family of preferences. Similarly, we need to restrict to closed sets for uniqueness, since if we start with a set B of sub-probability measures that is not closed, taking the closure of B would result in the same family of preferences.

While convexity is easy to define for a set of sub-probability measures, there seems to be no natural notion of convexity for a set $\mathcal{P}^+$ of weighted probabilities. Moreover, the requirement that $\mathcal{P}^+$ is closed is different from the requirement that $C(\mathcal{P}^+)$ is closed. The latter requirement seems more reasonable. For example, fix a probability measure Pr, and let $\mathcal{P}^+ = \{(1, \Pr)\} \cup \{(0, \Pr') : \Pr' \neq \Pr\}$. Thus, $\mathcal{P}^+$ essentially consists of a single probability measure, namely Pr, with weight 1; all the weighted probability measures $(0, \Pr')$ have no impact. This represents the uncertainty of an decision maker who is sure that that Pr is true probability. Clearly $\mathcal{P}^+$ is not closed, since we can find a sequence $\Pr_n$ such that $(0, \Pr_n) \to (0, \Pr)$, although $(0, \Pr) \notin \mathcal{P}^+$. But $C(\Pr^+)$ is closed.

Restricting to closed, convex sets of sub-probability measures does not suffice to get uniqueness; we also need to require downward-closedness. This is so because if $\mathbf{p}$ is in B, then adding any $\mathbf{q} \leq \mathbf{p}$ to the set leaves all regrets unchanged. Finally, the presence of a proper probability measure is also required, since for all $a \in (0,1]$, scaling each element in the set B by a leaves the family of preferences unchanged.

In summary, if we consider arbitrary sets of sub-probability measures, then the set of sub-probability measures that represent a given family of

MWER preferences is unique if we require it to be regular and contain a probability measure.

Although we have assumed that the set of menus is $\mathcal{M}_B$, other sets have been considered in the literature. In particular, Stoye considers the set $\mathcal{M}_C$ of menus that are convex hulls of a finite number of acts, and the set $\mathcal{M}_F$ of finite menus [28, 29]. As we now show, the representation theorem holds for both $\mathcal{M}_F$ and $\mathcal{M}_B$. In Appendix B, we show that if we consider preference orders as opposed to choice functions, then the corresponding representation theorem holds for $\mathcal{M}_C$ as well as $\mathcal{M}_F$ and $\mathcal{M}_B$.

Theorem 1. *For all Y, U, S, and $\mathcal{P}^+$, the choice function $C_{\mathcal{P}^+}^{S,Y,U}$ satisfies Axioms 1–7 for all $M \in \mathcal{M}_B$ (resp., $\mathcal{M}_F$). Conversely, if a choice function C satisfies Axioms 1–7 for all $M \in \mathcal{M}_B$ (resp., $\mathcal{M}_F$), then there exists a utility U on Y and a weighted set $\mathcal{P}^+$ of probabilities on S such that $C(\mathcal{P}^+)$ is regular and $C = C_{\mathcal{P}^+}^{S,Y,U}$. Moreover, U is unique up to affine transformations, and $C(\mathcal{P}^+)$ is unique in the sense that if $\mathscr{Q}^+$ represents C, and $C(\mathscr{Q}^+)$ is regular, then $C(\mathscr{Q}^+) = C(\mathcal{P}^+)$.*

Proof. Showing that $\succeq_{M,\mathcal{P}^+}^{S,Y,U}$ satisfies Axioms 1–7 is fairly straightforward; we leave details to the reader. Essentially the same proof works for $\mathcal{M}_B$ and $\mathcal{M}_F$. For the proof of the converse, we rely heavily on parts of the proof by Stoye [28]. Although Stoye considers $\mathcal{M}_F$, the arguments also work if we consider $\mathcal{M}_B$.

Stoye [28] shows that Axioms 1–7 imply that a menu-independent, revealed preference order $\succeq_C$ can be constructed based on the behavior of the choice function C on the set $\mathcal{M}_0$ of menus with nonpositive acts and utility frontier 0 (i.e., for every state, some act has a utility of 0). The preference order $\succeq_C$ can be thought of as the revealed preference order corresponding to the choice function C. Its definition is as follows:

$$f \succ_C g \Leftrightarrow \exists M \in \mathcal{M}_0 : f \in C(M), g \in M \backslash C(M),$$
$$f \sim_C g \Leftrightarrow \exists M \in \mathcal{M}_0 : f \in C(M), g \in C(M).$$

The arguments given by Stoye [28] to establish that $\succeq_C$ satisfies completeness, transitivity, nontriviality, monotonicity, mixture continuity, independence, INA, and ambiguity aversion apply verbatim to our setting here. We do not repeat Stoye's arguments here, but from here on we assume without further comment that $\succeq_C$ satisfies all these properties.

Theorem 1 can be completed by finding a MWER representation for $\succeq_C$. This follows from the following lemma. Let M_0 denote the set of all acts with nonpositive utilities.

Lemma 1. *If a preference order $\succeq$ on acts with nonpositive utilities satisfies completeness, transitivity, nontriviality, monotonicity, mixture continuity, independence, INA, and ambiguity aversion, then there exists a utility U on Y and a weighted set $\mathcal{P}^+$ of probabilities on S such that $C(\mathcal{P}^+)$ is regular and $\succeq = \succeq^{S,Y,U}_{M_0,\mathcal{P}^+}$. Moreover, U is unique up to affine transformations, and $C(\mathcal{P}^+)$ is unique in the sense that if $\mathscr{Q}^+$ represents $\succeq$, and $C(\mathscr{Q}^+)$ is regular, then $C(\mathscr{Q}^+) = C(\mathcal{P}^+)$.*

The proof of Lemma 1 is given in Appendix A. We use techniques in the spirit of those used by by Gilboa and Schmeidler [10] to represent (unweighted) MMEU. However, there are technical difficulties that arise from the fact that we do not have a key axiom that is satisfied by MMEU: *C-independence* (discussed below). The heart of the proof involves dealing with the lack of C-independence. With Lemma 1, the proof of Theorem 1 is complete. □

The axioms used in Theorem 1 can be adapted to describe choice functions over $\mathcal{M}_C$, the set of finitely generated convex menus. However, we do not know whether the equivalent of Theorem 1 holds if we restrict the domain of the choice function to be $\mathcal{M}_C$. However, in Appendix B we show that if we consider preference orders instead of a choice function, then there is a collection of axioms that provide a representation theorem for $\mathcal{M}_F$, $\mathcal{M}_B$, and $\mathcal{M}_C$. The result extends to $\mathcal{M}_C$ because the set of preference orders $\succeq_M$ with respect to the set $\mathcal{M}_C$ of all convex menus also determines the preference orders with respect to the set $\mathcal{M}_F$ of all finite menus. This is not the case when we take choice functions as the primitives. As Stoye [28, 30] points out, the main subtlety lies in the fact that the choice functions over convex menus can always return an interior point of the convex set, thus not providing observations of choice between the vertices of the set. Stoye believes that this is a technicality that can be overcome, so that Theorem 1 should hold even if we restrict the domain of the choice function to be $\mathcal{M}_C$. However, this conjecture has yet to be verified. In the case of preference orders, the preference orders with respect to the set $\mathcal{M}_B$ also determine the preference orders $\succeq_M$ with respect to the set $\mathcal{M}_C$, so once we have a proof for $\mathcal{M}_B$, it readily extends to $\mathcal{M}_F$ and $\mathcal{M}_C$.

As we observed, in general, we have ambiguity aversion (Axiom 7) for regret. *Betweenness* [5] is a stronger notion than ambiguity aversion, which states that if an decision maker is indifferent between two acts, then he must also be indifferent among all convex combinations of these acts. While betweenness does not hold for regret, Stoye [28] gives a weaker version that

	1 defect		10 defects	
	Payoff	Regret	Payoff	Regret
deliver	10,000	0	-10,000	10,000
$\frac{1}{2}$*deliver* + $\frac{1}{2}$*cancel*	5,000	5,000	-5,000	5,000
cancel	0	10,000	0	0
check	5,001	4,999	-4,999	4,999

Table 4: Payoffs and regrets for the quality-control problem, with *deliver* mixed with the constant act *cancel*.

does hold. A menu M has *state-independent outcome distributions* if the set $L(s) = \{y \in \Delta(Y) : \exists f \in M(f(s) = y(\}$ is the same for all states s.

Axiom 8 ([28])**.** *For all acts f, constant acts p, scalars $\lambda \in (0,1)$, and menus $M \supseteq \{p, f, \lambda f + (1-\lambda)p\}$ with state independent outcome distributions, if $p \notin C(M)$ and $f \notin C(M)$, then $\lambda f + (1-\lambda)p \notin C(M)$.*

The assumption that the menu has state-independent outcome distributions is critical in Axiom 8. Stoye [28] shows that Axioms 1–5 together with Axiom 8 characterize MER. Non-probabilistic regret (which we denote REG) can be viewed as a special case of MER, where $\mathcal{P}$ consists of all distributions. This means that it satisfies all the axioms that MER satisfies. As Stoye [29] shows, REG is characterized by Axioms 1–5 and one additional axiom, which he calls Symmetry. We omit the details here.

To see why Axiom 8 is needed, suppose that we change the payoffs in the quality-control problem so that *deliver* has the same maximum expected regret as *cancel* $(10,000)$. However, as seen in Table 4, $\frac{1}{2}$*deliver*+$\frac{1}{2}$*cancel* has lower maximum expected regret $(5,000)$ than *deliver* $(10,000)$, showing that the variant of Axiom 8 without the state-independent outcome distribution requirement does not hold.

Although Axiom 8 is sound for unweighted minimax expected regret, it is no longer sound once we add weights. For example, suppose that we modified the quality-control problem so that all states we care about have the same outcome distributions, as required by Axiom 8. Then the payoffs and regrets will be those shown in Table 5. Suppose that the weights on $\Pr_1$ and $\Pr_{10}$ are 1 and 0.5, respectively. Then *deliver* has the same maximum weighted expected regret as *cancel* $(10,000)$. However, $\frac{1}{2}$*deliver* + $\frac{1}{2}$*cancel* has lower maximum weighted expected regret $(7,500)$ than *deliver*, showing that Axiom 8 with weighted probabilities does not hold.

As mentioned in the introduction, Chateauneuf and Faro [4] axiomatize a weighted version of maxmin expected utility, when utilities are restricted

	1 defect		10 defects	
	Payoff	Regret	Payoff	Regret
deliver	10,000	0	-10,000	20,000
$\frac{1}{2}$*deliver* + $\frac{1}{2}$*cancel*	5,000	5,000	-5,000	15,000
cancel	0	10,000	0	10,000
*check*1	-5,000	15,000	5,000	5,000
*check*2	-10,000	20,000	10,000	0

Table 5: Payoffs and regrets for the quality-control problem, with state-independent outcome distributions.

to be nonnegative. The expected utilities are multiplied by the reciprocal of the weights, instead of the weights themselves. Preferences are then defined by using the maxmin expected utility rule with the weighted expected utilities. To obtain uniqueness of the representation, while we require (1) a measure with weight 1, (2) downward-closedness, (3) closedness, and (4) convexity of the sub-probability measures to get uniqueness, Chateauneuf and Faro [4] require slightly different properties. In particular, weights are represented by a function $\phi : \Delta(S) \to [0, 1]$, and Chateauneuf and Faro require that there be (1) a measure with weight 1, (2) upper semicontinuity of the function ϕ (i.e., the set $\{p \in \Delta : \phi(p) \geq \alpha\}$ is closed in the weak* topology, for all $\alpha \in [0, 1]$), and (3) quasi-concavity of ϕ (i.e., $\phi(\beta p_1 + (1 - \beta)p_2) \geq \min\{\phi(p_1), \phi(p_2)\}$ for all $\beta \in [0, 1]$). It is not hard to show that convexity of the sub-probability measures imply quasi-concavity of the weights, and closedness of the sub-probability measures implies weak* upper semicontinuity of the weights. However, the converse does not hold: weak* upper semicontinuity and quasi-concavity of the weights do not imply convexity of the sub-probability measures. Therefore, our conditions to obtain uniqueness of the representation are more stringent than those of Chateauneuf and Faro.

5 Characterizing likelihood updating for MWER

We next consider a more dynamic setting, where decision makers learn information. For simplicity, we assume that the information is always a subset E of the state space. If the decision maker is representing her uncertainty using a set $\mathcal{P}^+$ of weighted probability measures, then we would expect her to update $\mathcal{P}^+$ to some new set $\mathscr{Q}^+$ of weighted probability measures, and then apply MWER with uncertainty represented by$\mathscr{Q}^+$. In this section, we characterize what happens in the special case that the decision maker uses likelihood updating, so that $\mathscr{Q}^+ = (\mathcal{P}^+ \mid E)$.

For this characterization, we assume that the decision maker has a family of choice functions C_E indexed by the information E. Each choice function C_E satisfies Axioms 1–7, since the decision maker makes decisions after learning E using MWER. Somewhat surprisingly, all we need is one extra axiom for the characterization; we call this axiom MDC, for "menu-dependent dynamic consistency".

To explain the axiom, we need some notation. As usual, we take fEh to be the act that agrees with f on E and with h off of E; that is

$$fEh(s) = \begin{cases} f(s) & \text{if } s \in E \\ h(s) & \text{if } s \notin E. \end{cases}$$

In the quality-control problem, the act *check* can be thought of as $(\mathit{deliver})E(\mathit{cancel})$, where E is the set of states where there is only one defective item.

Roughly speaking, MDC says that you prefer f to g once you learn E if and only if, for all acts h, you also prefer fEh to gEh before you learn anything. This seems reasonable, since learning that the true state was in E is conceptually similar to knowing that none of your choices matter off of E.

To state MDC formally, we need to be careful about the menus involved. Let $MEh = \{fEh : f \in M\}$. We can identify unconditional preferences with preferences conditional on S; that is, we identify C with C_S. We also need to restrict the sets E to which MDC applies. Recall that conditioning using likelihood updating is undefined for an event such that $\overline{\mathcal{P}}^+(E) = 0$. That is, $\alpha_{\Pr} \Pr(E) = 0$ for all $\Pr \in \mathcal{P}$. As is commonly done, we capture the idea that conditioning on E is possible using the notion of a *non-null* event.

Definition 1 (Null event)**.** *An event E is* null *if, for all $f, g \in \Delta(Y)^S$ and menus M with $fEg, g \in M$, we have $fEg \in C(M) \Leftrightarrow g \in C(M)$.*

Axiom 9 (MDC)**.** *Let MEg denote the menu $\{hEg : h \in M\}$. For all $M \subseteq L, f \in M$,*

$$f \in C_E(M) \Leftrightarrow \exists g (fEg \in C(MEg)).$$

Theorem 2. *For all Y, U, S, and $\mathcal{P}^+$, the choice function $C^{S,Y,U}_{\mathcal{P}^+|E}$ for events E such that $\overline{\mathcal{P}}^+(E) > 0$ satisfies Axioms 1–7 and Axiom 9. Conversely, if a choice function C_E on the acts in $\Delta(Y)^S$ satisfies Axioms 1–7 and Axiom 9,* *then there exists a utility U on Y and a weighted set $\mathcal{P}^+$ of probabilities on S such that $C(\mathcal{P}^+)$ is regular, and for all non-null E, $C_E = C^{S,Y,U}_{\mathcal{P}^+|E}$. Moreover, U is unique up to affine transformations, and $C(\mathcal{P}^+)$ is unique in the sense that if $\mathscr{Q}^+$ represents C_E, and $C(\mathscr{Q}^+)$ is regular, then $C(\mathscr{Q}^+) = C(\mathcal{P}^+)$.*

Proof. Since $C = C_S$ satisfies Axioms 1–7, there must exist a weighted set $\mathcal{P}^+$ of probabilities on S and a utility function U such that $f \in C(M)$ iff $f \in C^{S,Y,U}(M)$. We now show that if E is non-null, then $\overline{\mathcal{P}}^+(E) > 0$, and $f \in C_E(M)$ iff $f \in C^{S,Y,u}_{\mathcal{P}^+|E}(M)$.

For the first part, it clearly is equivalent to show that if $\overline{\mathcal{P}}^+(E) = 0$, then E is null. So suppose that $\overline{\mathcal{P}}^+(E) = 0$. Then $\alpha_{\Pr} \Pr(E) = 0$ for all $\Pr \in \mathcal{P}$. This means that $\alpha_{\Pr} \Pr(s) = 0$ for all $\Pr \in \mathcal{P}$ and $s \in E$. Thus, for all acts f and g,

$$\begin{array}{rl} & \mathit{reg}_{M,\mathcal{P}^+}(fEg) \\ = & \sup_{\Pr\in\mathcal{P}} \left(\alpha_{\Pr} \sum_{s\in S} \Pr(s) \mathit{reg}_M(fEg, s)\right) \\ = & \sup_{\Pr\in\mathcal{P}} \left(\alpha_{\Pr} \left(\sum_{s\in E} \Pr(s) \mathit{reg}_M(f, s)\right) + \sum_{s\in E^c} \Pr(s) \mathit{reg}_M(g, s)\right) \\ = & \sup_{\Pr\in\mathcal{P}} \left(\alpha_{\Pr} \sum_{s\in S} \Pr(s) \mathit{reg}_M(g, s)\right) \\ = & \mathit{reg}_{M,\mathcal{P}^+}(g). \end{array}$$

Thus, $fEg \in C(M) \Leftrightarrow g \in C(M)$ for all acts f, g and menus M containing fEg and g, which means that E is null.

For the second part, we first show that if $\overline{\mathcal{P}}^+(E) > 0$, then for all $f, h \in M$, we have that

$$\mathit{reg}_{MEh,\mathcal{P}^+}(fEh) = \overline{\mathcal{P}}^+(E) \mathit{reg}_{M,\mathcal{P}^+|E}(f).$$

We proceed as follows:

$$\begin{array}{rl} & \mathit{reg}_{MEh,\mathcal{P}^+}(fEh) \\ = & \sup_{\Pr\in\mathcal{P}} \left(\alpha_{\Pr} \sum_{s\in S} \Pr(s) \mathit{reg}_{MEh}(fEH, s)\right) \\ = & \sup_{\Pr\in\mathcal{P}} \left(\alpha_{\Pr} \Pr(E) \sum_{s\in E} \Pr(s \mid E) \mathit{reg}_M(f, s) \ + \alpha_{\Pr} \sum_{s\in E^c} \Pr(s) \mathit{reg}_{\{h\}}(h, s)\right) \\ = & \sup_{\Pr\in\mathcal{P}} \left(\alpha_{\Pr} \Pr(E) \sum_{s\in E} \Pr(s|E) \mathit{reg}_M(s, f)\right) \\ = & \sup_{\Pr\in\mathcal{P}} \left(\overline{\mathcal{P}}^+(E) \alpha_{\Pr,E} \sum_{s\in E} \Pr(s|E) \mathit{reg}_M(f, s)\right) \\ & [\text{since } \alpha_{\Pr,E} = \sup_{\{\Pr'\in\mathcal{P}:\Pr'|E=\Pr|E\}} \frac{\alpha_{\Pr'} \Pr'(E)}{\overline{\mathcal{P}}^+(E)}] \\ = & \overline{\mathcal{P}}^+(E) \cdot \mathit{reg}_{M,\mathcal{P}^+|E}(f). \end{array}$$

Thus, for all $h \in M$,

$$\begin{aligned} & \mathit{reg}_{MEh,\mathcal{P}^+}(fEh) \le \mathit{reg}_{MEh,\mathcal{P}^+}(gEh) \\ \text{iff } & \overline{\mathcal{P}}^+(E) \cdot \mathit{reg}_{M,\mathcal{P}^+|E}(f) \le \overline{\mathcal{P}}^+(E) \cdot \mathit{reg}_{M,\mathcal{P}^+|E}(g) \\ \text{iff } & \mathit{reg}_{M,\mathcal{P}^+|E}(f) \le \mathit{reg}_{M,\mathcal{P}^+|E}(g). \end{aligned}$$

It follows that the choice function induced by $\mathcal{P}^+$ satisfies MDC. Moreover, if Axioms 1–7 and MDC hold, then for a weighted set $\mathcal{P}^+$ that represents C, we have

$$\begin{array}{ll} & f \in C_E(M) \\ \text{iff} & \text{for some } h, fEh \in C(MEh) \\ \text{iff} & \mathit{reg}_{M,\mathcal{P}^+|E}(f) \le \mathit{reg}_{M,\mathcal{P}^+|E}(g) \text{ for all } g, \end{array}$$

as desired.

Finally, the uniqueness of $C(\mathcal{P}^+)$ follows from Theorem 1, which says that C is already sufficient to guarantee the uniqueness of $C(\mathcal{P}^+)$. □

6 Conclusion

We proposed an alternative belief representation using *weighted sets of probabilities*, and described a natural approach to updating in such a situation and a natural approach to determining the weights. We also showed how weighted sets of probabilities can be combined with regret to obtain a decision rule, MWER, and provided an axiomatization that characterizes static and dynamic preferences induced by MWER.

One issue that must be dealt with when MWER is combined with likelihood updating is *dynamic inconsistency*. It is not hard to construct examples where a decision maker decides on a plan that says that if he learns E, he should perform act f, but then when he actually learns E, he performs f' instead. Such dynamic inconsistency arises with MMEU when using measure-by-measure updating as well. Siniscalchi [27] proposes an approach to dealing with dynamic consistency in the context of MMEU combined with measure-by-measure updating by using backward induction to decide which action to take. We believe that these ideas can be applied to MWER combined with likelihood updating as well. We hope to return to these issues in future work.

A Proof of Lemma 1

A.1 Defining a functional on utility acts

Stoye [28] also started his proof of a representation theorem for MER by reducing to a single preference order $\succeq_{M^*}$. He then noted that, the expected regret of an act f with respect to a probability Pr and menu M^* is just the negative of the expected utility of f. Thus, the worst-case expected regret of f with respect to a set $\mathcal{P}$ of probability measures is the negative of the worst-case expected utility of f with respect to $\mathcal{P}$. Thus, it sufficed for Stoye to show that $\succeq_{M^*}$ had an MMEU representation, which he did by showing that $\succeq_{M^*}$ satisfied Gilboa and Schmeidler's [10] axioms for MMEU, and then appealing to their representation theorem.

This argument does not quite work for us, because now $\succeq$ does not satisfy the C-independence axiom. (This is because our preference order is based on *weighted* regret, not regret.) However, we can get a representation theorem for weighted regret by using some of the techniques used by Gilboa and Schmeidler to get a representation theorem for MMEU, appropriately modified to deal with lack of C-independence. Specifically, like Gilboa and Schmeidler, we define a functional I on utility acts such that the preference order on utility acts is determined by their value according to I (see Lemma 3). Using I, we can then determine the weight of each probability in $\Delta(S)$, and prove the desired representation theorem.

By standard results, u represents $\succeq$ on constant acts, and $\succeq$ depends only on the utility achieved in each state (as opposed to the actual outcomes) of the acts. The space of all utility acts is the Banach space $\mathcal{B}$ of real-valued functions on S. Let $\mathcal{B}^-$ be the set of nonpositive functions in $\mathcal{B}$, where the function b is nonpositive if $b(s) \leq 0$ for all $s \in S$.

We now define a functional I on utility acts in $\mathcal{B}^-$ such that for all f, g with $b_f, b_g \in \mathcal{B}^-$, we have $I(b_f) \geq I(b_g)$ iff $f \succeq g$. Let

$$R_f = \{\alpha' : l^*_{\alpha'} \succeq f\}.$$

If $0^* \geq b \geq (-1)^*$, then f_b exists, and we define

$$I(b) = \inf(R_{f_b}).$$

For the remaining $b \in \mathcal{B}^-$, we extend I by homogeneity. Let $||b|| = |\min_{s \in S} b(s)|$. Note that if $b \in \mathcal{B}^-$, then $0^* \geq b/||b|| \geq (-1)^*$, so we define

$$I(b) = ||b|| I(b/||b||).$$

Lemma 2. *If $b_f \in \mathcal{B}^-$, then $f \sim l^*_{I(b_f)}$.*

Proof. Suppose that $b_f \in \mathcal{B}^-$ and, by way of contradiction, that $l^*_{I(b_f)} \prec f$. If $f \sim l^*_0$, then it must be the case that $I(b_f) = 0$, since $I(b_f) \leq 0$ by definition of inf, and $f \sim l^*_0 \succ l^*_\epsilon$ for all $\epsilon < 0$ by Lemma 9, so $I(b_f) > \epsilon$ for all $\epsilon < 0$. Therefore, $f \sim l^*_{I(b_f)}$. Otherwise, since $b_f \in \mathcal{B}^-$, by monotonicity, we must have $l^*_0 \succ f$, and thus $l^*_0 \succ f \succ l^*_{I(b_f)}$. By mixture continuity, there is some $q \in (0,1)$ such that $q \cdot l^*_0 + (1-q) \cdot l^*_{I(b_f)} \sim l_{(1-q)I(b_f)} \prec f$, contradicting the fact that $I(b)$ is the greatest lower bound of R_f.

If, on the other hand, $l^*_{I(b_f)} \succ f$, then $l^*_{I(b_f)} \succ f \succeq l^*_{\underline{c}}$ for some $\underline{c} \in \mathbb{R}$. If $f \sim l^*_{\underline{c}}$ then it must be the case that $I(b_f) = \underline{c}$. $I(b_f) \leq \underline{c}$ since $l^*_{\underline{c}} \succeq l^*_{\underline{c}}$, and $I(b_f) \geq \underline{c}$ since for all $c' < \underline{c}$, $l^*_{c'} \prec f \sim l^*_{\underline{c}}$.

Otherwise, $l^*_{I(b_f)} \succ f \succ l^*_{\underline{c}}$, and by mixture continuity, there is some $q \in (0,1)$ such that $q \cdot l^*_{I(b_f)} + (1-q) l^*_{\underline{c}} \succ f$. Since $qI(b_f) + (1-q)\underline{c} < I(b_f)$, this contradicts the fact that $I(b_f)$ is a lower bound of R_f. Therefore, it must be the case that $l^*_{I(b_f)} \sim f$. □

We can now show that I has the required property.

Lemma 3. *For all acts f, g such that $b_f, b_g \in \mathcal{B}^-$, $f \succeq g$ iff $I(b_f) \geq I(b_g)$.*

Proof. Suppose that $b_f, b_g \in \mathcal{B}^-$. By Lemma 2, $l^*_{I(b_f)} \sim f$ and $g \sim l^*_{I(b_g)}$. Thus, $f \succeq g$ iff $l^*_{I(b_f)} \succeq l^*_{I(b_g)}$, and by Lemma 9, $l^*_{I(b_f)} \succeq l^*_{I(b_g)}$ iff $I(b_f) \geq I(b_g)$. □

In order to invoke a standard separation result for Banach spaces, we extend the definition of I to the Banach space $\mathcal{B}$. We extend I to $\mathcal{B}$ by taking $I(b) = I(b^-)$ for $b \in \mathcal{B} - \mathcal{B}^-$, where for all $b \in \mathcal{B}$, b^- is defined as

$$b^-(s) = \begin{cases} b(s), \text{ if } b(s) \leq 0, \\ 0, \text{ if } b(s) > 0. \end{cases}$$

Clearly $b^- \in \mathcal{B}^-$ and $b = b^-$ if $b \in \mathcal{B}^-$.

We show that the axioms guarantee that I has a number of standard properties. Since we have artificially extended I to $\mathcal{B}$, our arguments require more cases than those in [10]. (We remark that such an "artificial" extension seem unavoidable in our setting.) Moreover, we must work harder to get the result that we want. We need different arguments from that for MMEU [10], since the preference order induced by MMEU satisfies C-independence, while our preference order does not.

Lemma 4. *(a) If $c \leq 0$, then $I(c^*) = c$.*

(b) I satisfies positive homogeneity: if $b \in \mathcal{B}$ and $c > 0$, then $I(cb) = cI(b)$.

(c) I is monotonic: if $b, b' \in \mathcal{B}$ and $b \geq b'$, then $I(b) \geq I(b')$.

(d) I is continuous: if $b, b_1, b_2, \ldots \in \mathcal{B}$, and $b_n \to b$, then $I(b_n) \to I(b)$.

(e) I is superadditive: if $b, b' \in \mathcal{B}$, then $I(b + b') \geq I(b) + I(b')$.

Proof. For part (a), If c is in the range of u, then it is immediate from the definition of I and Lemma 9 that $I(c^*) = c$. If c is not in the range of u, then since $[-1, 0]$ is a subset of the range of u, we must have $c < -1$, and by definition of I, we have $I(c^*) = |c|I(c^*/|c|) = c$.

For part (b), first suppose that $||b|| \leq 1$ and $b \in \mathcal{B}^-$ (i.e., $0^* \geq b \geq (-1)^*$). Then there exists an act f such that $b_f = b$. By Lemma 2, $f \sim l^*_{I(b)}$. We now need to consider the case that $c \leq 1$ and $c > 1$ separately. If $c \leq 1$, by Independence, $cf_b + (1-c)l_0^* \sim cl^*_{I(b)} + (1-c)l_0^*$. By Lemma 3, $I(b_{cf_b+(1-c)l_0^*}) = I(b_{cl^*_{I(b)}+(1-c)l_0^*})$. It is easy to check that $b_{cf_b+(1-c)l_0^*} = cb$, and $b_{cl^*_{I(b)}} + (1-c)l_0^* = cI(b)^*$. Thus, $I(cb) = I(cI(b)^*)$. By part (a), $I(cI(b)^*) = cI(b)$. Thus, $I(cb) = cI(b)$, as desired.

If $c > 1$, there are two subcases. If $||cb|| \leq 1$, since $1/c < 1$, by what we have just shown $I(b) = I(\frac{1}{c}(cb)) = \frac{1}{c}I(cb)$. Crossmultiplying, we have that $I(cb) = cI(b)$, as desired. And if $||cb|| > 1$, by definition, $I(cb) = ||cb||I(bc/||cb||) = c||b||I(b/||b||)$ (since $bc/||cb|| = b/||b||$). Since $||b|| \leq 1$, by what we have shows $I(b) = I(||b||(b/||b||) = ||b||I(b/||b||)$, so $I(b/||b||) = \frac{1}{||b||}I(b)$. Again, it follows that $I(cb) = cI(b)$.

Now suppose that $||b|| > 1$. Then $I(b) = ||b||I(b/||b||)$. Again, we have two subcases. If $||cb|| > 1$, then

$$I(cb) = ||cb||I(cb/||cb||) = c||b||I(b/||b||) = cI(b).$$

And if $||cb|| \leq 1$, by what we have shown for the case $||b|| \leq 1$,

$$I(b) = I(\frac{1}{c}(cb)) = \frac{1}{c}I(cb),$$

so again $I(cb) = cI(b)$.

For part (c), first note that if $b, b' \in \mathcal{B}^-$. If $||b|| \leq 1$ and $||b'|| \leq 1$, then the acts f_b and $f_{b'}$ exist. Moreover, since $b \geq b'$, we must have $(f_b(s))^* \succeq (f_{b'})^*(s)$ for all states $s \in S$. Thus, by Monotonicity, $f_b \succeq f_{b'}$. If either $||b|| > 1$ or $||b'|| > 1$, let $n = \max(||b||, ||b'||)$. Then $||b/n|| \leq 1$ and $||b'/n|| \leq$

1. Thus, $I(b/n) \geq I(b'/n)$, by what we have just shown. By part (b), $I(b) \geq I(b')$. Finally, if either $b \in \mathcal{B} - \mathcal{B}^-$ or $b' \in \mathcal{B} - \mathcal{B}^-$, note that if $b \geq b'$, then $b^- \geq (b')^-$. By definition, $I(b) = I(b^-)$ and $I(b') = I(b')^-$; moreover, $b^-, (b')^- \in \mathcal{B}^-$. Thus, by the argument above, $I(b) \geq I(b^-)$.

For part (d), note that if $b_n \to b$, then for all k, there exists n_k such that $b_n - (1/k)^* \leq b_n \leq b_n + (1/k)^*$ for all $n \geq n_k$. Moreover, by the monotonicity of I (part (c)), we have that $I(b - (1/k)^*) \leq I(b_n) \leq I(b + (1/k)^*)$. Thus, it suffices to show that $I(b - (1/k)^*) \to I(b)$ and that $I(b + (1/k)^*) \to I(b)$.

To show that $I(b - (1/k)^*) \to I(b)$, we must show that for all $\epsilon > 0$, there exists k such that $I(b - (1/k)^*) \geq I(b) - \epsilon$. By positive homogeneity (part (b)), we can assume without loss of generality that $||b - (1/2)^*|| \leq 1$ and that $||b|| \leq 1$. Fix $\epsilon > 0$. If $I(b - (1/2)^*) \geq I(b) - \epsilon$, then we are done. If not, then $I(b) > I(b) - \epsilon > I(b - (1/2)^*)$. Since $||b|| \leq 1$ and $||b - (1/2)^*|| \leq 1$, f_b and $f_{b-(1/2)^*}$ exist. Moreover, by Lemma 3, $f_b \succ f_{(I(b)-\epsilon)^*} \succ f_{b-(1/2)^*}$. By mixture continuity, for some $p \in (0,1)$, we have $pf_b + (1-p)f_{(b-(1/2)^*} \succ f_{(I(b)-\epsilon)^*}$. It is easy to check that $b_{pf_b+(1-p)f_{b-(1/2)^*}} = b - (1-p)(1/2)^*$. Thus, by Lemma 3, $f_{b-(1-p)(1/2)^*} \succeq f_{(I(b)-\epsilon)^*}$, and $I(b - (1-p)1/2)^*) > I(b) - \epsilon$. Choose k such that $1/k < (1-p)(1/2)$. Then

$$I(b - (1/k)^*) \geq I(b - (1-p)1/2)^*) > I(b) - \epsilon,$$

as desired.

The argument that $I(b + (1/k)^*) \to I(b)$ is similar and left to the reader.

For part (e), first suppose that $b, b' \in \mathcal{B}^-$. If $||b||, ||b^-|| \leq 1$, and $I(b), I(b') \neq 0$, consider $\frac{b}{-I(b)}$ and $\frac{b'}{-I(b')}$. Since $I(\frac{b}{-I(b)}) = I(\frac{b'}{-I(b')}) = -1$, it follows from Lemma 2 that $f_{\frac{b}{-I(b)}} \sim f_{\frac{b'}{-I(b')}}$. By ambiguity aversion, for all $p \in (0,1]$, $pf_{\frac{b}{-I(b)}} + (1-p)f_{\frac{b'}{-I(b')}} \succeq f_{\frac{b}{-I(b)}}$. Taking $p = I(b)/(I(b) + I(b'))$, we have that $(I(b)/(I(b) + I(b')))f_{b/I(b)} + (I(b')/(I(b) + I(b')))f_{b'/I(b')} \succeq f_{b/I(b)}$. Therefore, we have

$$I\left(\frac{-I(b)}{-I(b) - I(b')}\frac{b}{-I(b)} + \frac{-I(b')}{-I(b) - I(b')}\frac{b'}{-I(b')}\right) \geq I(\frac{b}{-I(b)}) = -1.$$

Simplifying, we have

$$I\left(\frac{-1}{I(b) + I(b')}b + \frac{-1}{I(b) + I(b')}b'\right) \geq -1,$$

which, together with positive homogeneity of I (part (b)), implies $I(b+b') \geq I(b) + I(b')$, as required.

If $b, b^- \in \mathcal{B}^-$ and either $||b|| > 1$ or $||b'|| > 1$, and both $I(b) \neq 0$ and $I(b') \neq 0$, then the result easily follows by positive homogeneity (property (b)).

If $b, b^- \in \mathcal{B}-$ and either $I(b) = 0$ or $I(b') = 0$, let $b_n = b - \frac{1}{n}^*$ and $b'_n = b' - \frac{1}{n}^*$. Clearly $||b_n|| > 0$, $||b'_n|| > 0$, $b_n \to b$, and $b'_n \to b'_n$. By our argument above, $I(b_n + b'_n) \geq I(b_n) + I(b'_n)$ for all $n \geq 1$. The result now follows from continuity.

Finally, if either $b \in \mathcal{B} - \mathcal{B}^-$ or $b' \in \mathcal{B} - \mathcal{B}^-$, observe that

$$(b + b')^-(s) \begin{cases} = b^-(s) + b'^-(s), & \text{if } b(s) \leq 0, b'(s) \leq 0 \\ = b^-(s) + b'^-(s), & \text{if } b(s) \geq 0, b'(s) \geq 0 \\ \geq b^-(s) + b'^-(s), & \text{if } b(s) > 0, b'(s) \leq 0 \\ \geq b^-(s) + b'^-(s), & \text{if } b(s) \leq 0, b'(s) > 0. \end{cases}$$

Therefore, $(b+b')^- \geq b^- + b'^-$. Thus, $I(b+b') = I((b+b')^-) \geq I(b^- + b'^-)$ by the monotonicity of I, and $I(b^- + b'^-) \geq I(b^-) + I(b'^-)$ by superadditivity of I on $\mathcal{B}^-$. Therefore, $I(b + b') \geq I(b) + I(b')$. □

A.2 Defining the weights

In this section, we use I to define a weight $\alpha_{\Pr}$ for each probability $\Pr \in \Delta(S)$. The heart of the proof involves showing that the resulting set $\mathcal{P}^+$ so determined gives us the desired representation.

Given a set $\mathcal{P}^+$ of weighted probability measures, for $b \in \mathcal{B}^-$, define

$$NWREG(b) = \inf_{\Pr \in \mathcal{P}} \alpha_{\Pr}(\sum_{s \in S} b(s) \Pr(s)).$$

Note that $NWREG$ is the negative of the weighted regret when the menu is $\mathcal{B}^-$. Define

$$NREG(b) = \inf_{\Pr \in \mathcal{P}} \sum_{s \in S} b(s) \Pr(s).$$

and

$$NREG_{\Pr}(b) = \sum_{s \in S} b(s) \Pr(s) = E_{\Pr} b.$$

For each probability $\Pr \in \Delta(S)$, define

$$\alpha_{\Pr} = \sup\{\alpha \in \mathbb{R} : \alpha NREG_{\Pr}(b) \geq I(b) \text{ for all } b \in \mathcal{B}^-\}. \qquad (1)$$

Note that $\alpha_{\Pr} \geq 0$ for all distributions $\Pr \in \Delta(S)$, since $0 \geq I(b)$ for $b \in \mathcal{B}^-$ (by monotonicity); and $\alpha_{\Pr} \leq 1$, since $NREG_{\Pr}((-1)^*) = I((-1)^*) = -1$

for all distributions Pr. Thus, $\alpha_{\Pr} \in [0,1]$. Moreover, it is immediate from the definition of $\alpha_{\Pr}$ that $\alpha_{\Pr} NREG_{\Pr}(b) \geq I(b)$ for all $b \in \mathcal{B}^-$. The next lemma shows that there exists a probability Pr where we have equality.

Lemma 5. *(a) For some distribution* Pr*, we have* $\alpha_{\Pr} = 1$.

(b) For all $b \in \mathcal{B}^-$*, there exists* Pr *such that* $\alpha_{\Pr} NREG_{\Pr}(b) = I(b)$.

Proof. The proofs of both part (a) and (b) use a standard separation result: If U is an open convex subset of $\mathcal{B}$, and $b \notin U$, then there is a linear functional λ that separates U from b, that is, $\lambda(b') > \lambda(b)$ for all $b' \in U$. We proceed as follows

For part (a), we must show that for some Pr, for all $b \in \mathcal{B}^-$, $NREG_{\Pr}(b) \geq I(b)$. Since $NREG_{\Pr}(b) = E_{\Pr} b$, it suffices to show that $E_{\Pr}(b) \geq I(b)$ for all $b \in \mathcal{B}^-$.

Let $U = \{b' \in \mathcal{B} : I(b') > -1\}$. U is open (by continuity of I), and convex (by positive homogeneity and superadditivity of I), and $(-1)^* \notin U$. Thus, there exists a linear functional λ such that $\lambda(b') > \lambda((-1)^*)$ for $b' \in U$. We want to show that λ is a positive linear functional, that is, that $\lambda(b) \geq 0$ if $b \geq 0^*$. Since $0^* \in U$, and $\lambda(0^*) = 0$, it follows that $\lambda((-1)^*) < 0$. Since λ is linear, we can assume without loss of generality that $\lambda((-1)^*) = -1$. Thus, for all $b' \in \mathcal{B}^-$, $I(b') > -1$ implies $\lambda(b') > -1$. Suppose that $c > 0$ and $b' \geq 0^*$. From the definition of I, it follows that $I(cb') = I(0^*) = 0 > -1$. So $c\lambda(b') = \lambda(cb') > -1$, so $\lambda(b') > -1/c$. (The fact that $I(cb') = I(0^*)$ follows from the definition of I on elements in $\mathcal{B} - \mathcal{B}^-$.) Since this is true for all $c > 0$, it must be the case that $\lambda(b') \geq 0$. Thus, λ is a positive functional.

Define the probability distribution Pr on S by taking $\Pr(s) = \lambda(1_s)$. To see that Pr is indeed a probability distribution, note that since $1_s \geq 0$ and λ is positive, we must have $\lambda(1_s) \geq 0$. Moreover, $\sum_{s \in S} \Pr(s) = \lambda(1^*) = 1$. In addition, for all $b' \in \mathcal{B}$, we have

$$\lambda(b') = \sum_{s \in S} \lambda(1_s) b'(s) = \sum_{s \in S} \Pr(s) b'(s) = E_{\Pr}(b').$$

Next note that, for $b \in \mathcal{B}^-$,

$$\text{for all } c < 0, \text{ if } I(b) > c, \text{ then } \lambda(b) > c. \tag{2}$$

For if $I(b) > c$, then $I(b/|c|) > -1$ by positive homogeneity, so $\lambda(b/|c|) > -1$ and $\lambda(b) > c$. The result now follows. For if $b \in \mathcal{B}^-$, then $I(b) \leq I(0^*) = 0$ by monotonicity. Thus, if $c < I(b)$, then $c < 0$, so, by (2), $\lambda(b) > c$. Since $\lambda(b) > c$ whenever $I(b) > c$, it follows that $E_{\Pr}(b) = \lambda(b) \geq I(b)$, as desired.

The proof of part (b) is similar to that of part (a). We want to show that, given $b \in \mathcal{B}^-$, there exists Pr such that $\alpha_{\Pr} NREG_{\Pr}(b) = I(b)$. First suppose that $||b|| \leq 1$. If $I(b) = 0$, then there must exist some s such that $b(s) = 0$, for otherwise there exists $c < 0$ such that $b \leq c^*$, so $I(b) \leq c$. If $b(s) = 0$, let $\Pr_s$ be such that $\Pr_s(s) = 1$. Then $NREG_{\Pr_s}(b) = 0$, so (b) holds in this case.

If $||b|| \leq 1$ and $I(b) < 0$, let $U = \{b' : I(b') > I(b)\}$. Again, U is open and convex, and $b \notin U$, so there exists a linear functional λ such that $\lambda(b') > \lambda(b)$ for $b' \in U$. Since $0^* \in U$ and $\lambda(0^*) = 0$, we must have $\lambda(b) < 0$. Since $(-1)^* \leq b$, $(-1)^*$ is not in U, and therefore we also have $\lambda((-1)^*) < 0$. Thus, we can assume without loss of generality that $\lambda((-1)^*) = -1$, and hence $\lambda((1)^*) = 1$. The same argument as above shows that λ is positive: for all $c > 0$ and $b' \geq 0^*$, $I(cb') = 0$ as before. Since $I(b) < 0$, it follows that $I(cb') > I(b)$, so $cb' \in U$ and $\lambda(cb') > \lambda(b) \geq \lambda((-1)^*) = -1$. Thus, as before, for all $c > 0$, $b' \geq 0^*$, $\lambda(b') > \frac{-1}{c}$, so λ is a positive functional.

Therefore, λ determines a probability distribution Pr such that, for all $b' \in \mathcal{B}^-$, we have $\lambda(b') = E_{\Pr}(b')$. This, of course, will turn out to be the desired distribution. To show this, we need to show that $\alpha_{\Pr} = I(b)/NREG_{\Pr}(b)$. Clearly $\alpha_{\Pr} \leq I(b)/NREG_{\Pr}(b)$, since if $\alpha > I(b)/NREG_{\Pr}(b)$, then $\alpha NREG_{\Pr}(b) < I(b)$ (since $NREG_{\Pr}(b) = \lambda(b) < 0$). To show that $\alpha_{\Pr} \geq I(b)/NREG_{\Pr} b$, we must show that $(I(b)/NREG_{\Pr}(b))NREG_{\Pr}(b') \geq I(b')$ for all $b' \in \mathcal{B}^-$. Equivalently, we must show that $I(b)\lambda(b')/\lambda(b) \geq I(b')$ for all $b' \in \mathcal{B}^-$.

Essentially the same argument used to prove (2) also shows

$$\text{for all } c > 0, \text{ if } I(b') > cI(b), \text{ then } \lambda(b') > c\lambda(b).$$

In particular, if $I(b') > cI(b)$, then by positive homogeneity, $\frac{I(b')}{c} > I(b)$, so $\frac{b'}{c} \in U$, and $\lambda(\frac{b'}{c}) > \lambda(b)$ and hence $\lambda(b') > c\lambda(b)$.

Thus, if $I(b')/(-I(b)) > c$ and $c < 0$, then $I(b') > -cI(b)$, and hence $\lambda(b')/(-\lambda(b)) > c$. It follows that $\lambda(b')/(-\lambda(b)) \geq I(b')/(-I(b))$ for all $b' \in \mathcal{B}^-$. Thus, $I(b)\lambda(b')/\lambda(b) \geq I(b')$ for all $b' \in \mathcal{B}^-$, as required.

Finally, if $||b|| > 1$, let $b' = b/||b||$. By the argument above, there exists a probability measure Pr such that $\alpha_{\Pr} NREG_{\Pr}(b/||b||) = I(b/||b||)$. Since $NREG_{\Pr}(b/||b||) = NREG_{\Pr}(b)/||b||$, and $I(b/||b||) = I(b)/||b||$, we must have that $\alpha_{\Pr} NREG_{\Pr}(b) = I(b)$. □

We can now complete the proof of Lemma 1. By Lemma 5 and the definition of $\alpha_{\Pr}$, for all $b \in \mathcal{B}^-$,

$$\begin{aligned} I(b) &= \inf_{\Pr \in \Delta(S)} \alpha_{\Pr} NREG(b) \\ &= \inf_{\Pr \in \Delta(S)} \left(\alpha_{\Pr} \sum_{s \in S} b(s) \Pr(s) \right) \\ &= \sup_{\Pr \in \Delta(S)} \left(-\alpha_{\Pr} \sum_{s \in S} b(s) \Pr(s) \right). \end{aligned} \tag{3}$$

Recall that, by Lemma 3, for all acts f, g such that $b_f, b_g \in \mathcal{B}^-$, $f \succeq g$ iff $I(b_f) \geq I(b_g)$. Thus, $f \succeq g$ iff

$$\sup_{\Pr \in \Delta(S)} \left(-\alpha_{\Pr} \sum_{s \in S} u(f(s)) \Pr(s) \right) \leq \sup_{\Pr \in \Delta(S)} \left(-\alpha_{\Pr} \sum_{s \in S} u(g(s)) \Pr(s) \right).$$

Note that, for $f \in M^* = \mathcal{B}^-$, we have $reg_{M^*,\Pr}(f) = \sup(-u(f(s)) \Pr(s)$, since 0^* dominates all acts in M^*. Thus, $\succeq = \succeq^{S,Y,U}_{M^*,\mathcal{P}^+}$, where $\mathcal{P}^+ = \{(\Pr, \alpha_{\Pr} : \Pr \in \Delta(S)\}$.

We have already observed that U is unique up to affine transformations, so it remains to show that $\mathcal{P}^+$ is maximal. This follows from the definition of $\alpha_{\Pr}$. If $\succeq_M = \succeq^{S,Y,U}_{M,(\mathcal{P}')^+}$, and $(\alpha', \Pr) \in (\mathcal{P}')^+$, then we claim that $\alpha' \in \{\alpha \in \mathbb{R} : \alpha NREG_{\Pr}(b) \geq I(b) \text{ for all } b \in \mathcal{B}^-\}$. If not, there would be some $b \in \mathcal{B}^-$ with $||b|| \leq \frac{1}{2}$, such that $\alpha' NREG_{\Pr}(b) < I(b)$, which, by the definition of $\prec^{S,Y,U}_{M^*,(\mathcal{P}')^*}$, means that $l^*_{-1} \prec^{S,Y,U}_{M^*,(\mathcal{P}')^+} f_b \prec^{S,Y,U}_{M^*,(\mathcal{P}')^+} l^*_{I(b)}$. Recall that $I(b_f) = \inf\{\gamma : l^*_\gamma \succeq_{M^*} f\}$. Moreover, since $\prec^{S,Y,U}_{M^*,(\mathcal{P}')^+}$ satisfies mixture continuity, there exists some $p \in (0,1)$ such that $f_b \prec^{S,Y,U}_{M^*,(\mathcal{P}')^+} pl^*_{-1} + (1-p) l^*_{I(b)} \prec^{S,Y,U}_{M^*,(\mathcal{P}')^+} \prec^{S,Y,U}_{M^*,(\mathcal{P}')^+} l^*_{I(b)}$. This contradicts the definition of $I(b)$. Therefore, $\alpha' \in \{\alpha \in \mathbb{R} : \alpha NREG_{\Pr}(b) \geq I(b) \text{ for all } b \in \mathcal{B}^-\}$, and hence $\alpha' \leq \alpha_{\Pr}$.

A.3 Uniqueness of Representation

In this section, we show that the canonical set of weighted probabilities we constructed, when viewed as a set of subnormal probability measures, is regular and includes at least one proper probability measure. Moreover, this set of sub-probability measures is the only regular set that induces the preference order $\succeq$ on nonpositive acts. Our uniqueness result is analogous to the uniqueness results of Gilboa and Schmeidler [11], who show that the convex, closed, and non-empty set of probability measures in their representation theorem for MMEU is unique. The argument is based on two lemmas: Lemma 6 says that the canonical set of sub-probability measures is regular; and Lemma 7 says that a set of sub-probability measures representing $\succeq$ over nonpositive acts that is regular and contains at least one proper probability measure is unique. The proof of this second lemma, like the proof of uniqueness in Gilboa and Schmeidler [11], uses a separating hyperplane theorem to show the existence of acts on which two different representations must 'disagree'. However, a slightly different argument is required in our case, since our acts must have utilities corresponding to nonpositive vectors in $\mathbb{R}^{|S|}$.

Lemma 6. *Let $\mathcal{P}^+$ be the canonical set of weighted probability measures representing $\succeq$. The set $C(\mathcal{P}^+)$ of sub-probability measures is regular.*

Proof. It is useful to note that, by definition, $\mathbf{p} \in C(\mathcal{P}^+)$ if and only if

$$E_{\mathbf{p}}(b) \geq I(b) \text{ for all } b \in \mathcal{B}^-$$

(where expectation with respect to a subnormal probability measure is defined in the obvious way).

Recall that a set is regular if it is convex, closed, and downward-closed. We first show that $C(\mathcal{P}^+)$ is downward-closed. Suppose that $\mathbf{p} \in C(\mathcal{P}^+)$ and $\mathbf{q} \leq \mathbf{p}$ (i.e., $\mathbf{q}(s) \leq \alpha \Pr(s)$ for all $s \in S$. Since $\mathbf{p} \in C(\mathcal{P}^+)$, $E_{\mathbf{p}}(b) \geq I(b)$ for all $b \in \mathcal{B}^-$. Since $\mathbf{q} \leq \mathbf{p}$ and, if $b \in \mathcal{B}^-$, we have $b \leq 0^*$, it follows that $E_{\mathbf{q}}(b) \geq E_{\mathbf{p}}(b) \geq I(b)$ for all $b \in \mathcal{B}^-$, and thus $\mathbf{q} \in C(\mathcal{P}^+)$.

To see that $C(\mathcal{P}^+)$ is closed, let $\mathbf{p} = \lim_{n\to\infty} \mathbf{p}_n$, where each $\mathbf{p}_n \in C(\mathcal{P}^+)$. Since $\mathbf{p}_n \in C(\mathcal{P}^+)$ it must be the case that $E_{\mathbf{p}_n}(b) \geq I(b)$ for all $b \in \mathcal{B}^-$. By the continuity of expectation, it follows that $E_{\mathbf{p}}(b) \geq I(b)$ for all $b \in \mathcal{B}^-$. Thus, $\mathbf{p} \in C(\mathcal{P}^+)$.

To show that $C(\mathcal{P}^+)$ is convex, suppose that $\mathbf{p}, \mathbf{q} \in C(\mathcal{P}^+)$. Then $E_{\mathbf{p}}(b) \geq I(b)$ and $E_{\mathbf{q}}(b) \geq I(b)$ for all $b \in \mathcal{B}^-$. It easily follows that for all $a \in (0,1)$, $E_{a\mathbf{p}+(1-a)\mathbf{q}}(b) \geq I(b)$ for all $b \in \mathcal{B}^-$. Thus, $a\mathbf{p} + (1-a)\mathbf{q} \in C(\mathcal{P}^+)$. □

Lemma 7. *A set of sub-probability measures representing $\succeq$ over nonpositive acts that is regular, and has at least one proper probability measure is unique.*

Proof. Suppose for contradiction that there exists two regular sets of subnormal probability distributions, C_1 and C_2, that represent $\succeq$ and have at least one proper probability measure.

First, without loss of generality, let $\mathbf{q} \in C_2 \backslash C_1$. We actually look at an extension of C_1 that is downward-closed in each component to $-\infty$. Let $\overline{C}_1 = \{\mathbf{p} \in \mathbb{R}^{|S|} : \mathbf{p} \leq \mathbf{p}'\}$. Note an element $\mathbf{p}$ of $\overline{C}_1$ may not be subnormal probability measures; we do not require that $\mathbf{p}(s) \geq 0$ for all $s \in S$. Since $\overline{C}_1$ and $\{\mathbf{q}\}$ are closed, convex, and disjoint, and $\{\mathbf{q}\}$ is compact, the separating hyperplane theorem [23] says that there exists $\theta \in \mathbb{R}^{|S|}$ and $c \in \mathbb{R}$ such that

$$\theta \cdot \mathbf{p} > c \text{ for all } \mathbf{p} \in \overline{C}_1, \text{ and } \theta \cdot \mathbf{q} < c. \tag{4}$$

By scaling c appropriately, we can assume that $|\theta(s)| \leq 1$ for all $s \in S$. Now we argue that it must be the case that $\theta(s) \leq 0$ for all $s \in S$ (so that θ corresponds to the utility profile of some act with nonpositive utilities). Suppose that $\theta(s') > 0$ for some $s' \in S$. By (4), $\theta \cdot \mathbf{p} > c$ for all $\mathbf{p} \in \overline{C}_1$. However, consider $\mathbf{p}^* \in \overline{C}_1$ defined by

$$\mathbf{p}^*(s) = \begin{cases} 0, \text{ if } s \neq s' \\ \frac{-|c|}{\theta(s)}, \text{ if } s = s'. \end{cases}$$

Clearly, $\theta \cdot \mathbf{p}^* \leq c$, contradicting (4). Thus it must be the case that $\theta(s) \leq 0$ for all $s \in S$.

Consider the θ given by the separating hyperplane theorem, and let f be an act such that $u \circ f = \theta$. By continuity, $f \sim l_d^*$ for some constant act l_d^*. Since C_1 and C_2 both represent $\succeq$, and C_1 and C_2 both contain a proper probability measure,

$$\min_{\mathbf{p} \in C_1} \mathbf{p} \cdot (u \circ f) = \min_{\mathbf{p} \in C_1} \mathbf{p} \cdot (u \circ l_d^*) = d = \min_{\mathbf{p} \in C_2} \mathbf{p} \cdot (u \circ f).$$

However, by (4),

$$\min_{\mathbf{p} \in C_1} \mathbf{p} \cdot (u \circ f) > c > \min_{\mathbf{p} \in C_2} \mathbf{p} \cdot (u \circ f),$$

which is a contradiction. □

B An Axiomatic characterization of MWER with Preference Relations

We consider an axiomatization based on primitive preference orders $\succeq_M$ indexed by menus. (Because regret is menu-dependent, we cannot consider a single preference order $\succeq$.) As Stoye [29] points out, one disadvantage of considering such preference orders is that they are not observable. For example, suppose that $f_1 \succ_M f_2 \succ_M f_3$. By presenting the menu $M = \{f_1, f_2, f_3\}$ to the decision maker, we can observe that he prefers f_1. But there is no way to observe that $f_2 \succ_M f_3$. The traditional approach (seeing which of f_2 and f_3 the decision maker prefers when presented with the menu $M' = \{f_2, f_3\}$) will not work, because the decision maker's preferences with menu M' may be different from those with menu M. (Bleichrodt [3] studies a similar problem.)

In Section 4, we provided a characterization of MWER with choice functions as the primitives. Despite the fact that a regret-based preference order is not observable, an axiomatization using menu-dependent preference orders allows us to compare the axioms for weighted regret to those for other decision rules.

We state the axioms in a way that lets us clearly distinguish the axioms for SEU, MMEU, MER, and MWER. The axioms are universally quantified over acts f, g, and h, menus M and M', and $p \in (0, 1)$. We assume that $f, g \in M$ when we write $f \succeq_M g$.[7] We use l^* to denote a constant act that maps all states to l.

Axiom 10. *(Transitivity)* $f \succeq_M g \succeq_M h \Rightarrow f \succeq_M h$.

Axiom 11. *(Completeness)* $f \succeq_M g$ *or* $g \succeq_M f$.

Let $\mathcal{M}_B$ denote the set of all menus that are bounded above; that is, $\mathcal{M}_B = \{M : \sup_{g \in M} u(g(s)) \text{is finite}\}$.

Axiom 12. *(Nontriviality)* $f \succ_M g$ *for some acts* f *and* g *and menu* $M \in \mathcal{M}_B$.

Up to now, we have taken the set of menus to be $\mathcal{M}_B$. This assumption is necessary (and sufficient) for regret to be well defined. Later, we use Axiom 12 in the context of different classes of menus. In particular, we are interested in the set of finite menus and the set of *finitely generated* convex menus, that is, the menus M such such that there is a finite set A_M of acts such that M consists of all the convex combinations of acts in A_M. We denote these sets $\mathcal{M}_F$ and $\mathcal{M}_C$, respectively. When we use Axiom 12 in

[7] Stoye [29] assumed that menus were convex, so that if $f, g \in M$, then so is $pf + (1-p)g$. We do not make this assumption, although our results would still hold if we did (with the axioms slightly modified to ensure that menus are convex). While it may seem reasonable to think that, if f and g are feasible for an decision maker, then so is $pf + (1-p)g$, this not always the case. For example, it may be difficult for the decision maker to randomize, or it may be infeasible for the decision maker to randomize with probability p for some choices of p (e.g., for p irrational).

such contexts, $\mathcal{M}_B$ in Axiom 12 is understood to be replaced by $\mathcal{M}_C$ and $\mathcal{M}_F$, respectively. Observe that $\mathcal{M}_C \subseteq \mathcal{M}_B$ and $\mathcal{M}_F \subseteq \mathcal{M}_B$.

Axiom 13. *(Monotonicity) If* $(f(s))^* \succeq_{\{(f(s))^*,(g(s))^*\}} (g(s))^*$ *for all* $s \in S$, *then* $f \succeq_M g$.

Axiom 14. *(Mixture Continuity) If* $f \succ_M g \succ_M h$, *then there exist* $q, r \in (0,1)$ *such that*

$$qf + (1-q)h \succ_{M\cup\{qf+(1-q)h\}} g \succ_{M\cup\{rf+(1-r)h\}} rf + (1-r)h.$$

Menu-independent versions of Axioms 10–14 are standard (for example, (menu-independent versions of) these axioms are in [11]). Clearly (menu-independent versions of) Axioms 10, 11, 13, and 14 hold for MMEU, and SEU; Axiom 12 is assumed in all the standard axiomatizations, and is used to get a unique representation.

Axiom 15. *(Ambiguity Aversion)*

$$f \sim_M g \Rightarrow pf + (1-p)g \succeq_{M\cup\{pf+(1-p)g\}} g.$$

Ambiguity aversion says that the decision maker weakly prefers to hedge her bets. It also holds for MMEU, MER, and SEU, and is assumed in the axiomatizations for MMEU and MER. It is not assumed for the axiomatization of SEU, since it follows from the Independence axiom, discussed next. Independence also holds for MWER, provided that we are careful about the menus involved. Given a menu M and an act h, let $pM + (1-p)h$ be the menu $\{pf + (1-p)h : p \in M\}$.

Axiom 16. *(Independence)*

$$f \succeq_M g \text{ iff } pf + (1-p)h \succeq_{pM+(1-p)h} pg + (1-p)h.$$

Independence holds in a strong sense for SEU, since we can ignore the menus. The menu-independent version of Independence is easily seen to imply ambiguity aversion. Independence does not hold for MMEU.

Although we have menu independence for SEU and MMEU, we do not have it for MER or MWER. The following two axioms are weakened versions of menu independence that do hold for MER and MWER.

Axiom 17. *(Menu independence for constant acts) If* l^* *and* $(l')^*$ *are constant acts, then* $l^* \succeq_M (l')^*$ *iff* $l^* \succeq_{M'} (l')^*$.

In light of this axiom, when comparing constant acts, we omit the menu.

An act h is *never strictly optimal relative to* M if, for all states $s \in S$, there is some $f \in M$ such that $(f(s))^* \succeq (h(s))^*$.

Axiom 18. *(Independence of Never Strictly Optimal Alternatives (INA)) If every act in M' is never strictly optimal relative to M, then $f \succeq_M g$ iff $f \succeq_{M \cup M'} g$.*

Theorem 3. *For all Y, U, S, and $\mathcal{P}^+$, the family of preference orders $\succeq^{S,Y,U}_{M,\mathcal{P}^+}$ for $M \in \mathcal{M}_B$ (resp., $\mathcal{M}_F$, $\mathcal{M}_C$) satisfies Axioms 10–18. Conversely, if a family of preference orders $\succeq_M$ on the acts in $\Delta(Y)^S$ for $M \in \mathcal{M}_B$ (resp., $\mathcal{M}_F$, $\mathcal{M}_C$) satisfies Axioms 10–18, then there exist a utility U on Y and a weighted set $\mathcal{P}^+$ of probabilities on S such that $C(\mathcal{P}^+)$ is regular and $\succeq_M = \succeq^{S,Y,U}_{M,\mathcal{P}^+}$ for all $M \in \mathcal{M}_B$ (resp., $\mathcal{M}_C$, $\mathcal{M}_F$). Moreover, U is unique up to affine transformations, and $C(\mathcal{P}^+)$ is unique in the sense that if $\mathcal{Q}^+$ represents $\succeq_M$, and $C(\mathcal{Q}^+)$ is regular, then $C(\mathcal{Q}^+) = C(\mathcal{P}^+)$.*

Proof. Showing that $\succeq^{S,Y,U}_{M,\mathcal{P}^+}$ satisfies Axioms 10–18 is fairly straightforward; we leave details to the reader. Essentially the same proof works for $\mathcal{M}_B$, $\mathcal{M}_C$, and $\mathcal{M}_F$. The proof of the converse is quite nontrivial, although it follows the lines of the proof of other representation theorems. We start by considering $\mathcal{M}_B$.

Using standard techniques, we can show that the axioms guarantee the existence of a utility function U on prizes that can be extended to lotteries in the obvious way, so that $l^* \succeq (l')^*$ iff $U(l) \geq U(l')$. We then use techniques of Stoye [29] to show that it suffices to get a representation theorem for a single menu, rather than all menus: the menu consisting of all acts f such that $U(f(s)) \leq 0$ for all states $s \in S$. This allows us to use techniques in the spirit of those used by by Gilboa and Schmeidler [10] to represent (unweighted) MMEU.

We show here that if a family of menu-dependent preferences $\succeq_M$ satisfies Axioms 10–18, then $\succeq_M$ can be represented as minimizing expected regret with respect to a set of weighted probabilities and a utility function. Since the proof is somewhat lengthy and complicated, we split it into several steps, each in a separate subsection.

Simplifying the Problem. Our proof starts in much the same way as the proof by Stoye [29] of a representation theorem for regret. Lemma 8 guarantees the existence of a utility function U on prizes that can be extended to lotteries in the obvious way, so that $l^* \succeq (l')^*$ iff $U(l) \geq U(l')$. In other words, preferences over all constant acts are represented by the maximization of U on the corresponding lotteries that the constant acts map to. Lemma 8 is a consequence of standard results. Our menus are arbitrary sets of acts, as opposed to convex hulls of a finite number of acts in [29]; Lemma 10 shows that Stoye's technique can be adapted to work when menus are arbitrary sets of acts. Finally, following Stoye [29], we reduce the proof of existence of a minimax weighted regret representation for the family $\succeq_M$ to the proof of existence of a minimax weighted regret representation for a single menu-independent preference order $\succeq$ (Lemma 11).

Lemma 8. *If Axioms 1-3, 5, 7, and 8 hold, then there exists a nonconstant function $U : X \to \mathbb{R}$, unique up to positive affine transformations, such that for all constant acts l^* and $(l')^*$ and menus M,*

$$l^* \succeq_M (l')^* \Leftrightarrow \sum_{\{y:\, l^*(y)>0\}} l(y)U(y) \geq \sum_{\{y:\, l'(y)>0\}} l'(y)U(y).$$

Proof. By menu independence for constant acts, the family of preferences $\succeq_M$ all agree when restricted to constant acts. The lemma then follows from standard results (see, e.g., [17]), since menu-independence for constant acts, combined with independence, gives the standard independence (substitution) axiom from expected utility theory. □

As is commonly done, given U, we define $u(l) = \sum_{\{y:\, l(y)>0\}} l(y)U(y)$. Thus, $u(l)$ is the expected utility of lottery l. We extend u to constant acts by taking $u(l^*) = u(l)$. Thus, Lemma 8 says that, for all menus M, $l^* \succeq (l')^*$ iff $u(l^*) \geq u((l')^*)$. If c is the utility of some lottery, let l_c^* be a constant lottery that $u(l_c^*) = c$. The following is now immediate. We state it as a lemma so that we can refer to it later.

Lemma 9. *$u(l_c^*) \geq u(l_{c'}^*)$ iff $l_c^* \succeq l_{c'}^*$; similarly, $u(l_c^*) = u(l_{c'}^*)$ iff $l_c^* \sim l_{c'}^*$, and $u(l_c^*) > u(l_{c'}^*)$ iff $l_c^* \succ l_{c'}^*$.*

The key step in showing that we can reduce to a single menu is to show that, roughly speaking, for each menu, there exists a menu-dependent function g_M such that $u(g_M(s)) = -\sup_{f \in M} u(f(s))$. Stoye [29] proved a similar result, but he assumed that all menus were obtained by taking the convex hull of a finite set of acts. Because we allow arbitrary bounded menus, this result is not quite true for us. For example, suppose that the range of u is $(-1, \infty]$. Then there may be a menu M such that $\sup_{f \in M} u(f(s)) = 5$, so $-\sup_{f \in M} u(f(s)) = -5$. But there is no act g such that $u(g(s)) = -5$, since u is bounded below by -1. The following weakening of this result suffices for our purpose.

Lemma 10. *There exists a utility function U such that for every menu M, there exists $\epsilon \in (0,1]$ and constant act l^* such that for all $f, g \in M$, $f \succeq_M g \Leftrightarrow t(f) \succeq_{t(M)} t(g)$, where t has the form $t(f) = \epsilon f + (1-\epsilon)l^*$ and $t(M) = \{t(f) : f \in M\}$. Moreover, there exists an act $g_{t(M)}$ such that $u(g_{t(M)}(s)) = -\sup_{f \in t(M)} u(f(s))$ for all $s \in S$.*

Proof. The nontriviality and monotonicity axioms imply there must exist prizes x and y such that $U(x) > U(y)$. We consider four cases.

Case 1: The range of U is bounded above and below. Then we can rescale so that the range of U is $[-1, 1]$. Thus, there must be prizes x and y such that $U(x) = 1$ and $U(y) = -1$. For all $c \in [-1, 1]$, there must be a prize x' that is a convex combination of x and y such that $u(x') = c$, so we can clearly define a function g_M such that, for all $s \in S$, we have $u(g_M(s)) = -\sup_{f \in M} u(f(s))$. Furthermore, we know that such a g_M exists because it can be formed as an act which maps each state to an appropriate lottery over the prizes x and y. More generally, we know that an act with a certain utility profile exists if its utility for each state is within the range of U. This fact will be used in the other cases as well. Thus, in this case we can take t to be the identity (i.e., $\epsilon = 1$).

Case 2: The range of U is $(-\infty, \infty)$. Again, for all $c \in (\infty, \infty)$, there must exist a prize x such that $u(x) = c$. Since menus are assumed to be bounded above, we can again define the required function g and take $\epsilon = 1$.

Case 3: The range of U is bounded above and unbounded below. Then we can assume without loss of generality that the range is $(-\infty, 1]$, and for all c in the range, there is a prize x such that $u(x) = c$. For all menus M, $\epsilon > 0$, and acts $f, g \in M$, by Independence, we have that

$$f \succeq_M g \Leftrightarrow \epsilon f + (1-\epsilon)l_1^* \succeq_{\epsilon M + (1-\epsilon)l_1^*} \epsilon g + (1-\epsilon)l_1^*.$$

There exists an $\epsilon > 0$ such that for all $s \in S$,

$$1 \geq \sup_{f \in M} \epsilon u(f(s)) + (1-\epsilon) \geq -1.$$

Let $t(f) = \epsilon f + (1-\epsilon)l_1^*$. Clearly there exists an act $g_{t(M)}$ such that $u(g_{t(M)}(s)) = -\sup_{f \in t(M)} u(f(s))$ for all $s \in S$.

Case 4: The range of U is bounded below and unbounded above. By the upper-boundedness axiom, every menu has an upper bound on its utility range. Therefore, for every menu M, $\epsilon > 0$, and all acts f and g in M, by Independence,

$$f \succeq_M g \Leftrightarrow \epsilon f + (1-\epsilon)l_{-1}^* \succeq_{\epsilon M + (1-\epsilon)l_{-1}^*} \epsilon g + (1-\epsilon)l_{-1}^*.$$

There exists $\epsilon > 0$ such that for all $s \in S$,

$$\sup_{f \in M} \epsilon u(f(s)) + (1-\epsilon) u(l^*_{-1}(s)) \leq 1.$$

Let $t(f) = \epsilon f + (1-\epsilon) l^*_{-1}$. Again, it is easy to see that $g_{t(M)}$ exists. $\square$

In light of Lemma 10, we henceforth assume that the utility function u derived from U is such that its range is either $(-\infty, \infty)$, $[-, 1, 1]$, $(-\infty, 1]$, or $[-1, \infty)$. In any case, its range always includes $[-1, 1]$.

Before proving the key lemma, we establish some useful notation for acts and utility acts (real-valued functions on S). Given a utility act b, let f_b, the act corresponding to b, be the act such that $f_b(s) = l_{b(s)}$, if such an act exists. Conversely, let b_f, the utility act corresponding to the act f, be defined by taking $b_f(s) = u(f(s))$. Note that monotonicity implies that if $f_b = g_b$, then $f \sim_M g$ for all menus M. That is, only utility acts matter. If c is a real, we take c^* to be the constant utility act such that $c^*(s) = c$ for all $s \in S$.

Lemma 11. *Let M^* be the menu consisting of all acts f such that $(-1)^* \leq b_f \leq 0^*$. Then $(U, \mathcal{P}^+)$ represents $\succeq_{M^*}$ (i.e., $\succeq_{M^*} = \succeq^{S,X,U}_{M^*,\mathcal{P}^+}$) iff $(U, \mathcal{P}^+)$ represents $\succeq_M$ for all menus M.*

Proof. Our arguments are similar in spirit to those of Stoye [29].

By Lemma 10, there exists t such that $t(f) = \epsilon f + (1-\epsilon)h$ for a constant function h such that

$$f \succeq_M g \text{ iff } t(f) \succeq_{t(M)} t(g);$$

moreover, for this choice of t, the act $g_{t(M)}$ defined in Lemma 10 exists.

By Independence,

$$t(f) \succeq_{t(M)} t(g) \text{ iff } \frac{1}{2}t(f) + \frac{1}{2}g_{t(M)} \succeq_{\frac{1}{2}t(M)+\frac{1}{2}g_{t(M)}} \frac{1}{2}t(g) + \frac{1}{2}g_{t(M)}.$$

Let M^* be the menu that contains all acts with utilities in $[-1, 0]$. By INA, we know that for all acts f and g, and menus M for which g_M is defined, we have

$$f \succeq_M g \text{ iff } \frac{1}{2}f + \frac{1}{2}g_M \succeq_{M^*} \frac{1}{2}g + \frac{1}{2}g_M.$$

This is because acts of the form $\frac{1}{2}f + \frac{1}{2}g_M$ are never strictly optimal with respect to the menu $\frac{1}{2}M + \frac{1}{2}g_M$. At every state s there must be some act in

$\frac{1}{2}M + \frac{1}{2}g_M$ that has utility 0 at s (namely, the mixture that involves an act $f \in M$ whose utility at s is maximal; that is, $u(f(s)) \geq \max_{f' \in M} u(f'(s))$. Thus,

$$f \succeq_M g \text{ iff } \frac{1}{2}t(f) + \frac{1}{2}g_{t(M)} \succeq_{M^*} \frac{1}{2}t(g) + \frac{1}{2}g_{t(M)}.$$

Since the MWER representation also satisfies Independence and INA, we know that for all menus M, and acts f and g in M,

$$f \succeq^{S,X,U}_{M,\mathcal{P}^+} g \Leftrightarrow t(f) \succeq^{S,X,U}_{t(M),\mathcal{P}^+} t(g) \Leftrightarrow \frac{1}{2}t(f) + \frac{1}{2}g_{t(M)} \succeq^{S,X,U}_{M^*,\mathcal{P}^+} \frac{1}{2}t(g) + \frac{1}{2}g_{t(M)}.$$

Therefore, to show that $\succeq_M$ has a MWER representation with respect to $(U, \mathcal{P}^+)$, it suffices to show that $\succeq_{M^*}$ has a MWER representation with respect to $(U, \mathcal{P}^+)$. □

In the sequel, we drop the menu subscript when we refer to the family of preferences, and just write $\succeq$ (to denote $\succeq_{M^*}$); by Lemma 11, it suffices to consider $\succeq_{M^*}$.

It is straightforward to check that $\succeq_{M^*}$ satisfies completeness, transitivity, nontriviality, monotonicity, mixture continuity, independence, INA, and ambiguity aversion. Therefore, by Lemma 1, there exists some $(U, \mathcal{P}^+)$ representing $\succeq_{M^*}$. By Lemma 11, $(U, \mathcal{P}^+)$ represents $\succeq_M$ for all menus M, as required.

Since the axioms hold for all menus in $\mathcal{M}_B$, they clearly continue to hold if we restrict to $\mathcal{M}_F$ and $\mathcal{M}_C$. To prove the converse in the case of $\mathcal{M}_F$ we first argue that if the preference orders $\succeq_M$ for $M \in \mathcal{M}_F$ satisfy the axioms, then they uniquely determine preference orders $\succeq_M$ for menus $M \in \mathcal{M}_B$ that also satisfy the axioms. Clearly, it also then follows that the set of preference orders $\succeq_M$ for $M \in \mathcal{M}_C$ determines $\succeq_M$ for $M \in \mathcal{M}_B$. The proof immediately follows from this observation and the proof in the case of $\mathcal{M}_B$.

Consider a bounded menu M. The *utility frontier* of menu M is a function mapping each state to the maximum utility achieved in that state by any act in M. Since S is assumed to be finite, there exists a finite subset $M' \subseteq M$ such that the utility frontier of M' is the same as the utility frontier of M. Therefore, for all acts $f, g \in M$,

$$f \succeq_M g \Leftrightarrow f \succeq_{M' \cup \{f,g\}} g,$$

by Axiom 5. Since M' is finite, we have shown what we need.

	SEU	REG	MER	MWER	MMEU
Ax. 1-6,8-10	✓	✓	✓	✓	✓
Ind	✓	✓	✓	✓	
C-Ind	✓				✓
Ax. 12	✓	✓	✓		
Symmetry	✓	✓			

Table 6: Characterizing axioms for several decision rules.

Finally, for $\mathcal{M}_C$, suppose that M is a convex set of acts generated by the finite set A_M. Then, for all $f, g \in A_M$,

$$f \succeq_{A_M} g \Leftrightarrow f \succeq_M g,$$

by Axiom 5, since no interior points in M can be strictly optimal; hence, interior points can be removed from M without changing preferences. Thus, the result for $\mathcal{M}_C$ follows from the result for $\mathcal{M}_F$. □

It is instructive to compare Theorem 3 to other representation results in the literature. Anscombe and Aumann [2] showed that the menu-independent versions of axioms 10–14 and 16 characterize SEU. The presence of Axiom 16 (menu-independent Independence) greatly simplifies things. Gilboa and Schmeidler [11] showed that axioms 10–15 together with one more axiom that they call *certainty-independence* characterizes MMEU. Certainty-independence, or C-independence for short, is a weakening of independence (which, as we observed, does not hold for MMEU), where the act h is required to be a constant act. Since MMEU is menu-independent, we state it in a menu-independent way.

Axiom 19. *(C-Independence) If h is a constant act, then $f \succeq g$ iff $pf + (1-p)h \succeq pg + (1-p)h$.*

Table 6 describes the relationship between the axioms characterizing the decision rules.

C Characterizing MWER with Likelihood Updating

We can in fact directly translate the MDC axiom into a setting with preference relations instead of choice functions.

Definition 2 (Null event)**.** *An event E is* null *if, for all $f, g \in \Delta(Y)^S$ and menus M with $fEg, g \in M$, we have $fEg \sim_M g$.*

MDC. For all non-null events E, $f \succeq_{E,M} g$ iff $fEh \succeq_{MEh} gEh$ for some $h \in M$.[8]

The key feature of MDC is that it allows us to reduce all the conditional preference orders $\succeq_{E,M}$ to the unconditional order $\succeq_M$, to which we can apply Theorem 3.

Theorem 4. *For all Y, U, S, $\mathcal{P}^+$, and $M \in \mathcal{M}_B$, the family of preference orders $\succeq^{S,Y,U}_{\mathcal{P}^+|E,M}$ for events E such that $\overline{\mathcal{P}}^+(E) > 0$ satisfies Axioms 10–18 and MDC. Conversely, if a family of preference orders $\succeq_{E,M}$ on the acts in $\Delta(Y)^S$ satisfies Axioms 10–18 and MDC for $M \in \mathcal{M}_B$ (resp., $\mathcal{M}_F$, $\mathcal{M}_C$), then there exists a utility U on Y and a weighted set $\mathcal{P}^+$ of probabilities on S such that $C(\mathcal{P}^+)$ is regular, and for all non-null E, $\succeq_{E,M} = \succeq^{S,Y,U}_{\mathcal{P}^+|E,M}$. Moreover, U is unique up to affine transformations, and $C(\mathcal{P}^+)$ is unique in the sense that if $\mathscr{Q}^+$ represents $\succeq_{E,M}$, and $C(\mathscr{Q}^+)$ is regular, then $C(\mathscr{Q}^+) = C(\mathcal{P}^+)$.*

Proof. Since $\succeq_M = \succeq_{S,M}$ satisfies Axioms 10–18, there must exist a weighted set $\mathcal{P}^+$ of probabilities on S and a utility function U such that $f \succeq_M g$ iff $f \succeq^{S,Y,U}_{M,\mathcal{P}^+} g$. The rest of the proof is identical to that of Theorem 2; we do not repeat it here. □

Analogues of MDC have appeared in the literature before in the context of updating preference orders. In particular, Epstein and Schneider [7] discuss a menu-independent version of MDC, although they do not characterize updating in their framework. Ghirardato [9] characterizes update for a menu-independent version of DC. Sinischalchi [27] also uses an analogue of MDC in his axiomatization of measure-by-measure updating of MMEU. Like us, he starts with an axiomatization for unconditional preferences, and adds an axiom called *constant-act dynamic consistency* (CDC), somewhat analogous to MDC, to extend the axiomatization of MMEU to deal with conditional preferences. CDC in the form in [27] was first proposed by Pires [22], based on an observation of Jaffray [14].

[8] Although we do not need this fact, it is worth noting that the MWER decision rule has the property that $fEh \succeq_{MEh} gEh$ for some act h iff $fEh \succeq_{MEh} gEh$ for all acts h. Thus, this property follows from Axioms 10–18.

References

[1] M. Abdellaoui, A. Baillon, L. Placido, and P. P. Wakker. The Rich Domain of Uncertainty: Source Functions and Their Experimental Implementation. *American Economic Review*, 101(2):695–723, April 2011.

[2] F. Anscombe and R. Aumann. A definition of subjective probability. *Annals of Mathematical Statistics*, 34:199–205, 1963.

[3] H. Bleichrodt. Reference-dependent expected utility with incomplete preferences. *Journal of Mathematical Psychology*, 53(4):287 – 293, 2009.

[4] A. Chateauneuf and J. Faro. Ambiguity through confidence functions. *Journal of Mathematical Economics*, 45:535 – 558, 2009.

[5] S. H. Chew. A generalization of the quasilinear mean with applications to the measurement of income inequality and decision theory resolving the allais paradox. *Econometrica*, 51(4):1065–92, July 1983.

[6] G. de Cooman. A behavioral model for vague probability assessments. *Fuzzy Sets and Systems*, 154(3):305–358, 2005.

[7] L. G. Epstein and M. Le Breton. Dynamically consistent beliefs must be Bayesian. *Journal of Economic Theory*, 61(1):1–22, 1993.

[8] L. G. Epstein and M. Schneider. Learning under ambiguity. *Review of Economic Studies*, 74(4):1275–1303, 2007.

[9] P. Ghirardato. Revisiting savage in a conditional world. *Economic Theory*, 20(1):83–92, 2002.

[10] I. Gilboa and D. Schmeidler. Maxmin expected utility with a non-unique prior. *Journal of Mathematical Economics*, 18:141–153, 1989.

[11] I. Gilboa and D. Schmeidler. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18(2):141–153, 1989.

[12] T. Hayashi. Regret aversion and opportunity dependence. *Journal of Economic Theory*, 139(1):242–268, 2008.

[13] T. Hayashi. Regret aversion and opportunity dependence. *Journal of Economic Theory*, 139(1):242–268, 2008.

[14] J.-Y. Jaffray. Bayesian updating and belief functions. *Systems, Man and Cybernetics, IEEE Transactions on*, 22(5):1144–1152, Sep 1992.

[15] J.-Y. Jaffray. Dynamic decision making with belief functions. In R. R. Yager, J. Kacprczyk, and M. Fedrizzi, editors, *Advances in the Dempster-Shafer Theory of Evidence*, pages 331–352. Wiley, New York, 1994.

[16] P. Klibanoff, M. Marinacci, and S. Mukerji. A smooth model of decision making under ambiguity. *Econometrica*, 73(6):1849–1892, November 2005.

[17] D. M. Kreps. *Notes on the Theory of Choice*. Westview Press, Boulder, Colo., 1988.

[18] G. Loomes and R. Sugden. Regret Theory: An Alternative Theory of Rational Choice under Uncertainty. *Economic Journal*, 92(368):805–24, 1982.

[19] F. Maccheroni, M. Marinacci, and A. Rustichini. Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica*, 74(6):1447–1498, 2006.

[20] S. Moral. Calculating uncertainty intervals from conditional convex sets of probabilities. In *Proc. Eighth Conference on Uncertainty in Artificial Intelligence (UAI '95)*, pages 199–206. 1992.

[21] J. Niehans. Zur preisbildung bei ungewissen erwartungen. *Schweizerische Zeitschrift für Volkswirtschaft und Statistik*, 84(5):433–456, 1948.

[22] C. P. Pires. A rule for updating ambiguous beliefs. *Theory and Decision*, 53(2):137–152, 2002.

[23] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, N. J., 1970.

[24] T. Sarver. Anticipating regret: Why fewer options may be better. *Econometrica*, 76(2):263–305, 2008.

[25] L. Savage. The theory of statistical decision. *Journal of the American Statistical Association*, 46:55–67, 1951.

[26] L. Savage. *The Foundations of Statistics*. Wiley, New York, 1954.

[27] M. Siniscalchi. Dynamic choice under ambiguity. *Theoretical Economics*, 6(3):379–421, 2011.

[28] J. Stoye. Axioms for minimax regret choice correspondences. *Journal of Economic Theory*, 146(6):2226 – 2251, 2011.

[29] J. Stoye. Statistical decisions under ambiguity. *Theory and Decision*, 70(2):129–148, 2011.

[30] J. Stoye. Choice theory when agents can randomize. Unpublished manuscript, 2013.

[31] P. Walley. Statistical inferences based on a second-ordr possibility distribution. *International Journal of General Systems*, 26(4):337–383, 1997.