Asymmetric Gain-Loss Preferences: 
Endogenous Determination of Beliefs and Reference Effects *

Juan Sebastián Lleras†

April 2013 ‡

Abstract

This paper characterizes a model of reference-dependence, where a state-contingent contract (act) is evaluated by its expected value and its expected gain-loss utility. The expected utility of an act serves as the reference point, hence gains occur at states where the act provides a better-than-expected outcome, and losses occur at states where the act provides an outcome that is worse than expected. Beliefs, preferences over outcomes, and a degree of reference-dependence characterize the utility representation, and all are uniquely identified from behavior. The difference between this utility representation and subjective expected utility is captured by a one-dimensional parameter, which identifies the magnitude and direction of the reference effect independently from risk attitudes and beliefs. A link between reference-dependence and attitudes towards uncertainty is established.

Keywords: Reference Dependent Preferences, Endogenous Reference Points, Gain-Loss Attitudes, Subjective Expected Utility, Belief Distortion.

*I would like to thank David Ahn, Haluk Ergin, Shachar Kariv, Jian Li, Omar Nayeem, Luca Rigotti, Chris Shannon, and seminar participants at UC Berkeley, Pitt, CMU, University of Manchester, NC State University, and Florida State University, for helpful discussion and comments. I am particularly grateful to Roee Teper for constant discussions and comments throughout all stages of this project. All remaining errors are mine.

†Contact Information: j.sebastian.lleras@gmail.com

‡First Draft: September 5, 2012
1 Introduction

This paper studies the identification of reference points and reference effects. It provides a model of reference-dependence for choices under uncertainty in which reference points and reference-dependent attitudes are uniquely pinned down. The model differs from the standard subjective expected utility (SEU) model by one parameter that captures the direction and the magnitude of reference dependence. The parameter is one-dimensional and it is completely identified from behavior, which makes the model tractable and easily applicable.

Our main result is the Asymmetric Gain-Loss (AGL) representation of preferences. This representation is characterized by three components: $U$, $\mu$, and $\lambda$. Here $U$ is a utility function over outcomes, $\mu$ represents the subjective belief of the decision-maker (henceforth DM), and $\lambda$ is a new parameter that captures reference-dependence. The parameter $\lambda \equiv \lambda_l + \lambda_g$ is the sum of the weights given to losses ($\lambda_l$) and to gains ($\lambda_g$).

An act, which provides different outcomes depending on the realization of some uncertain state, is evaluated by the sum of its expected utility and its expected gain-loss utility. In our model, the expected utility of the act serves as the reference point that determines which outcomes are considered gains and which outcomes are considered losses.

1.1 Asymmetric Gain-Loss Preferences

Kahneman and Tversky [1979] introduced the notion of reference-dependent preferences in the seminal Prospect Theory. This theory aims to explain experimental violations of expected utility from the premise that risky outcomes are evaluated relative to a reference point rather than in absolute terms.\(^1\) The deviations from the reference point are weighted by a gain-loss value function, which has the feature that losses have more negative value than equal sized gains positive value.\(^2\) Prospect Theory, and most other models of reference dependence, do not provide a way to identify the reference point or the attitudes towards gains and losses from observable behavior. Work on reference dependence generally assumes an exogenous reference point rather than identifying one. In this paper, we identify reference points and reference effects from behavior. This identification is studied

\(^1\)Prospect Theory also introduced the notions of non-linear probability weighting and diminishing sensitivity.

\(^2\)This feature has been called loss aversion. Besides the experiments from Kahneman and Tversky [1979], many empirical findings have been attributed to the asymmetric assessment of gains and losses, such as the endowment effect [Kahneman et al., 1990], daily income targeting [Camerer et al., 1997], stock market participation [Barberis et al., 2006], to name a few.
here. Thus, in our model, these features of preferences are no longer exogenous.

The type of reference-dependent preferences studied in this paper are called asymmetric gain-loss preferences. A DM with asymmetric gain-loss preferences evaluates any state-contingent contract (act) as a combination of its expected utility and its expected gain-loss utility. The DM holds a probabilistic belief about the states. Based on this belief and her preferences over outcomes she forms the reference point, which is given by the expected utility of the act. The gain-loss utility measures how deviations from the reference point affect the DM. She realizes a gain in the states where the act provides an outcome that is better than the expected utility of that act, and a loss when the outcome provided by the act is worse than its expected utility. In the evaluation of expected gains and losses, she assigns a negative value to losses and a positive value to gains. The absolute value assigned to losses and gains can be different. Our model, unlike most work on reference dependence, can accommodate not only situations when losses have higher value than gains, but also instances where gains are valued more than equally sized losses.

The AGL representation provides a tool to behaviorally characterize gain-loss attitudes independently from risk attitudes and beliefs. The asymmetry between the valuation of gains and losses characterizes the degree of reference-dependence. Thus, the parameter $\lambda$ provides a simple “gain-loss index”, or “reference-dependence index”, which can be easily estimated. In addition, $\lambda$ provides a way to compare reference-dependence across agents independently of beliefs.

1.1.1 Beliefs, Dispersion, and Reference Dependence

Our representation captures reference-dependence in a way that deviates minimally from subjective expected utility (SEU); thus our model is as easy to estimate and apply as the standard model. The main behavioral difference with SEU is the identification of the subjective beliefs from observation; the fact that the reference point is the expected utility of the act confounds this identification.

Beliefs cannot be easily determined from behavior because the reference point is determined by beliefs, and this reference point affects the evaluation of the act. Thus the “willingness to pay” (i.e the certainty equivalent) of an act is not as transparently related to the expected utility of the act as in the standard model. The presence of reference-dependence can appear to contaminate the beliefs of the DM, which makes the

---

3This normalization has the natural intuition that losses are hurtful and gains are helpful.

4This is where the “asymmetric gain-loss preferences” name comes from.

5Recall that the SEU model is completely characterized by a utility function and a unique belief over the uncertainties, as in Savage [1954] or Anscombe and Aumann [1963].
identification of an unique prior from behavior challenging. This complication does not exist for the SEU model, where the certainty equivalent of an act is completely determined by beliefs.\footnote{To identify beliefs in the standard model we use the fact that the certainty equivalent of the act is equal to its expected utility.}

This contamination of beliefs due to the presence of gain-loss consideration establishes a link between reference dependence and ambiguity. Models of ambiguity typically consider DMs who have multiple priors.\footnote{See Siniscalchi [2008] for a survey of the literature.} Our model explains that reference dependence can be interpreted as a source of ambiguity. Due to the contamination of beliefs, a DM with gain-loss preferences might behave as if she has multiple priors in mind when making decisions, despite being probabilistically sophisticated with an unique prior.\footnote{Probabilistic sophistication is defined by Machina and Schmeidler [1995]. It characterizes DMs who behave as if they have a unique prior over the state space.} In addition, since the expected utility serves as the reference point and deviations from the reference point influence the DM’s preferences, our model is related to the models that study attitudes towards dispersion from the mean, such as Siniscalchi [2009] and Grant and Polak [2011]. The relationship between this paper, attitudes towards uncertainty, and mean dispersion preferences is studied in sections 7 and 7.

\subsection*{1.1.2 Reference Point Determination}

The model of reference-dependence studied here differs conceptually from the few existing models where some reference point is endogenously identified, such as Shalev [2000], Brunnermeier and Parker [2005], Kőszegi and Rabin [2006] or Sarver [2011].\footnote{Brunnermeier and Parker [2005] study “optimal expectations”, not reference-dependence. The relationship between their model and this paper, is that expectations and preferences over certain outcomes completely determine the reference point in this model.} The key difference is that in our model the DM has no control over the reference point, so that reference point does not depend on any actions taken by the DM. The reference point is identified from beliefs and preferences over outcomes, neither of which are affected by reference dependence. In the models of Shalev [2000], Brunnermeier and Parker [2005], Kőszegi and Rabin [2006], or Sarver [2011], the DM is aware of her reference dependence and how her choices affect the reference point. These papers require that the DM’s choices are optimal given the reference point they induce, so these models require an equilibrium condition to account for the mutual relationship between reference points and choices.\footnote{This requires a fixed point condition, which can be computationally challenging.} The model presented here has the benefit that it captures gain-loss attitudes and elicits
the reference point, reference effects, and beliefs simultaneously from choices in a way that is easy to compute or estimate without the need of equilibrium conditions.\footnote{The approach in Ok et al. [2011] also tackles the reference point determination problem under a very general framework, where they do not need an equilibrium condition to characterize reference dependence. Nonetheless in their framework is impossible to identify reference points and reference effects uniquely.}

1.2 Example of Asymmetric Gain-Loss Preferences

The formal framework and the functional representation are introduced in sections 3 and 4. The following numerical example using insurance decisions explains the intuition behind the representation, and shows how asymmetric gain-loss preferences can explain different types of behavior regarding uncertainty.

Suppose that a car can have two mutually exclusive types of damage: comprehensive or collision damage. Consider a risk neutral DM (i.e. one who cares only about expected monetary value) who believes that the probability of either type of accident is .25, and the probability that nothing happens to the car is .5. Her car is worth $20,000, and any type of damage will cost $10,000. Taking the probability of accidents into account, the expected value of owning a car is $15,000. Expected utility would predict that a risk neutral DM is willing to pay a premium that is lower than the expected damage to the car. In this case, the maximum willingness to pay is $2,500 for partial insurance (collision only or comprehensive only coverage), and $5,000 for full insurance.

Now suppose that the DM has gain-loss preferences: in addition to the expected value she wants to avoid losses, so she subtracts any expected losses from the expected value to determined the valuation. A loss for her is any situation (state) where her outcome is worse than the expected value, therefore the outcome of any accident is a loss to her. The magnitude of the loss is $5,000 for either type of accident because it is the net loss of the accident with respect to expectations. Given that she believes that some accident, collision or comprehensive, will happen with probability .5, her expected losses are $2,500. So her valuation for the car is $12,500, instead of its expected value of $15,000. Thus, her willingness to pay for full insurance is $7,500 rather than $5,000 because she is willing to pay up to the point where she gets $12,500 in every possible state. Following this intuition it is possible to calculate the willingness to pay for partial insurance. Table 1.2 summarizes the relevant elements for each decision: the value of each possible plan for each realization of the uncertainty, its expected value and the expected losses (given by the beliefs that the probability of each type of accident is .25), and the total value of each plan which is given by expected value minus expected losses.
Table 1: Willingness to Pay for Insurance for a Risk Neutral Agent

<table>
<thead>
<tr>
<th></th>
<th>no damage</th>
<th>collision</th>
<th>comp.</th>
<th>ex. value</th>
<th>ex. loss</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>no ins.</td>
<td>20,000</td>
<td>10,000</td>
<td>10,000</td>
<td>15,000</td>
<td>2,500</td>
<td>12,500</td>
</tr>
<tr>
<td>full</td>
<td>20,000 − (P)</td>
<td>20,000 − (P)</td>
<td>20,000 − (P)</td>
<td>20,000 − (P)</td>
<td>0</td>
<td>20,000 − (P)</td>
</tr>
<tr>
<td>coll.</td>
<td>20,000 − (P_1)</td>
<td>20,000 − (P_1)</td>
<td>10,000 − (P_1)</td>
<td>17,500 − (P_1)</td>
<td>1,875</td>
<td>15,625 − (P_1)</td>
</tr>
<tr>
<td>comp.</td>
<td>20,000 − (P_2)</td>
<td>10,000 − (P_2)</td>
<td>20,000 − (P_2)</td>
<td>17,500 − (P_2)</td>
<td>1,875</td>
<td>15,625 − (P_2)</td>
</tr>
</tbody>
</table>

\(P\) is the price for full insurance. The maximum such price that she is willing to pay is \(P = 7,500\), which equates value of the insurance and the value of no insurance.

\(P_1\) is the price for collision only insurance. The maximums such price is such that the willingness to pay for insurance equals the no insurance price, i.e. \(P_1 = 3,125\).

\(P_2\) is the price for comprehensive only insurance, and it is symmetric to the collision only insurance because the cost and the likelihood of the events is the same. Hence \(P_2 = 3,125\) as well.

The results from table 1.2 show that if there are gain-loss considerations, a risk neutral DM could be willing to pay more to fully insure ($7,500) than the sum of the partial insurances ($3,125 each). It is easy to see that if the DM cares only about gains and not about losses, and assigns positive value to expected gains, she will be willing to pay more to partially insure than to fully insure.

Asymmetric gain-loss preferences generalize this example, by allowing gains and losses to have different values, as well as allowing DMs to have different beliefs and risk attitudes. In general, as analysts we do not observe beliefs, utility functions, or how much decision-makers care about gains and losses; all we can observe is choices. The behavioral identification of the elements of the representation, gains, losses, and the decision value attached to expected gains and losses, is the focus of this paper.

1.3 Representation as Model of Reference Dependence

The general idea behind the AGL representation is not entirely new. A similar functional form was introduced by Gul [1991] in a theory of disappointment aversion, where outcomes that the DM considers elating and outcomes that the DM considers disappointing are valued differently in choices under risk. In terms of reference-dependence, a similar representation was first introduced by Kőszegi and Rabin [2006] (henceforth KR). Nonetheless, despite of the similarity between our functional representation and Gul and KR, our domain of choice is different, and our behavioral characterization has different implications than these two models.
Gul [1991] studies deviations from expected utility under risk, where the certainty equivalent of a lottery is used to characterize outcomes as disappointing or elating. After a lottery is decomposed into elating and disappointing outcomes, the DM weights expected elation and expected disappointment differently. In our paper the DM weights gains and losses differently, much like the DM in Gul’s model. However Gul’s paper studies choices under risk where objective probabilities are exogenously given and used to determine elation and disappointment. Whereas here beliefs are subjective and have to be elicited from preferences. This elicitation is the key element of the representation because beliefs determine what is a gain and what is a loss, and hence beliefs determine how gain and losses are aggregated.

KR extend the insights of Prospect Theory and provide a model of reference-dependent preferences that values consumption utility as well as gain-loss utility, given by utility deviations from the reference point. KR use “recent expectations” about outcomes as their reference point. Using this reference point can unify most of the different theories of reference points (status-quo, lagged consumption, social comparisons, etc). Our paper follows the general idea that beliefs and expectations determine the reference point by considering the expected utility of an act to be the reference point.

Although the general representation of reference-dependence in this paper follows the idea introduced by KR, the two models are conceptually different. First and foremost, in this model the DM has no control over the reference point, and her choices do not affect the reference point. Thus there is no need for an equilibrium condition as KR. Second, the KR model does not provide a model of how beliefs are determined; rather their model shows how a reference point can be endogenously determined given that DMs hold correct beliefs. For us, the formation of beliefs is an essential part of the model because the identification of beliefs is central for the determination of the reference point endogenously. Finally the domain over which the beliefs defined is different. In KR, the uncertainty is over the possible choice sets that the DM will see once she has to make a choice; in this paper the uncertainty is over the realization payoff-relevant states. A consequence of this difference in the domain of uncertainty is that in KR there is a choice to be made after the uncertainty is realized, while in this model all decisions are made before a state is observed. This difference implies that the assessment of gains and losses is different in this model and KR. In KR, gains and losses are assessed after the uncertainty is realized and the DM has made a choice, whereas in this model gains and losses as assessed before any state is realized. A lengthy discussion about the relation of this paper with KR and Gul’s disappointment aversion model is presented in section 7.
1.4 Structure of the Paper

The rest of the paper is structured as follows. Section 2 reviews the literature related to this paper. Section 3 presents the setup and model, and describes the concept of alignment of acts, which is instrumental for the endogenous determination of a reference point. Section 4 describes the utility representation, and provides a characterization. Section 5 explains the links between gain-loss and ambiguity, and section 6 provides comparative statics results. Section 7 discusses the results and relations to the literature, particularly the reference point determination in KR, and the relationship between this model and models that study preferences that depend on dispersions from the mean. Section 8 concludes. All proofs are in the appendix.

2 Related Literature

This section briefly surveys the literature. A lengthy discussion about some of the related literature is presented in section 7. This paper links two topics in the literature that have not been formally linked yet: reference-dependent preferences, and attitudes towards variation and ambiguity. We show that the idea of Köszegi and Rabin [2006] about reference dependent preferences, where the reference point depends on expectations, provides a clean way to link these two concepts in the domain of choice under uncertainty.

In many decision theory models, the status quo has been interpreted as a reference point. Giraud [2004a], Masatlioglu and Ok [2005], Sugden [2003], Sagi [2006], Rubinstein and Salant [2007], Apesteguia and Ballester [2009], Ortoleva [2010], Riella and Teper [2012] and Masatlioglu and Ok [2012] provide models of reference dependence, where the reference point is exogenously given. Along with Köszegi and Rabin [2006] and Köszegi and Rabin [2007], three other papers that tackle the problem of endogenous reference point determination are Giraud [2004b], Ok et al. [2011] and Sarver [2011].

For gain-loss preferences the evaluation of acts depends on the state by state variation of the act. Although some papers have studied attitudes toward variation in the context of risk and uncertainty, none relates such attitudes to reference-dependence. In the risk domain, Quiggin and Chambers [1998, 2004] measure attitudes towards risk, which depend on the expectation of the lottery and a risk index of the lottery that depends on the variation of the distribution. A different approach is taken by Gul [1991], who provides a theory of disappointment aversion. In Gul [1991], variation depends on the composition of a lottery among disappointment and elation outcomes.

In the uncertainty domain some models use the idea of a reference prior, which is
a baseline for adjusting imprecise information, or to measure different specifications of a model like in the multiplier preferences of Hansen and Sargent [2001]. Wang [2003], Gajdos et al. [2004], Gajdos et al. [2008], Siniscalchi [2009], Strzalecki [2011] and Chambers et al. [2012] take this approach. In contrast, in this paper the DM has a unique prior over the states, but there is some contamination due to gain-loss considerations.

Grant and Polak [2011] provide a general model of mean-dispersion preferences, where deviations from the expectation affect utility as well. They show that many well-known families of preferences such as Choquet EU [Schmeidler, 1989], Maxmin EU [Gilboa and Schmeidler, 1989], invariant biseparable preferences [Ghirardato et al., 2004], variational preferences [Maccheroni et al., 2006], and Vector EU [Siniscalchi, 2009], belong to this family of preferences. The asymmetric gain-loss preferences belong to this family of preferences as well, since deviations from the reference point are in fact deviations from the expected utility of each act. However, the general mean-dispersion preferences are so general that it is not possible to provide clear comparative statics results like the ones presented here. Chambers et al. [2012] model mean-dispersion preferences where absolute uncertainty aversion is allowed to vary across acts, unlike in the mean-dispersion preferences of Grant and Polak [2011] where absolute uncertainty aversion is constant. After presenting the model and main results, we return in section 7 to give a longer discussion of the relationship between this paper and the literature.

3 Model

We use the framework employed by Anscombe and Aumann [1963]. Let \( S = \{s_1, s_2, ..., s_n\} \) be a finite set of states of the world that represent all possible payoff-relevant contingencies for the DM; any \( E \subseteq S \) is called an event. Define \( \mathcal{E} = \mathcal{P}(S) \setminus \{\emptyset, S\} \) as the set of all nonempty events (subsets of \( S \)) that have non-empty complement. Let \( X \) be a prize space, and let \( \mathcal{L} \) be the set of distributions on \( X \) with finite support. Denote by \( \mathcal{F} = \mathcal{L}^S \) the set of all acts, that is, functions from \( S \) to \( \mathcal{L} \), \( f : S \rightarrow \mathcal{L} \). Take the mixture operation on \( \mathcal{F} \) as the standard pointwise mixture, where for any \( \alpha \in [0,1] \), \( \alpha f + (1 - \alpha)g \in \mathcal{F} \) gives \( \alpha f(s) + (1 - \alpha)g(s) \in \mathcal{L} \) for any \( s \in S \). Abusing notation, any \( c \in \mathcal{L} \) can be identified with the constant act that yields \( c \) for all \( s \in S \), i.e. \( c \in \mathcal{F} \) where \( c(s) = c \) for all \( s \in S \). Let \( \mathcal{F}_c \) be the set of constant acts, which is identified with \( \mathcal{L} \). Model preferences on \( \mathcal{F} \) by a binary relation \( \succsim \); \( \succ \) and \( \sim \) denote respectively the asymmetric and symmetric components of \( \succsim \).

\[ ^{12} \text{Other models where the uncertainty aversion is allowed to vary are Klibanoff et al. [2005] and Cerreia-Vioglio et al. [2011].} \]
For each $f \in \mathcal{F}$, if there is some $c_f \in \mathcal{F}_c$, such that $f \sim c_f$ call $c_f$ the certainty equivalent of $f$.

**Balanced Pairs of Acts**

A particularly important type of act, is those acts that provide prefect hedges against uncertainty. Hedging not only gets rid of uncertainty but also removes all possible gain-loss considerations from the act. Siniscalchi [2009] calls a pair of acts that provide perfect hedging as *complementary acts*. These are pairs of acts such that for every state, the 50-50 mixture of the two acts gives the DM the same constant act. Formally, $f$, and $f'$ are complementary if for every $s, s' \in S$, \( \frac{1}{2} f(s) + \frac{1}{2} f'(s) = \frac{1}{2} f(s') + \frac{1}{2} f'(s') \). We strengthen the definition of complementary acts to further require the acts to be indifferent. Call a pair of acts $(f, \bar{f})$ balanced if they provide a perfect hedge and are indifferent to each other.\(^{13}\) The importance of balanced acts is that eliminating subjective gain-loss considerations allows an analyst to identify beliefs from preferences.

**Definition 1.** Two acts $f$ and $\bar{f}$ are balanced if $f \sim \bar{f}$, and for any states $s, s' \in S$

\[
\frac{1}{2} f(s) + \frac{1}{2} \bar{f}(s) \sim \frac{1}{2} f(s') + \frac{1}{2} \bar{f}(s')
\]

If there exists $e_f \in \mathcal{F}_c$ such that $e_f = \frac{1}{2} f(s) + \frac{1}{2} \bar{f}(s)$ for all $s \in S$, we call $e_f$ the hedge of $f$. $(f, \bar{f})$ is referred to as a balanced pair, and $\bar{f}$ is a balancing act of $f$ (and vice-versa).

When the notation $\bar{f}$ is used, it is always in reference to the balancing act of $f \in \mathcal{F}$. The conditions imposed on preferences below guarantee that $c_f$ and $e_f$ are unique and well defined for each $f$.\(^ {14}\)

**Act Alignment: Separating Positive and Negative States**

When a DM cares about expected gains and losses she must have a way of determining what is a gain and what is a loss. The certainty equivalent of an act (willingness to pay for it) does not provide a measure of what is a gain and what is a loss, since gain-loss considerations affect the assessment of the acts. For instance, in the insurance example from section 1, the willingness to pay for insurance is $12,500, even though the expected

\(^{13}\) For example, in sports betting if the DM is indifferent between betting on either side of the line, those bets are complementary. Since regardless of the result, the DM will get the same outcome.

\(^{14}\) Grant and Polak [2011] use complementary acts to define a baseline prior. Under different conditions this paper uses balanced pairs to define gains and losses, and the conditions guarantee that the baseline prior is related to $e_f$, in the sense that the expected utility is the utility of $e_f$. 

value of the car is $15,000, which is the value used by the DM to determine gains and losses.

Balanced acts provide a behavioral way of separating gains and losses. We require that when the outcome in state \( s \) is considered a gain for \( f \), the outcome on state \( s \) is considered a loss for \( \bar{f} \). This is a natural requirement given that \( f \) and \( \bar{f} \) provide a perfect hedge to the DM. Therefore \( \bar{f} \) must have the opposite gain-loss composition of \( f \). For an act define positive states as those states that deliver gains, and negative states as those states that deliver losses.

**Definition 2.** Let \((f, \bar{f})\) be a balanced pair. Say \( s \in S \) is a positive state for \( f \) if \( f(s) \succ \bar{f}(s) \), and a negative state for \( f \) if \( \bar{f}(s) \succ f(s) \). If a state is both positive and negative (i.e. \( f(s) \sim \bar{f}(s) \)) say \( s \) is a neutral state for \( f \).

Any act induces a set of partitions of the state space into two events. Each partition is a way of separating states into positive and negative states which is induced by the act. This separation is done in a way that there is one event that contains only positive states, and one event that contains only negative states. The neutral states can be labeled as either positive or negative but not both. Because of neutral states, each act that has neutral states induces multiple partitions since there is freedom to label neutral states as either positive or negative. However, when there are no neutral states, each act has a unique way of partitioning the states into positive and negative. These partitions associated with each act are called the alignment of the act. We use the convention that the alignment of the act is represented by the event that includes the positive states \( E \) (the complement is the negative states) rather than saying that the alignment is represented by the partition \( \{E, E^c\} \).

**Definition 3.** For any \( f \in \mathcal{F} \), say \( f \) is aligned with the event \( E \in \mathcal{E} \) if for all \( s \in E \), \( f(s) \succeq \bar{f}(s) \) and for all \( s \in E^c \), \( \bar{f}(s) \succeq f(s) \).

From the definition, an act can be aligned with more than one event since whenever a state is neutral it can be part of \( E \) or of \( E^c \). For every \( E \in \mathcal{E} \), there is a set of acts that is aligned with \( E \).

**Definition 4.** Given any event \( E \subset S \), define \( \mathcal{F}^E \) be the set of acts where the positive

---

\[\text{This makes the notation easier because there is no need to specify which cell in the partition represents the positive states and which cell represents the negative states.}\]

\[\text{Besides constants, from the definition of an alignment no act can be aligned with } S \text{ or } \emptyset \text{ if preferences are monotonic.}\]
states are contained in \( E \), i.e.

\[
\mathcal{F}^E = \{ f \in H : \forall s \in E, f(s) \succsim \tilde{f}(s) \text{ and } \forall s \not\in E, \tilde{f}(s) \succsim f(s) \}\]

Note that any constant act is its own balancing act, therefore constant acts are aligned with all partitions of the state space. It is useful to consider acts that have only one alignment, which are called single alignment acts. These acts are important because they are acts where small perturbations on outcomes do not change the alignment.

**Definition 5.** Given \( f \in \mathcal{F} \), \( f \) is single-alignment act if for no \( s \in S \), \( f(s) \sim \tilde{f}(s) \).

If the event \( E \) represents an alignment of \( f \), every subset of \( E \) or \( E^c \) is called a non-overlapping event. These are the events where all the states are either all positive or all negative for \( f \), so there is no overlap between positive and negative states for \( f \). Non-overlapping events provide a way of specifying situations where there is no tradeoff between positive and negative states, only across one type of state.

**Definition 6.** Given \( f \in \mathcal{F} \), event \( F \subset S \) is a non-overlapping event for \( f \) if every state in \( F \) is aligned in the same way. \( F \) is non-overlapping for \( f \in \mathcal{F}^E \) if

\[
F \subseteq E \text{ or } F \subseteq E^c
\]

Abusing terminology, we say that \( F \) is non-overlapping for \( E \) whenever \( F \) is non-overlapping for \( f \in \mathcal{F}^E \).

### 4 Representation of Gain-Loss Preferences

This section formally presents the AGL representation first, and then gives the necessary and sufficient conditions on \( \succsim \) to represent preferences with an AGL representation \((U, \mu, \lambda)\).

#### 4.1 AGL Functional

The main result of the paper is the behavioral characterization of asymmetric gain-loss preferences, which are preferences that can be represented by the functional

\[
V(f) = \int_S (U \circ f)\mu(ds) + \int_S \eta^\lambda(U \circ f, \mu)\mu(ds)
\]

\( E \mu[U \circ f] \) Expected Utility

\( E \eta^\lambda(U \circ f, \mu) \) Expected Gain-Loss Utility
Here $U : \mathcal{L} \to \mathbb{R}$ is an expected utility function, $\mu$ is a distribution over $S$, and $\eta^\lambda_{\mathcal{L}} : [U(\mathcal{L})]^{[S]} \times \Delta(S) \to \mathbb{R}^{[S]}_+$ is the gain-loss function. The function $\eta^\lambda_{\mathcal{L}}$ depends on a utility vector and a distribution over the state-space $S$. In addition, $E_\mu[U \circ f]$, the expected utility of $f$ for belief $\mu$, serves as the reference point for $\eta^\lambda_{\mathcal{L}}$. The utility in states that provide gains relative to the reference point is valued differently than those states that provide losses. The parameters $\lambda_g$ and $\lambda_l$ are the weights given to gains and losses respectively. Formally, for all $s \in S$, the gain/loss on state $s$, $\eta^\lambda_{\mathcal{L}}(s) : U(\mathcal{L}) \times \Delta(S) \to \mathbb{R}_+$, is defined by

$$
\eta^\lambda_{\mathcal{L}}(U \circ f, \mu)(s) = \begin{cases} 
\lambda_g (U(f(s)) - E_\mu[U \circ f]) & \text{if } U(f(s)) \geq E_\mu[U \circ f] \\
\lambda_l (E_\mu[U \circ f] - U(f(s))) & \text{if } U(f(s)) < E_\mu[U \circ f] 
\end{cases}
$$

(4.2)

All the elements in the representation are identified from choice behavior: $\mu$ is unique, $U$ is cardinally unique, and $\lambda = \lambda_g + \lambda_l$ is identified uniquely. Note that in this model any gains-loss considerations take place because of uncertainty. The gain-loss function $\eta$ captures how the DM will feel for every realization of uncertainty compared to the reference point. For strictly risky choices, the DM behaves as an expected utility maximizer.\(^{17}\)

The parameter $\lambda = \lambda_l + \lambda_g$ is the new element of the representation. It captures reference dependence and it represents the difference between the weight of gains and losses. The parameters $\lambda_l$ and $\lambda_g$ are not uniquely identified but their sum is. Therefore for simplicity of exposition, the normalization that gains are always considered positive and losses negative is used.\(^{18}\) So we use $\lambda_g \geq 0$ and $\lambda_l \leq 0$, which implies that whenever $\lambda \leq 0$ losses are weighted more than gains, and when $\lambda \geq 0$ gains are weighted more than losses. The reference point for each act is its expected utility, which is determined by the beliefs and the utility function. This reference point does not affect how the reference point is determined, unlike other models of reference-dependence that require an equilibrium condition to deal with the feedback between choices and reference points. Also, unlike traditional models of reference-dependence, where losses always have higher value than equally sized gains, asymmetric gain-loss preferences allow gains to have more value than losses as well. Attitudes towards gains and losses are a feature of the decision

\(^{17}\)For instance, the model of disappointment aversion in Gul [1991] studies the case where outcomes of lotteries might be disappointing or elating (a very similar concept to gains and losses), and the DM exhibits an aversion to disappointment. This model could be used as the baseline case for risk as well, this is possible because of the separation of risk and uncertainty in the Anscombe and Aumann [1963] framework. For simplicity this paper assumes expected utility maximization for acts involving only risk.

\(^{18}\)It is behaviorally impossible to distinguish between situations where losses are weighted negatively and gains positively, and situations where losses are weighted positively and gains negatively.
process that should ideally be derived from behavior. If $\succsim$ admits an AGL representation $V(\cdot)$, we say that the triple $(U, \mu, \lambda)$ represents $\succsim$, where $\lambda = \lambda_g + \lambda_l$.

The AGL representation differs from the subjective expected utility representation, which is characterized by a utility function and an unique belief about the uncertainty, just by the parameter $\lambda$. Since the reference point is the expected utility of the act, the key element for the identification of the reference point and the reference effects is $\mu$. The presence of any reference-dependence, where expected gains and losses have an impact on the evaluation of the act, implies that the identification of beliefs a-la Anscombe and Aumann [1963], where the certainty equivalent captures information about beliefs, is no longer possible. The effect of gain-loss considerations implies that the willingness to pay for an act is not its expected utility regardless of the risk attitudes of the DM. In other words, the certainty equivalent of an act might not determine its expected value. The certainty equivalent of an act can be determined from data by eliciting the “willingness to pay”, whereas the expectations given any gain-loss consideration cannot be easily identified from behavior. Nonetheless we show that by identifying hedging behavior it is possible to pin down the beliefs.

4.2 Axioms

Standard Axioms

The first 4 conditions, A1-A4, are standard axioms in the literature of choice under uncertainty.

A 1. **Weak Order.** $\succsim$ is complete and transitive.

A 2. **Continuity.** For all $f \in H$, the sets $\{g \in H : g \succ f\}$ and $\{g \in H : f \succ g\}$ are closed.

A 3. **Strict Monotonicity.** If for all $s$, $f(s) \succsim g(s)$ then $f \succsim g$. If $f(s) \succ g(s)$ for some $s$ and $f(s) \succsim g(s)$ for all $s$, then $f \succsim g$.

A 4. **Unboundedness.** For any $f$, $g$ in $\mathcal{F}$, and any $\alpha \in (0,1)$, there exists $w$ and $z$ in $\mathcal{F}_c$ satisfying $g \succ \alpha w + (1-\alpha)f$ and $\alpha z + (1-\alpha)g \succ f$.

Unboundedness implies non-triviality of preferences (see Kopylov [2007] or Grant and Polak [2011]), and it is a technical condition that guarantees that the range of the utility function over outcomes is $(-\infty, \infty)$. In addition, strong monotonicity implies state-independence and that no state is null, i.e. the DM puts positive probability on every state occurring.
New Axioms: Mixture Conditions

The standard subjective expected utility model from Anscombe and Aumann [1963] is characterized by some version of A1-A4, plus the independence axiom. Independence requires that \( f \succeq g \) if and only if \( \alpha f + (1 - \alpha)h \succeq \alpha g + (1 - \alpha)h \) for any \( h \in \mathcal{F} \), and any \( \alpha \in (0, 1) \).

The independence axiom does not hold for asymmetric gain-loss preferences because it does not allow for gains and losses to be evaluated differently. Independence can be violated with asymmetric gain-loss preference because convex combinations of acts change the gain-loss composition of acts differently, and therefore changes the assessment of acts differently as well. Asymmetric gain-loss preferences relax independence, but impose three consistency requirements on how independence fails depending on how the gain-loss composition of acts changes when acts are mixed.

The first new axiom states that as long as the alignment of acts remains the same when mixing, then independence is preserved. In situations where two acts have the same alignment, taking any mixture of them does not change the composition of gains and losses, so the tradeoff between gains and losses should not change. For acts that are aligned the same way, the gains or losses of the mixture is the mixture of the gains or the losses of each act respectively.

**A 5. Mutual Pairwise Alignment Independence (MPA-independence).** If \( f, h \in \mathcal{F}^E \) and \( g, h \in \mathcal{F}^F \) for some \( E, F \in \mathcal{E} \), then \( f \succeq g \) if and only if \( \alpha f + (1 - \alpha)h \succeq \alpha g + (1 - \alpha)h \) for all \( \alpha \in [0, 1] \).

The second condition imposes structure on mixtures for pairs of acts that are not mutually aligned. Before explaining the new condition note two things. First, when any act \( h \) is neutral on an event \( E \) then mixing \( f \) with \( h \) should not change the gain-loss consideration of \( f \) within \( E \) directly.\(^{19}\) Second, for a single alignment act \( f \), mixing with any other act, where the weight on \( f \) is close to 1, does not change the alignment of \( f \) if preferences are continuous and monotonic.

Consider an event \( E \) that is non-overlapping for \( f \) and \( g \). The new axiom states that whenever mixing \( f \) and \( g \) with any \( h \) that is neutral on \( E^c \) is such that the alignment of \( \alpha f + (1 - \alpha)h \) and \( \alpha g + (1 - \alpha)h \) is the same as the alignment of \( f \) and \( g \) respectively, then the mixture with \( h \) perturbs \( f \) and \( g \) in the same direction. The condition requires that \( f \) and \( g \) agree that there is no tradeoff between gains and losses on \( E \), so if the alignment of \( f \) and the alignment of \( g \) are the same as the alignment of the respective mixtures, then

\(^{19}\)It might change the gain-loss considerations by changes outside of \( E \).
mixing $f$ and $g$ with $h$ keeps the same gain-loss consideration on $E^c$ for $f$ and $g$, therefore
the effect of mixing with $h$ should be consistent for both acts. In other words, the DM
must not change the direction of the assessment of these particular perturbations where
the alignment of the act does not change and it does not induce any tradeoff between
positive and negative states (only among one type of state for each act). This means that
$\alpha f + (1 - \alpha)h$ is preferred to $f$ if and only if $\alpha g + (1 - \alpha)h$ is preferred to $g$ locally (when
$\alpha$ is close to 1, so that the alignment of $f$ and $g$ does not change).

Using the intuition of the representation, if the DM cares about expected gains and
losses, then beliefs determine the DMs aggregation of gains and losses. So if there is a
change in the act that keeps the composition of gains and losses the same, but induces a
tradeoff among some states that are aligned the same way, the assessment of the change
with respect to the original act depends just on the beliefs. This condition is called
Non-overlapping Event Local Mixture Consistency.

For any single-alignment acts $f \in \mathcal{F}^E$ and $g \in \mathcal{F}^{E'}$, for any event $F$, which is non-
overlapping for both $E$, and $E'$, and any $h \in \mathcal{F}$ such that $h(s) \sim \bar{h}(s)$ for all $s \notin F$, there
exists $\alpha^* < 1$ such that

$$f \succcurlyeq \alpha f + (1 - \alpha)h \iff g \succcurlyeq \alpha g + (1 - \alpha)h$$

for all $\alpha \in (\alpha^*, 1)$.

The last axiom imposes a consistency condition on failures of independence when
reversing the role of gains and losses. Suppose that for two indifferent acts $f$ and $g$, the
DM strictly prefers mixing $f$ with $h$ than mixing $g$ with $h$ (with the same weights $\alpha$ and
$1 - \alpha$). This would be a failure of independence attributed to the fact that mixing $f$
with $h$ improves the gain-loss considerations more than mixing $g$ with $h$. If the roles of
gains and losses are reversed for $f$ and $g$, situation which is given by the balancing acts $\bar{f}$
and $\bar{g}$, then mixing $\bar{g}$ with $h$ should be strictly preferred to mixing $\bar{f}$ with $h$. So mixing
must worsen the gain-loss considerations for $\bar{f}$ more than for $\bar{g}$ because it improves $f$
more than $g$. If the gain-loss composition of acts is reversed, then the preferences over
mixing the acts with a fixed $h$ should reverse as well. This condition is called balanced
act antisymmetry.

A 7. Balanced Act Antisymmetry (BA-antisymmetry). For any acts $f$ and $g$ such that
$f \sim g$, for every $h \in H$, and for any $\alpha \in (0, 1)$, $\alpha f + (1 - \alpha)h \succeq \alpha g + (1 - \alpha)h$ implies
(i) \( \alpha \bar{f} + (1 - \alpha)h \ll \alpha \bar{g} + (1 - \alpha)h \) and

(ii) \( \alpha f + (1 - \alpha)\bar{h} \ll \alpha g + (1 - \alpha)\bar{h} \)

where \( \bar{f} \) and \( \bar{g} \) are balancing acts of \( f \) and \( g \) respectively.

These three axioms are related to relaxations of the independence axioms seen in the literature. First, MPA-independence implies certainty independence [Gilboa and Schmeidler, 1989]. Recall certainty independence requires that \( f \succeq g \) if and only if for all \( \alpha \in [0, 1] \), \( \alpha f + (1 - \alpha)c \succeq \alpha g + (1 - \alpha)c \) where \( c \in \mathcal{F}_c \). This follows from the fact that for any constant \( c \in \mathcal{F}_c \), \((c, c)\) constitutes a balanced pair. Therefore constants are aligned with all \( E \in \mathcal{E} \). Second, MPA-independence and BA-antisymmetry imply the condition called complementary independence [Siniscalchi, 2009]. Complementary independence states that for all complementary pairs \((f, f')\) and \((g, g')\) where \( f \succ f' \) and \( g \succ g' \), then for all \( \alpha \in [0, 1] \), \( \alpha f + (1 - \alpha)g \succeq \alpha f' + (1 - \alpha)g' \). These results are obvious, and they are key in understanding asymmetric gain-loss preferences in relation to models of ambiguity preferences and dispersion preferences. These relations explain results in sections 4.3 and 6, and discussed in section 7.

### 4.3 Representation Results

This section provides two utility representation and uniqueness results. Theorem 1 and Proposition 2, highlight the role of MPA-independence. Theorem 3 characterizes AGL representation with the addition of NOEL-mixture consistency (A6) and BA-antisymmetry (A7).

**Theorem 1.** (SEU Representation for \( \succ \) restricted to \( \mathcal{F}^E \)) For any \( E \in \mathcal{E} \), let \( \succ_E \) be the restriction of \( \succ \) to \( \mathcal{F}^E \). The following conditions are equivalent:

1. \( \succ \) satisfies A1-A4, and MPA-independence (A5).

2. There exists an affine function \( U_E : \mathcal{L} \to \mathbb{R} \), and a probability distribution \( \mu_E \), such that for \( f, g \in \mathcal{F}^E \),

\[
\begin{align*}
    f \succ_E g \iff \int_S (U_E \circ f) \mu_E(ds) & \geq \int_S (U_E \circ g) \mu_E(ds) \\
    \text{E}_{\mu_E}[U \circ f] & \geq \text{E}_{\mu_E}[U \circ g]
\end{align*}
\]

(a) \( U_E \) is unique up to positive affine transformations.
(b) \( \mu_E : 2^S \to [0, 1] \) is a unique probability distribution over \( S \), where \( \mu_E(s) > 0 \) for all \( s \).

Theorem 1 shows the implication of MPA-independence. The axiom states that along mutually aligned acts, independence is preserved. Hence for the subspace of all acts with an alignment represented by \( E, \mathcal{F}^E \), the DM behaves as a SEU maximizer with the unique prior \( \mu_E \). Mixing acts that have the same alignment keeps the proportion of gains and losses the same therefore the behavior maintains the linearity associated with expected utility. Notice that if the DM has asymmetric gain-loss preferences and the only choice data available is for acts that share one particular alignment, then the analyst might conclude that the DM is a SEU maximizer, with belief \( \mu_E \), and will not be able to identify any gain-loss attitudes. To be able to identify reference-dependence in this setting requires acts with different alignments.

Since constants can be aligned with any events, i.e. \( \mathcal{F}_c \subset \mathcal{F}^E \) for all \( E \in \mathcal{E} \), the representations for \( \{ \succsim^E \} \subset \mathcal{E} \) can be used to find a representation for \( \succsim \). Therefore preferences over \( \mathcal{F} \) can be represented with the family of functionals \( \{ V^E(\cdot) \} \subset \mathcal{E} \), where each \( V^E(\cdot) \) represents \( \succsim^E \) (from Theorem 1). Preferences over constants are the same across all alignments, hence the utility function \( U^E \) can be normalized to be the same for all families of mutually aligned acts, i.e. \( U^E = U^{E'} \) for all \( E, E' \in \mathcal{E} \).

**Proposition 2.** The following conditions are equivalent:

1. \( \succsim \) satisfies A1-A5.
2. There exists an affine function \( U : \mathcal{L} \to \mathbb{R} \), and a set of probability distributions over \( S \), \( \{ \mu^E \} \subset \mathcal{E} \), such that for \( f \in \mathcal{F}^E \) and \( g \in \mathcal{F}^F \),
   \[
   f \succsim g \iff \int_S (U \circ f) \mu^E(ds) \geq \int_S (U \circ g) \mu^F(ds)
   \]
   Moreover \( U \) is unique up to positive affine transformations, and the set \( \{ \mu^E \} \subset \mathcal{E} \) is unique.

Every prior in \( \{ \mu^E \} \subset \mathcal{E} \) is different and MPA-independence does not imply any structure on the priors. To derive the main result from the representation of Proposition 2, NOEL-mixture consistency and BA-antisymmetry are used to guarantee that every prior in the set \( \{ \mu^E \} \subset \mathcal{E} \) can be written as functions of one unique prior \( \mu \). NOEL-mixture consistency implies that there is a unique prior that generates the conditional distributions of \( \mu^E \), for all events that are non-overlapping with \( E \). Furthermore, BA-antisymmetry implies that the difference between \( \mu^E \) and the baseline prior \( \mu \), is the same as the difference between \( \mu^{E'} \) and \( \mu \). This leads to the main representation result of this paper.
Theorem 3. The following conditions are equivalent


2. There exists an affine function $U : \mathcal{L} \to \mathbb{R}$, a probability distribution $\mu : 2^S \to [0, 1]$, and a real number $\lambda = \lambda_g + \lambda_l$ such that

$$f \succeq g \iff \int_S \left( (U \circ f) - \eta^\lambda_{\mathcal{N}}(U \circ f) \right) \mu(ds) \geq \int_S \left( (U \circ f) - \eta^\lambda_{\mathcal{N}}(U \circ f) \right) \mu(ds)$$

where $\eta^\lambda_{\mathcal{N}} : (U(\mathcal{L}))^{[S]} \times \Delta(S) \to \mathbb{R}$ is given by

$$\eta^\lambda_{\mathcal{N}}((U \circ f), \mu)(s) = \begin{cases} 
\lambda_l (U(f(s)) - E_\mu[U \circ f]) & \text{if } U(f(s)) \geq E_\mu[U \circ f] \\
\lambda_g (E_\mu[U \circ f] - U(f(s))) & \text{if } U(f(s)) < E_\mu[U \circ f]
\end{cases}$$

Moreover

(a) $\mu$ is a unique probability distribution.

(b) $\lambda = \lambda_g + \lambda_l \in \mathbb{R}$ is unique, and $|\lambda| < \min_{s \in S} \left( \frac{1}{\mu(s)} \right)$.\(^{20}\)

(c) $U$ is unique up to positive affine transformations.

The parameter $\lambda$ characterizes gain-loss asymmetry uniquely. It is impossible to identify a gain weight and a loss weight uniquely. For instance for the cases where $|\lambda_l| = |\lambda_g|$, the behavior will be the same as a SEU maximizer. So an analyst cannot distinguish some DM who has gain-loss considerations that are symmetric (gains and losses are weighted equally) from a DM who is a SEU maximizer from choice behavior. The intuition for the result is natural since the evaluation of gains and losses is not absolute. The parameter $\lambda$ captures the difference between the weight placed on gains and the weight placed on losses, which for the representation is unique. An important application of the representation result from Theorem 3 is that it provides an index for reference dependence ($\lambda$) that is completely decoupled from risk attitudes, and can be easily estimated and applied because it is a 1-dimensional parameter.\(^{21}\)

\(^{20}\)Strict monotonicity puts a bound on gain-loss attitudes, since an act is never be deemed worse than the outcome on the worst possible state.

\(^{21}\)For situations of choice under risk, Köbberling and Wakker [2005], characterize an similar index in the context of “loss aversion” for choices under risk based on the cumulative prospect theory Tversky and Wakker [1993], Chateauneuf and Wakker [1999] representation. More recently, Ghossoub [2012] provides
4.3.1 Sketch of the Proof

The result is proven in several steps, building on the SEU representation on $\mathcal{F}^E$ (Theorem 1), and the aggregation of these preferences across families of mutually aligned acts (Proposition 2).

First, adding NOEL-mixture consistency guarantees that for all $E \in \mathcal{E}$, the conditional distributions are the same for all $\mu_E$ when conditioned on any $F$ which is non-overlapping with $E$. Hence for any $E, E' \in \mathcal{E}$, $\mu_E(\cdot|F) = \mu_{E'}(\cdot|F)$ whenever $F$ is non-overlapping for $E$ and $E'$. Moreover, the axioms imply that there exists a unique distribution $\mu \in \Delta(S)$ that generates these conditionals. Moreover this $\mu$ is the unique distribution that represents $e_f$ for all $f \in \mathcal{F}$. In other words, if $e_f$ is the hedge of $f$ then $\int_S (U \circ f) \mu(ds) = U(e_f)$. This implies that for any $E \in \mathcal{E}$, $\mu_E$ can be written as a constant distortion $\gamma_E^+$ of $\mu$ on all $s \in E$ (positive states), and a constant distortion $\gamma_E^-$ of $\mu$ on all $s \in E^c$ (negative states), i.e.

$$
\mu_E(s) = \begin{cases} 
\gamma_E^+ \mu(s) & \text{if } s \in E \\
\gamma_E^- \mu(s) & \text{if } s \in E^c 
\end{cases}
$$

Then BA-antisymmetry implies a particular relationship between distributions indexed by complementary alignments, which is that for any $E, F \in \mathcal{E}$,

$$
\mu_E + \mu_{E^c} = \mu_F + \mu_{F^c}
$$

Using this observation we show that of $s \in E$ the distributions on $\mu_E$ and $\mu_{E\setminus s}$ depend only on $s$. So whenever $s \in E \cap F$ (and $s \neq F$ or $s \neq E$),

$$
\mu_E - \mu_{E \setminus s} = \mu_F - \mu_{F \setminus s}
$$

Then, using these conditions about the family $\{\mu_E\}_{E \in \mathcal{E}}$ we show that the difference between the distortion on negative states and positive states is always constant, thus $\gamma_E^- - \gamma_E^+ = \lambda$ for all $E \in \mathcal{E}$. Therefore it is possible to can characterize any $\mu_E$, as functions of $\mu$, and this constant $\lambda$ that captures the difference between the negative and positive distortions. That is,

$$
\mu_E(s) = \mu(s) (1 - \lambda (1 - \mu(E))) \quad \text{if } s \in E \\
\mu_E(s) = \mu(s) (1 + \lambda (\mu(E))) \quad \text{if } s \in E^c
$$

a behavioral characterization of absolute and comparative loss aversion of decision-making which also provides indices of loss aversion that just depend on the functional that represents the preferences without the assumptions of any model.
Finally this representation of $\mu_E$ is used to rewrite the representation from Proposition 2 in terms of $\mu$, which yields the desired result.

5 Gain-Loss Attitudes and Ambiguity Attitudes

The presence of reference dependence contaminates the perceived beliefs of the DM. This contamination of the beliefs is represented by the family of distributions $\{\mu_E\}_{E \in \mathcal{E}}$ from Proposition 2, where each $\mu_E$ is a distortion of the original prior $\mu$. This section discusses the structure of the contamination of the beliefs, and the relationship between asymmetric gain-loss preferences and attitudes towards uncertainty.

As previously discussed, MPA independence implies certainty independence because constant acts are mutually aligned with all acts (every state is neutral for a constant act). Certain independence, along with a condition called uncertainty aversion, which requires that for all $f, g$ such that $f \sim g$ for any $\alpha \in (0, 1)$, $\alpha f + (1 - \alpha)g \succeq f$, are the key conditions that characterize the Gilboa and Schmeidler [1989] Maxmin Expected Utility (MMEU) representation. The MMEU utility representation is characterized by a utility function $U$ and a set of priors $C$, such that

$$f \succeq g \iff \min_{\mu \in C} \int_S (U \circ f)\mu(ds) \geq \min_{\mu \in C} \int_S (U \circ g)\mu(ds)$$

If uncertainty aversion is changed for uncertainty seeking preferences, which states that for all $f, g \in \mathcal{F}$, $f \sim g$ implies for all $\alpha \in (0, 1)$ $f \succsim \alpha g + (1 - \alpha)g$, then the representation is a Maxmax representation.\footnote{Instead of taking the minimum over the whole set of priors, the DM evaluates an act by the prior that maximizes the expectations.} Since asymmetric gain-loss preferences satisfy certainty independence, a natural question to ask is whether there is a relationship between these preferences and attitudes towards uncertainty from the multiple priors model of Gilboa and Schmeidler [1989].

The AGL representation does not explicitly require preferences to be uncertainty averse or uncertainty seeking, nonetheless there is a close relationship between the Maxmin and Maxmax representations and the AGL representation. The relationship is captured by the following result, which states that asymmetric gain-loss preferences always admit a Maxmin or Maxmax representation where the set of priors $C$ has a specific structure that is related to the contamination of beliefs. The next result establishes the link between uncertainty attitudes and gain-loss attitudes with an equivalent representation result for
The AGL representation \((U, \mu, \lambda)\).

**Theorem 4.** Suppose \(\succeq\) admits an AGL representation \((U, \mu, \lambda)\). Then there exists a closed and convex set \(C\) of probability distributions on \(S\) such that

\[
\text{(i.) If } \lambda \geq 0 \\
\quad f \succeq g \iff \max_{\nu \in C} \int_S (U \circ f) \nu(ds) \geq \max_{\nu \in C} \int_S (U \circ g) \nu(ds)
\]

\[
\text{(ii.) If } \lambda \leq 0 \\
\quad f \succeq g \iff \min_{\nu \in C} \int_S (U \circ f) \nu(ds) \geq \min_{\nu \in C} \int_S (U \circ g) \nu(ds)
\]

Here \(C\) is the unique convex set \(C\) defined by \(C = \text{conv}(\{\nu_E\}_{E \in \mathcal{E}})\), where for every \(E \in \mathcal{E}\) and all \(s \in S\)

\[
\nu_E(s) = \begin{cases} 
\mu(s) \left(1 - \lambda \mu(E^c)\right) & \text{if } s \in E \\
\mu(s) \left(1 + \lambda \mu(E)\right) & \text{if } s \in E^c
\end{cases}
\]  

(5.1)

**Proof.** In appendix. \(\square\)

Theorem 4 has several implications. Firstly, it shows that this form of reference-dependence is always tied to a particular attitude towards uncertainty. The preferences studied in this paper will always be either uncertainty averse or uncertainty seeking. Secondly, it gives a precise form to the belief contamination that takes place when gain-loss consideration affect a probabilistically sophisticated DM. The belief contamination keeps the relative likelihood of states among gains and among losses unchanged as the original prior; however depending on the sign of \(\lambda\), it increases or decreases the total weight given to gains (and losses) proportionally to the baseline belief.\(^{23}\) This distortion is a function of the degree of reference dependence, \(\lambda\), and the total probability that the DM assigns to gains and losses for each case. Whenever \(\lambda < 0\), then losses are weighted more than gains, hence the DM will behave as if she is a SEU maximizer who always thinks losses are more likely than “she originally thought” (i.e. the probability given by \(\mu\)). When \(\lambda > 0\) she will behave as someone who thinks losses are always less likely than they actually are.\(^{24}\) Finally, the structure given to the set of priors makes it possible to establish a comparative notion of reference-dependence that is related to the comparative notion of ambiguity aversion of Ghirardato and Marinacci [2002] even when

\(^{23}\) In a similar exercise Grant and Kajii [2007], provide a link between special case Maxmin EU which is called the \(\epsilon\)-Gini contamination and the classical mean-variance preferences.

\(^{24}\) This suggests that there is an informal relationship between gain-bias and optimism, and loss-bias and pessimism.
DMs have different baseline priors and sets of beliefs. The relationship between the AGL representation and comparative ambiguity attitudes is explored in section 6.

6 Comparative Statics

This section provides comparative statics results relating behavior to elements of the AGL representation. The first result provides a characterization of gain-biased and loss-biased preferences as absolute notions of reference dependence. The second result provides comparative notions of reference dependence with a characterization of “more gain-biased” and “more loss-biased” preferences.

Recall that whenever there are gain-loss considerations the certainty equivalent of an act, does not provide all the information required to identify beliefs. Instead, it is the hedge of an act what includes the information necessary for the identification of beliefs. Therefore a natural measure for the degree and direction of reference effects is the gap between \( e_f \) and \( c_f \), the hedge and the certainty equivalent. In the standard SEU model, \( e_f \) and \( c_f \) are the same, so using the gap between \( c_f \) and \( e_f \), establishes the SEU model as the baseline case for reference effects. This suggests a natural definition of gain-biased preferences as those where the hedge is always (weakly) dispreferred to the certainty equivalent, and loss-biased preferences as preferences where the hedge is always (weakly) preferred to the certainty equivalent.

**Definition 7.** Let \( \succsim \) be a preference over \( F \). Say \( \succsim \) is gain-biased if for all \( f \in F \), \( c_f \succsim e_f \). Say \( \succsim \) is loss-biased if for all \( f \in F \), \( e_f \succsim c_f \).

For a DM with AGL preferences, the parameter \( \lambda \) captures the bias of the preferences uniquely. In addition, this bias towards gains or losses is related to attitudes towards ambiguity (see Theorem 4). Therefore for a DM with preferences that admit an AGL representation, gain-bias is equivalent to uncertainty seeking preferences, and loss-bias is equivalent to uncertainty averse preferences. This observation establishes a clear connection between the idea of “loss aversion” that has been prevalent since Prospect Theory, and uncertainty aversion.

**Proposition 5.** Let \( \succsim \) admit an AGL representation. Then the following are equivalent:

1. \( \succsim \) is gain-biased.
2. \( \succsim \) is uncertainty seeking.
3. $\lambda \geq 0$.

The following are equivalent:

1. $\succsim$ is loss-biased.
2. $\succsim$ is uncertainty averse.
3. $\lambda \leq 0$.

Proof. This follows directly from the representations of Theorems 3 and 4.

Beliefs influence the assessment of $f$, captured by $c_f$, and also the expectations of $f$, captured by the hedge $e_f$. If we want to be able to compare two DM’s degree of reference dependence who can hold different beliefs we want to disentangle reference dependence from beliefs. To do this, we define two acts, the join (maximum) and the meet (minimum) of a balanced pair $(f, \bar{f})$, and show that any balanced pair, its join, and its meet, provide information to distinguish the reference effects from beliefs. For any act $f \in \mathcal{F}$, let $f \lor \bar{f} \in \mathcal{F}$ be the act that gives the DM the best outcome between $f$ and $\bar{f}$ for each $s \in S$, and $f \land \bar{f}$ the act that gives the worst outcome between $f$ and $\bar{f}$ for all $s \in S$.

Definition 8. Given any balanced pair $(f, \bar{f})$, define the act $f \lor \bar{f}$, the join of $(f, \bar{f})$ as

$$(f \lor \bar{f})(s) = \begin{cases} f(s) & \text{if } f(s) \succsim \bar{f}(s) \\ \bar{f}(s) & \text{if } \bar{f}(s) \succ f(s) \end{cases}$$

Define the act $f \land \bar{f}$, the meet of $(f, \bar{f})$, as

$$(f \land \bar{f})(s) = \begin{cases} \bar{f}(s) & \text{if } f(s) \succsim \bar{f}(s) \\ f(s) & \text{if } \bar{f}(s) \succ f(s) \end{cases}$$

From the AGL representation, gain-loss utility depends on how much the act deviates state by state from $e_f$. The meet and the join, $f \lor \bar{f}$ and $f \land \bar{f}$, provide a way to capture how much $f$ deviates from $e_f$ from preferences because these two acts capture the absolute value of the state by state deviations of $f$ from $e_f$.\footnote{That is $(f \lor \bar{f})(s) \succsim e_f$ for all $s$ and $(f \land \bar{f})(s) \precsim e_f$ for all $s$.} To determine the subjective assessment of these deviations for each DM (which is given by the beliefs), we consider the hedge of the meet and the join, $e_{f \lor \bar{f}}$ and $e_{f \land \bar{f}}$. Since we know that $e_f$ is determined exclusively by beliefs, these two constant acts, $e_{f \lor \bar{f}}$ and $e_{f \land \bar{f}}$, provide way to measure how large is the deviation for each act for a DM.
To capture reference-dependence behaviorally we need to focus on acts that have the same hedge (expectation) for both DMs. This is the only way to isolate reference effects across DMs from preferences: if acts have different hedges the reference effects can be confounded by the beliefs. When acts have the same hedge, $e_{f\lor f}$ provides a positive measure of the deviation of $f$ from $e_f$.\(^{26}\) Likewise, and $e_{f\land f}$ provides a negative measure of the deviation of $f$ from $e_f$. So we can use these measures of deviation to compare gain-loss attitudes of DMs similar to the comparative notion of ambiguity aversion from Ghirardato and Marinacci [2002].

The intuition behind the definition is that conditional on having the same hedge, a DM with gain-bias prefers an act $f$ with larger deviations from $e_f$ because the expected gains are larger. Conversely, a DM with loss-bias prefers acts with smaller deviation condition on having the same hedge. Since we want to consider acts that have the same hedge, the comparative notions of “more gain-biased” and “more loss-biased” depends on a possibly different act for each DM, $f$ for DM 1 and $g$ for DM 2.\(^{27}\) If the hedge of $f$ for DM 1 and the hedge of $g$ for DM 2 are the same, DM 1 is more gain-biased than DM 2 if the fact that $e_{f\lor f}$ is larger than $e_{g\lor g}$ provides a sufficient information to conclude that DM 1 is willing to pay more for $f$ than DM 2 is willing to pay for $g$. Here $e_{f\lor f}$ provides a measure for the size of the deviation of $f$ from $e_f$ for DM 1, and $e_{g\lor g}$ a measure for the size of the deviation of $g$ from $e_g$ from DM 2. Therefore DM who is more gain-biased prefers to have larger deviation, if the hedge of the acts is the same. Similar to the comparative definition of more gain-biased, we say that DM 1 is more loss biased than DM 2 if the fact that $e_{f\land f}$ is larger than $e_{g\land g}$ always provides sufficient information to conclude that DM 2 is willing to pay more for $g$ than DM 1 is willing to pay for $f$. Therefore the DM who is more loss-biased prefers to have a smaller deviation, conditional on acts having the same hedge. We use the notation where $e_i^f$ denotes the hedge of $f$ for DM $i$.

**Definition 9.** Given two preference orders $\succeq_1$ and $\succeq_2$, say that $\succeq_1$ is *more gain-biased* than $\succeq_2$ if for any $f, g$ with $e_1^f = e_2^g$ and $e_{f\lor f}^{i} \succeq_i e_{g\lor g}^{i}$ for $i = 1, 2$, then for any $c \in \mathcal{F}_c$, $c \succeq_1 f$ implies $c \succeq_2 g$ and $c \succ_1 f$ implies $c \succ_2 g$.

Say that $\succeq_1$ is *more loss-biased* then $\succeq_2$ if for any $f, g$ with $e_1^f = e_2^g$ and $e_{g\land g}^{i} \succeq_i e_{f\land f}^{i}$ for $i = 1, 2$, then for any $c \in \mathcal{F}_c$, $c \succeq_2 g$ implies $c \succeq_1 f$ and $c \succ_2 g$ implies $c \succ_1 f$.

The comparative notion of “more gain bias” or “more loss bias” is similar to the notion

\(^{26}\)This is a positive measure because for all $s \in S$, $f \lor \bar{f}(s) \succeq e_f$.

\(^{27}\)If beliefs are different for most $f \in \mathcal{F}$, the hedge of $f$ for DM 1 is different than the hedge of $f$ for DM 2.
of uncertainty aversion in Ghirardato and Marinacci [2002]. For gain-loss preference the
implication of comparative loss bias is the same as comparative uncertainty aversion,
nonetheless in this case it holds only for acts that have the same hedge. The previous
definition of comparative gain-bias and loss bias provides a behavioral way to compare
reference effects across DMs.

**Theorem 6.** Let $\succsim_i$ admit an AGL representation given by $(U_i, \mu_i, \lambda_i)$ for $i = 1, 2$. If $U_1$ and $U_2$ are cardinally equivalent, then the following are equivalent.

(i) $\succsim_1$ is more gain-biased than $\succsim_2$.

(ii) $\succsim_2$ is more loss-biased than $\succsim_1$.

(iii) $\lambda_1 \geq \lambda_2$.

When DMs have the same belief, then for any $f \in \mathcal{F}$, $e^1_f = e^2_f$ because the belief completely determines the hedge of any act.28 Therefore, two immediate corollaries of Theorem 6 are the following results.

**Corollary 7.** Let $\succsim_i$ admit an AGL representation given by $(U_i, \mu_i, \lambda_i)$ for $i = 1, 2$. If $U_1$ and $U_2$ are cardinally equivalent and $\mu_1 = \mu_2$, then the following are equivalent.

(i) $\succsim_1$ is more gain-biased than $\succsim_2$.

(ii) $\succsim_2$ is more loss-biased than $\succsim_1$.

(iii) For all $f \in \mathcal{F}$ and all $c \in \mathcal{F}_c$, $f \succsim_2 c$ implies $f \succsim_1 c$.

(iv) $\lambda_1 \geq \lambda_2$.

**Corollary 8.** Let $\succsim_i$ admit an AGL representation for $i = 1, 2$. If $\succsim_i$ is gain-biased for $i = 1, 2$, $U_1$ and $U_2$ are cardinally equivalent, and $\mu_1 = \mu_2$, then the following are equivalent.

1. $\succsim_1$ is more gain-biased than $\succsim_2$.

2. $\lambda_1 \geq \lambda_2$.

3. $C_2 \subseteq C_1$ for the equivalent Maxmin/Maxmax representation from Theorem 4.

If $\succsim_i$ is loss-biased for $i = 1, 2$, $U_1$ and $U_2$ are cardinally equivalent, and $\mu_1 = \mu_2$, then the following are equivalent.

---

28See the proof of Theorem 3
1. $\succsim_1$ is more loss-biased than $\succsim_2$.

2. $\lambda_2 \geq \lambda_1$.

3. $C_2 \subseteq C_1$ for the equivalent Maxmin/Maxmax representation from Theorem 4.

The implication of Corollary 7 is that when DMs have the same prior, loss bias is equivalent to the comparative notion of ambiguity aversion from Ghirardato and Marinacci.\(^{29}\) This is because the priors determine the hedge of an act and also the meet and the join, hence $e^1_f = e^2_f$ for all $f \in F$. Corollary 8 implies that whenever $\succsim_i$ is gain-biased or loss-biased for both DMs, the notion of loss bias is consistent with the representation of comparative ambiguity aversion derived from Gilboa and Schmeidler [1989]. This notion states that on a Maxmin representation the more ambiguity averse DM should have a larger set of priors.

These comparative statics results establish an unexplored link between the absolute and comparative notions of gain or loss bias, and existing notions of uncertainty aversion which is worth further exploring. The initial motivation for studying uncertainty was due to the Ellsberg [1961] idea that DMs are not able to formulate unique probabilities over uncertain events. Many models with multiple priors have been developed to capture what is considered as “Ellsbergian behavior”. Nonetheless even if the DM is able to form a unique prior, having gain-loss considerations can appear to contaminate her prior in a way that gives rise to behavior embodied by some multiple priors model. Hence, for AGL preference a probabilistically sophisticated DM can appear to have multiple priors due to gain-loss asymmetry.

7 Discussion

This section discusses the relationship between this paper and the literature on three areas related to asymmetric gain-loss preferences: reference point determination, disappointment aversion, and dispersion preferences.

7.1 Endogenous Reference Point Determination

There are a few other papers that study reference points that are endogenously identified within the model like the AGL representation. However, the AGL model has two important features that no other model of reference point determination has. First, it uniquely

\(^{29}\)This is the definition (iii) in the Corollary.
identifies a reference point from behavior as the expected utility of the act together with a measure of reference-dependence. Second, the choices do not affect the reference point, which allows for identification without the necessity of an equilibrium condition.

Giraud [2004b] also studies a form of reference-dependence in a model of framing under risk, where the effect of the frame is determined endogenously in the model. In this model, since the reference frame is interpreted as states of mind (subjective state-dependent utility functions like in Dekel et al. [2001] and Dekel et al. [2007]), it is not possible to uniquely identify the reference point from the reference effect.

A similar problem is present in the work of Ok et al. [2011]. The authors consider a general choice setting and characterize choice problems where the DM uses a reference point, even for instances where there is no uncertainty. However, given that their framework is so general (choice functions), they cannot provide uniqueness results. Ok et al. [2011] can identify uniquely situations susceptible to reference dependence, but they show that the reference map that can rationalize the choices is not unique.

Közsegi and Rabin [2006] provide to my knowledge the first model of reference dependence that tackles the determination of an endogenous reference point. Although the general idea behind the functional representation in this paper (4.1) follows KR, the two models are conceptually different. The main difference, previously mentioned, is that the choices of the DM do not affect the reference point. This makes it possible to identify a reference point without the need to appeal to equilibrium notions, where the reference point is induced by the choices, and the choices are optimal given the reference point they induce. This notion is instrumental for the determination of a personal equilibrium in KR. Personal equilibrium states that conditional on a reference point, the choices of the DM are optimal, and that the distribution over optimal choices induce a reference point.

Besides this key conceptual issue, the fundamental technical difference between the two models is the domain of uncertainty in each model and how beliefs about the uncertainty affect the reference point and choices. In KR model, the uncertainty about the distribution over possible choice problems, whereas here uncertainty is over states that determine the outcomes of acts.

In the KR model the reference point (“recent expectations”) is determined by the probabilistic beliefs of the DM about the choice sets she will face, and the decision she will make for each choice set. The reference point is determined by the distribution of optimal choices induced by the beliefs about the choice sets. The formation of expectations is central in the KR model, but how the beliefs that determine expectations are formed is not part of their model. Identifying beliefs about the possible choice problems that a
DM might face poses a challenge from a theoretical and empirical point of view. If the DM is uncertain about the problems she might face, she must form expectations about the environment long before seeing the choice set, let alone making any choices. There is no standard choice setting that allows for the identification of these beliefs. Any observed choice behavior is always tied to a particular choice set so it is not possible to identify beliefs about facing that choice set. Any observed choice for a DM in the KR framework will be the choice given a reference point, which is the choice itself by the equilibrium requirement. The feedback between choices and the reference point, where reference point is always the chosen alternative, can lead to intransitivity of the revealed preference when considering choices [Gul and Pesendorfer, 2006]. Therefore under standard choice settings beliefs about the possible choice sets (and hence the endogenous reference point) cannot be independently identified from behavior.

Finally, given that choices do not affect the reference point in the AGL model there is only one reference point for each act, which is uniquely determined by beliefs and preferences. In KR the reference point is uniquely determined by the beliefs about the optimal choices in each potential problem, and the expected distribution over optimal outcomes is the unique stochastic reference point for all problems.

Sarver [2011] tackles the problem of endogenous reference point determination, in a dynamic setting where the DM has reference-dependence preferences and she knows the extent of how choices at each period affect the future reference point. In Sarver [2011], the reference point is a choice variable, and it requires, like Kőszegi and Rabin [2006], a notion of equilibrium since choices affect the reference point.

### 7.2 Disappointment Aversion

Gul [1991] provides a model of preferences over lotteries (risk) that exhibit disappointment aversion. Gul characterizes lotteries by the elation disappointment decomposition, which separates outcomes for each lottery into these two categories. Elating outcomes are outcomes that are better than the certainty equivalent of the lottery, and disappointing outcomes are outcomes that are worse.

Gul’s main axiom is a relaxation of independence which states if an outcome does not chance from elating to disappointing in two lotteries, then independence is preserved. This property shares a similar flavor to the mutual alignment independence axiom from this paper. An important difference between the two approaches is that in Gul’s work, the definition of elation and disappointment decomposition of a lottery is based on the certainty equivalent of the lottery. Those outcomes that are better than the certainty equivalent are
elating and those that are worse are disappointing. In Gul’s model is possible to use the certainty equivalent of a lottery to determine elating and disappointing outcomes because all the information is objective and observable to the analyst. As explained in section 3, the certainty equivalent of an act does not provide sufficient information to separate outcomes in each state according as gains and losses. The effect of beliefs on gain-loss utility can imply that for some act, a positive state gives an outcome that is worse than the certainty equivalent. Therefore since subjective beliefs determine gains and losses, the separation of gains and losses for AGL preference is different than in Gul [1991] where all the information is objective.

Nonetheless in the model presented in this paper, the preference over objective lotteries satisfy independence, so there is no disappointment aversion in the sense of Gul [1991]. Unlike in Gul’s model where the preferences over outcomes completely determine the disappointment aversion from the comparison of outcomes of an objective lottery to the certainty equivalent of the lottery. In this model the gain-loss attitudes are determined by beliefs as well as preferences over outcome, therefore for objective lotteries in this model there is no disappointment (i.e. preferences satisfy independence) to be able to meaningfully capture the difference between subjective gains and losses. So the AGL representation can be thought of as a subjective version of Gul [1991] disappointment aversion model, where the deviation from independence are due only to uncertainty and not at all to risk. In addition, Both models also share the property that the deviation from the standard model is captured by just one parameter, which makes the models tractable and easy estimate and use in applied work. The parameters $\beta$, in Gul’s model, and $\lambda$, in this model, completely capture attitudes toward disappointment and gains and losses.

### 7.3 Reference Dependent Subjective Expected Utility

The other approach to capture reference dependence in choice under uncertainty in a way that does not require an equilibrium condition is given by the work of Sugden [2003]. Sugden also studies preferences that measure the expected utility deviation of an act from the reference act. Sugden studies reference dependence in a SEU model, but not the identification of the reference point. The reference point is exogenously given, and not determined within the model.

He derives a representation for all pairs of acts given a third act as a reference point. Preferences in Sugden [2003] are defined over triples of acts, where triple $(f,g,h) \in \mathcal{F} \times \mathcal{F} \times \mathcal{F}$ is interpreted as $f$ is preferred to $g$ when $h$ is the reference point. He derives
necessary and sufficient conditions for a representation where \( f \succsim g \mid h \) if and only if 
\[ E_p[v(f, h) - v(g, h)] \geq 0, \]
p is a unique probability distribution and \( v \) is a relative value function which takes the act \( h \) as a reference act. His focus on the representation is the value function and the probability distribution, which like in this paper is considered as the unique beliefs of the DM about the uncertainties.

Sugden works on a Savage [1954] domain, where the state space is infinite and more importantly there is no mixing operation. The axiomatization in Sugden [2003] is mostly based on many of the insights and results of Regret Theory [Sugden, 1993, Loomes and Sugden, 1982]. Therefore his axiomatization is completely different to the one on this model and it is not to evident how to draw a parallel between the two.

### 7.4 Deviation Preferences

For asymmetric gain-loss preferences, the gains and losses are assessed as deviations from the expected utility of the act. A branch of the literature on choice under uncertainty studies dispersion preferences, which depend on the deviations from the mean of acts.

Grant and Polak [2011] present a general model of dispersion preferences. Their model is characterized by the triple \( \{U, \pi, \rho\} \), where \( U \) is a utility function, \( \pi \) is a distribution over \( S \), and \( \rho \) is a dispersion measure. The utility representation is given by

\[
V(f) = E_{\pi}(U \circ f) - \rho(d(f, \pi))
\]

where \( d(f, \pi) \) is the vector of utility deviations from the mean, and \( \rho \) is a function that aggregates the deviation. The main condition of the representation is called constant absolute risk aversion (CAUA), which states that for every \( f \in \mathcal{F} \), and \( x, y, z \in \mathcal{F}_c \),
\[
\alpha f + (1 - \alpha)x \succsim \alpha z + (1 - \alpha)x \text{ implies } \alpha f + (1 - \alpha)y \succsim \alpha z + (1 - \alpha)y.
\]
This is a weak condition which is implied by some recurrent weakening of the independence axiom from the literature like comonotonic independence, certainty independence, and weak certainty independence. Hence the representation includes many well known families of preferences such as Choquet EU [Schmeidler, 1989], Maxmin EU [Gilboa and Schmeidler, 1989], invariant biseparable preferences [Ghirardato et al., 2004], and variational preferences [Maccheroni et al., 2006]. Since the independence relaxation of this paper is stronger than certainty independence, the class of asymmetric gain-loss preferences will be a subset of the mean-dispersion preferences from Grant and Polak [2011]. The dispersion function \( \rho \) for the asymmetric gain-loss preferences is given by
\[
\rho(d(f, \pi)) = \lambda E_{\pi}[\eta^\lambda_{\eta}(U \circ f)]
\]
where \( \eta \) is defined as in (4.2).
The model presented by Grant and Polak does not identify the components from the representation uniquely. In particular, any $\pi$ is associated with the mean-dispersion representation of preference, and $\rho$ is never unique since no conditions are required. Grant and Polak show that further restrictions on the preferences can identify properties of $\rho$ such as non-negativity, convexity, linear homogeneity and symmetry. To identify a unique $\pi$ it is necessary to add a convexity condition, and a local smoothness condition.\[^{30}\] The local smoothness condition ensures that around certainty, preferences are locally approximated by either SEU or MMEU preferences. There is a similarity between Grant and Polak’s smoothness conditions and NOEL mixture consistency in the sense that they guarantee on the existence of some open set of $\alpha$’s such that an independence-like condition holds. Nonetheless the conditions have different implications. The condition from Grant and Polak [2011] is used to approximate preferences around certainty with a SEU, whereas NOEL-mixture consistency is used to guarantee that the perceived contamination of beliefs is consistent with the baseline prior.

A particularly relevant subset of the mean-dispersion preferences, is the vector expected utility (VEU) of Siniscalchi [2009]. The intuition behind the VEU model is that DMs cannot assign probabilities easily, so they set an “anchor”, which is the expected utility of the act, then a set of adjustment factors determine how preferences deviate from this anchor.

Siniscalchi [2009] characterizes VEU by the tuple $(u, p, n, \zeta, A)$, where $u$ is a utility function over outcomes, $p$ is a baseline prior (the “anchor”), $n$ is the number of adjustment factors, and $\zeta$ is the actual adjustment for each factor, and $A$ is how those factors are aggregated. The interaction between $A$ and the set of adjustment factors, lead to some non-uniqueness in the VEU model. His main axioms are complementary act independence, and complementary translation invariance. Complementary translation invariance requires that for all complementary pairs $(f, f')$, and any constants $x, x'$, if $f \sim x$ and $f' \sim x'$ then $\frac{1}{2} f + \frac{1}{2} x' \sim \frac{1}{2} f' + \frac{1}{2} x$.\[^{31}\] Both conditions are implied by the mutual MPA-independence and balanced act symmetry axioms in this paper because complementary acts from Siniscalchi [2009] are always aligned with complementary events in this paper. Hence it follows that the asymmetric gain-loss preferences are also a subset of the set of preferences that admit a VEU representation.

Chambers et al. [2012] also study a class of preferences that depart from CAUA, called

\[^{30}\]Either convexity, or weaker version of convexity, called preference for perfect hedges which was developed by Chateauneuf and Tallon [2002]. Or alternatively complementary independence, which is discussed in the Siniscalchi [2009].

\[^{31}\]Complementary independence is explained in section 3.
invariant symmetric preferences. Their analysis is motivated by the tradeoff between risk and expected return of assets, and the seminal idea that DMs cannot assign probabilities to events. These preferences are characterized as a tuple consisting of \((U, \pi, \rho, \varphi)\), where \(\rho\) is a dispersion measure and \(\varphi\) determines how the dispersion measure and the expected utility (according to \(\pi\) and \(U\)) are aggregated, where \(\rho\) is a symmetric function. The identification of the baseline probability measure is due to complementary independence following Siniscalchi [2009], which leads to the definition of the “mean” of an act. Then the novel conditions Chambers et al. impose on preferences, which are relaxations of independence, depend on “common-mean” acts. These are acts that share the same mean according to the baseline probability \(\pi\). The conditions are common-mean independence and common mean uncertainty aversion. The property that the asymmetric gain-loss preferences might fail, is common-mean uncertainty aversion. The asymmetric gain-loss preferences can either be uncertainty averse or uncertainty loving, and do not need to be averse to “dispersion”.

Siniscalchi [2009], Grant and Polak [2011], and Chambers et al. [2012] focus on the domain of choice under uncertainty and study attitudes towards variation. This paper is a special case of the general representations in Siniscalchi [2009] and Grant and Polak [2011]. Hence by the result in Grant and Polak [2011], the family of AGL preferences that satisfy gain-loss aversion will be part of the variational family of preferences from Maccheroni et al. [2006] as well. The gain-loss loving preferences will not be variational since they do not satisfy uncertainty aversion.\(^{32}\) The benefit of this model of asymmetric gain-loss preferences is that despite being less general than Grant and Polak [2011] and Siniscalchi [2009], it is possible to derive meaningful comparative statics results, as shown in section 6.

Theorem 4 shows that the AGL representation exhibits a Maxmin or a Maxmax form, depending on the sign of \(\lambda\). A natural question is what is the relationship between gain-loss attitudes, and the general \(\alpha\)-Maxmin representation from Ghirardato et al. [2004]. This is an open question. In the domain of this paper, a partial answer is provided by Eichberger et al. [2011] which show for that \(\alpha\)-MMEU on a finite state space, Ghirardato et al. [2004] axioms are satisfied for a constant \(\alpha\) if and only if \(\alpha = 1\) and \(\alpha = 0\), which are equivalent to Maxmin and Maxmax representations.

\(^{32}\)Asymmetric gain-loss preferences are not part of the family of multiplier preferences of Hansen and Sargent [2001], since they do not satisfy the sure thing principle. See Strzalecki [2011] for an axiomatization of multiplier preferences.
8 Conclusion

This paper provides a model of reference dependent preferences in the domain of choice under uncertainty, which is characterized by the family of preferences called asymmetric gain-loss preferences. This model provides a behavioral way to identify a unique reference point, and uniquely capture reference-dependent attitudes from choice behavior. The identification of the reference point has the feature that the choices do not affect the reference point. Hence this model provides a simple tool to capture reference dependence from behavior without the need to consider equilibrium conditions which can be problematic to identify from choice data.

For asymmetric gain-loss preferences an act is evaluated according to its expected utility, and gain-loss utility which is measured using the expectations as a reference point. The AGL model is characterized by a triple \((U, \mu, \lambda)\) such that it captures reference-dependence in a way that deviates minimally from the standard SEU model. The only difference between the standard model and the AGL model is the parameter \(\lambda\), which provides all the information about reference-dependence. In addition, the preferences studies in this model, allow for the behavioral comparison of gain-loss attitudes of different DMs who might have different beliefs, and also offer an clear behavioral characterization of gain-bias and loss-bias.

Moreover the AGL model establishes a simple and intuitive link between reference dependence and uncertainty attitudes. Whenever a DM exhibits gain-biased preferences she will be exhibit uncertainty averse preferences, and if she exhibits loss-biased preferences then she must exhibit uncertainty averse preferences. In addition even though the DM in this model has a unique prior over the states, which she uses to assess the reference point, gain-loss asymmetries can contaminate the beliefs in a way that she will behave as if she has multiple priors in mind.
References


A Appendix: Proofs

A.1 Proofs of the Main Results

This section provides proofs for the main results. The proofs for the auxiliary lemmas and propositions are in appendix A.2.

Proof of Theorem 1.

This is an obvious consequence of the Herstein and Milnor [1953] Mixture Space Theorem. Fix $E \in \mathcal{E}$. Since $\succsim$ satisfies MPA-independence, $\mathcal{F}^E$ is convex, and it includes all the constant acts. Therefore $\succsim_E$ and $\mathcal{F}^E$ define a mixture space, so by the Mixture Space Theorem, the conditions for a SEU representation of $\succsim_E$ are satisfied. Therefore there exists a cardinally unique expected utility function $U_E : \mathcal{L} \to \mathbb{R}$, and an unique probability distribution $\mu_E : 2^S \to [0, 1]$, such that for any $f, g \in \mathcal{F}^E$,

$$f \succsim g \iff V_E(f) \geq V_E(g)$$

Where

$$V_E(f) = \int_S (U_E \circ f)\mu_E(ds)$$

By strict monotonicity, $\mu_E(s) > 0$ for all $s$, so every state is non-null.

Proof of Proposition 2.

From Theorem 1, for every $E \in \mathcal{E}$, $\mathcal{F}^E$ admits a subjective expected utility (SEU) representation given by $(U_E, \mu_E)$ [Anscombe and Aumann, 1963]. Since any constant $c \in \mathcal{F}_c$ is in $\mathcal{F}^E$ for all $E \in \mathcal{E}$, every $U_E$ induces the same cardinal ranking of $X$. So it is possible to normalize all representations to have the same utility function $U$. Therefore for every $E \in \mathcal{E}$, preferences over $\mathcal{F}^E$ are represented by $(U, \mu_E)$ for any $E \in \mathcal{E}$. Let $V_E : \mathcal{F}^E \to \mathbb{R}$ be given as $V_E(f) = \int_S (U \circ f)\mu_E(ds)$. For $E \in \mathcal{E}$, and any $c \in \mathcal{F}_c$, $c \in \mathcal{F}^E$ $f \succsim c$ if and only if $V_E(f) = V_E(c) = U(c)$. Since every $f \in \mathcal{F}$ has a certainty equivalence $c_f$, and $\succsim$ is complete and transitive, for any $f \in \mathcal{F}^E$, $g \in \mathcal{F}^{E'}$, $f \succsim g$, if and only if $c_f \succsim c_g$, hence

$$f \succsim g \iff \int_S (U \circ f)\mu_E(ds) \geq \int_S (U \circ g)\mu_{E'}(ds)$$

proving the result.

Proof of Theorem 3.

Start with the representation of Theorem 1 and Proposition 2, which is guaranteed by A1-A5. Hence there is a utility function $U : \mathcal{L} \to \mathbb{R}$ and a set of probability distributions
over $S$, indexed by $\mathcal{E}$, $\{\mu_E\}_{E \in \mathcal{E}}$.

STEP 1: Show that for every $E, E'$ the conditional distributions of $\mu_E$ and $\mu_{E'}$-conditional on any event $F$ which is non-overlapping for $E$ and $E'$, are the same. And show that there is a unique distribution $\mu$ over $S$, that generates all the conditionals. In other words, there exists a unique $\mu$, such that $\mu_E(\cdot|F) = \mu(\cdot|F)$ for all $E \in \mathcal{E}$ and $F$ as long as $F$ is non-overlapping for $E$. This is achieved by the addition of NOEL-mixture consistency (A6).

**Proposition 9.** Suppose $(U, \mu_E)$ is a SEU representation $\succsim_E$ on $\mathcal{F}^E$ for all $E \in \mathcal{E}$. If $\succsim$ satisfies A1-A7, then there exists a unique distribution $\mu : 2^S \to [0, 1]$, such that for any $F, E \in \mathcal{E}$ such that $F \subseteq E$ or $F \cap E = \emptyset$, $\mu_E(\cdot|F) = \mu(\cdot|F)$, where for all $s \in F$, $\mu(s|F) = \frac{\mu_E(s)}{\mu(F)}$.

**Proof.** In appendix A.2. □

By Proposition 9, given $E \in \mathcal{E}$, for any $s, s' \in E$, $\frac{\mu(s)}{\mu(s')} = \frac{\mu_E(s)}{\mu_E(s')}$. This holds if for all $s \in E$, $\mu(s) = \gamma \mu_E(s)$, where $\gamma \in \mathbb{R}_{++}$. So on $E$, $\mu_E$ is just a constant perturbation of the original prior $\mu$. Then the distribution $\mu_E$ can be written in the following way

$$\mu_E(s) = \begin{cases} \gamma_E^+ \mu(s) & \text{if } s \in E \\ \gamma_E^- \mu(s) & \text{if } s \in E^c \end{cases}$$

where $\gamma_E^+$ represents how the original prior is perturbed on the positive states (i.e. $E$), and $\gamma_E^-$ represents how the original prior is modified on the negative states. Both $\gamma_E^+$ and $\gamma_E^-$ are positive numbers from monotonicity of $\succsim$, where $\gamma_E^+ \geq 1 \iff \gamma_E^- \leq 1$ for every $E$ because $\mu_E$ is a probability distribution and the sum need to add up to 1. Therefore distributions if $\mu_E$ and $\mu_{E'}$ agree on the probability assigned to a state they are exactly the same.

**Lemma 10.** Given $E \neq E' \in \mathcal{E}$. Suppose $\mu_E(s) = \mu_{E'}(s)$ for some $s$, then $\mu_E = \mu_{E'}$.

**Proof.** In appendix A.2. □

STEP 2: Adding BA-Antisymmetry, implies some consistency between the distributions induced on $\mathcal{F}^E$ and $\mathcal{F}^{E^c}$. In which the average probability attached to each $s$ is always the same for the pair $\mu_E, \mu_{E^c}$, or in other words the distortions on $E$ and $E^c$ exactly balance out.

**Proposition 11.** Let $\succsim$ satisfy A1-A7, then for any $E, F \in \mathcal{E}$,

$$\mu_E + \mu_{E^c} = \mu_F + \mu_{F^c}$$

**Proof.** In appendix A.2. □

**Lemma 12.** Let $\succsim$ satisfy A1-A7, then for all $E \in \mathcal{E}$,

$$\frac{\mu_E + \mu_{E^c}}{2} = \mu$$

40
Where \( \mu_E \) is the distribution from Theorem 2 that represents preferences over \( \mathcal{F}^E \).

**Proof.** This is an immediate consequence of Propositions 9 and 11. \( \square \)

From Lemma 12, further conclude that \( \gamma^+_E + \gamma^-_{E^c} = 2 \) for all \( E \in \mathcal{E} \). A more relevant implication is that \( \mu \), uniquely defines \( e_f \) for all \( f \in \mathcal{F} \). Recall that \( e_f \) is defined as the constant where \( e_f = \frac{1}{2}f(s) + \frac{1}{2}\tilde{f}(s) \) for all \( s \in S \). Let \( f \in \mathcal{F}^E \) and hence \( \tilde{f} \in \mathcal{F}^{E^c} \).

**Proposition 13.** Let \( \succeq \) satisfy A1-A7, then for every \( f \in \mathcal{F} \), \( e_f \in \mathcal{F}_c \) is an act such that \( U(e_f) = \int_S(U \circ f)\mu(ds) \).

**Proof.** In appendix A.2. \( \square \)

**Step 3:** Show that the distribution induced on \( \mathcal{F}^E \) and \( \mathcal{F}^{E^c} \) does not depend on \( E \). Hence, the difference in measures only depend on the states they do not have in common.

**Proposition 14.** Let \( \succeq \) satisfy A1-A7, then for any \( E,F \in \mathcal{E} \) and \( s \in S \), such that \( S \in E \cap F \),

\[
\mu_E - \mu_{E \setminus s} = \mu_F - \mu_{F \setminus s} \quad (A.1)
\]

**Proof.** In appendix A.2. \( \square \)

An immediate consequence of Proposition 14, is the following extension of the result.

**Lemma 15.** Let \( \succeq \) satisfy A1-A7, then for any \( T \subseteq E \cap F \),

\[
\mu_E - \mu_{E \setminus T} = \mu_F - \mu_{F \setminus T}
\]

**Step 4:** Based on the previous results, provide a characterization of the distortions \( \gamma^+_E \) and \( \gamma^-_E \) (5.1), as functions of \( \mu \) and \( \mu_E \). And show that for any particular \( E \in \mathcal{E} \), the difference between the negative and the positive distortion is always constant.

**Proposition 16.** If \( \succeq \) satisfies A1-A7, then for any \( E,F \in \mathcal{E} \), \( \gamma^-_E - \gamma^+_E = \gamma^-_F - \gamma^+_F \). Where \( \gamma^+_E, \gamma^-_E \) are defined as

\[
\gamma^+_E = \frac{\mu_E(s)}{\mu(s)} \quad \text{for} \ s \in E
\]

\[
\gamma^-_E = \frac{\mu_E(s)}{\mu(s)} \quad \text{for} \ s \in E^c
\]

**Proof.** In appendix A.2. \( \square \)

Therefore \( \gamma^-_E - \gamma^+_E \) is a constant (independence of \( E \)), which can be defined as

\[
\gamma^-_E - \gamma^+_E \equiv \lambda
\]

The next step is to characterize \( \lambda \).
Proposition 17. If \( \gamma_E^- - \gamma_E^+ = \lambda \) for all \( E \in \mathcal{E} \), then for any \( E \in \mathcal{E} \),

\[
\mu_E(s) = \mu(s) (1 - \lambda (1 - \mu(E))) \quad \text{if } s \in E \quad \text{and} \\
\mu_E(s) = \mu(s) (1 + \lambda (\mu(E))) \quad \text{if } s \in E^c
\] (A.2)

Proof. In appendix A.2.

STEP 5: Use the definition of \( \mu_E \) from Proposition 17 into the representation from Proposition 2.

For any \( f \in \mathcal{F}^E \), then \( V(f) = \int_S (u \circ f) \mu_E(s) \) which is equivalent to

\[
\int_S (U \circ f) \mu_E(s) = \int_E (U \circ f) \mu(s) (1 - \lambda (1 - \mu(E)) + \int_{E^c} (U \circ f) \mu(s) (1 + \lambda \mu(E))
\]

\[
= \int_E (U \circ f) \mu(s) - \lambda (1 - \mu(E)) \int_E (U \circ f) \mu(s)
\]

\[
+ \int_{E^c} (U \circ f) \mu(s) + \lambda (\mu(E)) \int_{E^c} (U \circ f) \mu(ds)
\]

\[
= \int_S (U \circ f) \mu(ds) + \lambda \left( \mu(E) \int_E (U \circ f) \mu(ds) - \int_E (U \circ f) \mu(ds) \right)
\]

\[
+ \lambda \left( \int_{E^c} (U \circ f) \mu(ds) - \mu(E^c) \int_{E^c} (U \circ f) \mu(ds) \right)
\]

\[
= \mathbb{E}_\mu[U \circ f] + \lambda \left( \mu(E) \mathbb{E}_\mu[U \circ f] - \int_E (U \circ f) \mu(ds) + \int_{E^c} (U \circ f) \mu(ds) - \mu(E^c) \mathbb{E}_\mu[U \circ f] \right)
\]

\[
= \mathbb{E}_\mu[U \circ f] + \lambda \left( \int_E (\mathbb{E}_\mu[U \circ f] - (U \circ f)) \mu(ds) + \int_{E^c} ((U \circ f) - \mathbb{E}_\mu[U \circ f]) \mu(ds) \right)
\] (A.3)

STEP 6: Show that \( \lambda \) is unique, but the specific weights on gains and losses are not unique. Since representation is the same for any \( \lambda_l, \lambda_g \) and \( \lambda'_l, \lambda'_g \) such that \( \lambda_g + \lambda_l = \lambda'_g + \lambda'_l \). So \( \lambda \) captures everything that can be identified about gain-loss asymmetry.

Let the function \( \eta'_\lambda : \mathcal{U}(\mathcal{L})^{[S]} \times \Delta(S) \to [0, 1]^{[S]} \) be defined as

\[
\eta'_\lambda(v, \mu) = -\lambda |v_s - \mathbb{E}_\mu[v]|
\]

For any \( f \in \mathcal{F}^E \), for all \( s \in E \), \( \mathbb{E}_\mu[U \circ f] - U(f(s)) \leq 0 \) and for all \( s \in E^c \), \( U(f(s)) -
\(E_\mu[U \circ f] \leq 0\). Moreover,

\[
\int_S \left( (U \circ f) - \left( \int_S (U \circ f) \mu(ds) \right) \right) \mu(ds) = 0
\]

Hence (A.3) can be written as,

\[
\int_S (U \circ f) \mu_E(ds) = E_\mu \left[ U \circ f + \frac{\lambda}{2} (U \circ f) - E_\mu[U \circ f] \right]
\]

(A.4)

Moreover, it follows from the representation that \(\eta^\prime\) can be written as \(\eta^\lambda\) as defined in (4.3) for any pair \((\lambda_g, \lambda_l)\), such that \(\lambda_g + \lambda_l = \lambda\). To see this, note that

\[
\int_E ((U \circ f) - E_\mu[U \circ f]) \mu(ds) = \int_{E^c} (E_\mu[U \circ f] - (U \circ f)) \mu(ds)
\]

Therefore for any \(\lambda = \lambda_g + \lambda_l\), whenever \(f \in \mathcal{F}^E\),

\[
\frac{\lambda}{2} E_\mu[(U \circ f) - E_\mu(U \circ f)] = \left( \frac{\lambda_g + \lambda_l}{2} \right) \left( \int_E ((U \circ f) - E_\mu[U \circ f]) \mu(ds) \right) + \left( \frac{\lambda_g + \lambda_l}{2} \right) \left( \int_{E^c} (E_\mu[U \circ f] - (U \circ f)) \mu(ds) \right)
\]

\[
= (\lambda_g + \lambda_l) \int_E ((U \circ f) - E_\mu[U \circ f]) \mu(ds)
\]

\[
= \lambda_g \int_E ((U \circ f) - E_\mu[U \circ f]) \mu(ds) + \lambda_l \int_{E^c} (E_\mu[U \circ f] - (U \circ f)) \mu(ds)
\]

\[
= E_\mu[\eta^\lambda(U \circ f, \mu)]
\]

\(U\) is cardinally unique since it represents preferences over constant acts as a result of the Mixture Space Theorem. \(\mu\) is unique, and moreover every \(\mu_E\) is unique as well. This implies that \(\lambda = \gamma^E_\lambda - \gamma^E_\mu\) is unique as well from the definition of \(\gamma\)'s from (A.26) and (A.27).

\[\Box\]

**Proof of Theorem 4.**

It suffices to show that for any \(f \in \mathcal{F}^E\), if \(\lambda < 0\),

\[
\mu_E = \arg \min_{\nu \in \{\mu_E\} \in \mathcal{E}} \left( \int_S (U \circ f) \nu(ds) \right)
\]

whereas if \(\lambda > 0\),

\[
\mu_E = \arg \max_{\nu \in \{\mu_E\} \in \mathcal{E}} \left( \int_S (U \circ f) \nu(ds) \right)
\]
If this is true, then the result follows from Proposition 2.

Consider the case where $\lambda > 0$, let $f \in \mathcal{F}^E$. Suppose $v \in \mathbb{R}^{|S|}$ is the utility vector assigned to $f$, i.e. $U \circ f = v$. Since $\mu$ is the unique distribution that determines $e_f$ as shown in the proof of Theorem 3, then for all $s \in E$, $v_s \geq \int_S v \mu(ds)$, and for all $s \in E^c$, $v_s \leq \int_S v \mu(ds)$. From Theorem 3, for any $E \in \mathcal{E}$,

$$
\mu_E(s) = \begin{cases} 
\mu(s) (1 - \lambda \mu(E^c)) & \text{if } s \in E \\
\mu(s) (1 + \lambda \mu(E)) & \text{if } s \in E^c
\end{cases}
$$

(A.5)

Want to show that $\int_S v \mu_E(ds) \leq \int_S v \mu_F(ds)$, for any $F \in \mathcal{E}, F \neq E$. This is equivalent to showing that for all $F$, $\int_S v (\mu_E - \mu_F)(ds) \leq 0$. For any $E$, $F$, $\mu_E(s) - \mu_F(s)$ depends on whether $s \in E \cap F$, $E^c \cap F$, $E \cap F^c$, or $E^c \cap F^c$. To calculate $\mu_E - \mu_F$, use the following table

<table>
<thead>
<tr>
<th>$s$</th>
<th>$\mu_E(s)$</th>
<th>$\mu_F(s)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E \cap F^c$</td>
<td>$(1 - \lambda \mu(E^c))$</td>
<td>$(1 - \lambda \mu(F^c))$</td>
</tr>
<tr>
<td>$E \cap F$</td>
<td>$(1 - \lambda \mu(E^c))$</td>
<td>$(1 + \lambda \mu(F))$</td>
</tr>
<tr>
<td>$E^c \cap F$</td>
<td>$(1 + \lambda \mu(E))$</td>
<td>$(1 - \lambda \mu(F^c))$</td>
</tr>
<tr>
<td>$E^c \cap F^c$</td>
<td>$(1 + \lambda \mu(E))$</td>
<td>$(1 + \lambda \mu(F))$</td>
</tr>
</tbody>
</table>

Hence, $\mu_E - \mu_F$ can be calculated by simplifying the expressions above.

<table>
<thead>
<tr>
<th>$s$</th>
<th>$\frac{\mu_E(s)}{\mu(s)} - \frac{\mu_F(s)}{\mu(s)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E \cap F$</td>
<td>$\lambda(\mu(E) - \mu(F))$</td>
</tr>
<tr>
<td>$E \cap F^c$</td>
<td>$\lambda(\mu(E) - \mu(F) - 1)$</td>
</tr>
<tr>
<td>$E^c \cap F$</td>
<td>$\lambda(\mu(E) - \mu(F) + 1)$</td>
</tr>
<tr>
<td>$E^c \cap F^c$</td>
<td>$\lambda(\mu(E) - \mu(F))$</td>
</tr>
</tbody>
</table>

So for $v = U \circ f$, if $\lambda > 0$ (assume $\lambda = 1$ for notational convenience).

$$
\int_S v (\mu_E - \mu_F)(ds) = \sum_{E \cap F} v_i \mu(t) (\mu(E) - \mu(F)) + \sum_{E \cap F^c} v_i \mu(t) (\mu(E) - \mu(F)) + \sum_{E^c \cap F^c} v_i \mu(t) (\mu(E) - \mu(F) - 1) + \sum_{E^c \cap F} v_i \mu(t) (\mu(E) - \mu(F) + 1)
$$

$$
= \sum_{E \cap F} v_i \mu(t) (\mu(E) - \mu(F)) - \sum_{E \cap F^c} v_i \mu(t) + \sum_{E^c \cap F} v_i \mu(t)
$$

$$
\leq \sum_{E \cap F} v_i \mu(t) (\mu(E) - \mu(F)) - \sum_{E \cap F^c} \left( \sum_{E \cap F} v_i \mu(t) \right) \mu(t) + \sum_{E^c \cap F} \left( \sum_{E \cap F} v_i \mu(t) \right) \mu(F) + \sum_{E^c \cap F} v_i \mu(t) (\mu(E^c \cap F^c)) + \sum_{E^c \cap F} v_i \mu(t) (\mu(E^c \cap F)) = 0
$$

This holds because $\mu(E) - \mu(F) = \mu(E \cap F^c) - \mu(E^c \cap F)$. For the case where $\lambda < 0$ the inequality are just reversed, hence $\mu_E$ is the maximizer rather than the minimizer. It follows that for $f \in \mathcal{F}^E$, for any $\nu \in \text{conv} \{ \mu_E \}_{E \in \mathcal{E}}$, $\int_S (U \circ f) \nu(ds) \geq (\leq) \int_S (U \circ f) \mu_E(ds)$
for $\lambda > 0 (\lambda < 0)$. Then representation from Proposition 2, implies that for $\lambda > 0$,

$$f \succeq g \iff \min_{\nu \in C} \int_S (U \circ f) \nu(ds) \geq \min_{\nu \in C} \int_S (U \circ g) \nu(ds) \tag{A.6}$$

and for $\lambda < 0$,

$$f \succeq g \iff \max_{\nu \in C} \int_S (U \circ f) \nu(ds) \geq \max_{\nu \in C} \int_S (U \circ g) \nu(ds) \tag{A.7}$$

Where $C = \text{conv} \{(\mu_E)_{E \in \mathcal{E}}\}$, and $\mu_E$ is defined as in (A.5). Proving the result.

\[ \square \]

**Proof of Proposition 5.**

Show only the first set of equivalences. The proof for the second set is exactly the same, where the inequalities are switched. Note that from the proof of Theorem 3, equation (A.4) implies that the functional $V(f)$ from the AGL representation can be rewritten as

$$V(f) = \int_S (U \circ f) \mu(ds) + \frac{\lambda}{2} \int_S |(U \circ f) - (U \circ e_f)| \mu(ds) \tag{A.8}$$

(i) $\Rightarrow$ (iii). Let $\prec$ be gain-biased, then for any $f \in \mathcal{F}$, $c_f \succeq e_f$. From the AGL representation and equation (A.8) for all $f \in \mathcal{F}$,

$$\int_S (U \circ f) \mu(ds) + \frac{\lambda}{2} \int_S |(U \circ f) - (U \circ e_f)| \mu(ds) \geq U(e_f) = \int_S (U \circ f) \mu(ds)$$

Since $U(e_f) = \int_S (U \circ f) \mu(ds)$, this implies $\frac{\lambda}{2} \int_S |U \circ f - U \circ e_f| \mu(ds) \geq 0$, which holds only if $\lambda \geq 0$.

(iii) $\Rightarrow$ (ii). Let $\succeq$ be uncertainty seeking. From the equivalent representation in Theorem 4, $\succeq$ also admits a Maxmax representation. Therefore from $\succeq$ is uncertainty seeking, since it is a necessary condition for this representation.

(ii) $\Rightarrow$ (i). Let $\preceq$ be uncertainty seeking. Then for all $f$,

$$e_f = \frac{1}{2} f + \frac{1}{2} \bar{f} \preceq f \sim c_f$$

since $f \sim \bar{f}$ by definition. Hence $e_f \preceq c_f$, which implies that $\preceq$ is gain-biased.

\[ \square \]

**Proof of Theorem 6.**
Given that $V(f)$ from the AGL representation can be written as

$$V(f) = \int_S (U \circ f) \mu(ds) + \frac{\lambda}{2} \int_S |(U \circ f) - (U \circ e_f)| \mu(ds)$$

For any $s \in S$, $f \lor \bar{f}(s) \geq e_f$, and $f \land \bar{f}(s) \leq e_f$, then $U(f \lor \bar{f}(s)) = U(e_f) + d_s$, and $U(f \land \bar{f}(s)) = U(e_f) - d_s$, since $e_f \geq 0$, where $d_s = |U(f(s)) - U(e_f)| = |U(\bar{f}(s) - U(e_f))|$. Use the notation that superscripts denote the DM, e.g., $e_i^f$ is the hedge of $f$ for DM $i$. Moreover since $U_1$ and $U_2$ are cardinally equivalent, can assume $U_1 = U_2 = U$.

(i) $\Leftrightarrow$ (ii). The equivalence between gain bias and loss bias is straightforward. It follows from definition and the fact that $U((f \lor \bar{f})(s)) = U(e_f) + d_s$, and $U((f \land \bar{f})(s)) = U(e_f) - d_s$. Hence just need to prove the equivalence between (i) and (iii).

(i) $\Rightarrow$ (iii). Let $\succeq_1$ be more gain-biased than $\succeq_2$. Consider any $f, g \in \mathcal{F}$ such that $e_i^f = e_i^g$.

Suppose in addition that $e_i^f \succeq_i e_i^{2g\lor g}$ for $i = 1, 2$. By the AGL representation result, $\mu_i$ determines $U(e_i^f)$ for all $f \in \mathcal{F}$. Since $U_1 = U_2 = U$, then $e_i^f \succeq_i e_i^{2g\lor g}$ for $i = 1, 2$ implies

$$\int_S (U \circ e_i^f) \mu_1(ds) \geq \int_S (U \circ e_i^{2g\lor g}) \mu_2(ds)$$

From the previous observation about $d_s$, for all $s \in S$, $U(f \lor \bar{f}) = U(e_f) + |U(f(s)) - U(e_f)|$, hence

$$\int_S ((U \circ e_i^f) + |U(f(s)) - U(e_f)|) \mu_1(ds) \geq \int_S ((U \circ e_i^g) + |U(f(s)) - U(e_f)|) \mu_2(ds)$$

$$\Rightarrow U(e_i^f) + \int_S |U(f(s)) - U(e_i^f)| \mu_1(ds) \geq U(e_i^g) + \int_S |U(f(s)) - U(e_i^g)| \mu_2(ds)$$

$$\Rightarrow \int_S |U(f(s)) - U(e_i^f)| \mu_1(ds) \geq \int_S |U(f(s)) - U(e_i^g)| \mu_2(ds)$$

Now suppose that $g \succeq_2 c$ for some $c \in \mathcal{F}_c$, implies that $f \succeq_1 c$. Using equation (A.8)

$$V(g) = \int_S (U \circ g) \mu_2(ds) + \frac{\lambda}{2} \int_S |U(g(s)) - U(e_i^g)| \mu_2(ds) \geq U(c)$$

implies

$$V(f) = \int_S (U \circ g) \mu_1(ds) + \frac{\lambda}{2} \int_S |U(g(s)) - U(e_i^g)| \mu_1(ds) \geq U(c)$$

This implies that $V(f) \geq V(g)$ for all such $f$ and $g$, where $e_i^f \succeq_i e_i^{2g\lor g}$.

---

33 By monotonicity and $U_1, U_2$ being cardinally equivalent we can simplify $e_i^f \sim_i e_i^{2g\lor g}$ for $i = 1, 2$. 

---

46
Since $e_f^1 = e_g^2$, for any $f, g \in \mathcal{F}$,

$$V(g) = \int_S (U \circ g)\mu_2(ds) + \frac{\lambda_2}{2} \int_S |U(g(s)) - U(e_g^2)|\mu_2(ds)$$

$$= U(e_g^2) + \frac{\lambda_2}{2} \int_S |U(g(s)) - U(e_g^2)|\mu_2(ds)$$

$$= U(e_f^1) + \frac{\lambda_2}{2} \int_S |U(g(s)) - U(e_g^2)|\mu_2(ds)$$

$$\leq U(e_f^1) + \frac{\lambda_2}{2} \int_S |U(f(s)) - U(e_f^1)|\mu_1(ds)$$

$$\leq U(e_f^1) + \frac{\lambda_1}{2} \int_S |U(f(s)) - U(e_f^1)|\mu_1(ds) = V(f)$$

This holds even in the case of $e_{f \vee f} = e_{g \vee g}$, which given the previous result implies that

$$\int_S |U(g(s)) - U(e_g^2)|\mu_2(ds) = \int_S |U(f(s)) - U(e_f^1)|\mu_1(ds).$$

Hence $\lambda_1 \geq \lambda_2$.

(iii) $\Rightarrow$ (i). Let $\lambda_1 \geq \lambda_2$. Let $f, g \in \mathcal{F}$ be such that $e_f^1 = e_g^2$, and $e_{f \vee f} \succsim e_{g \vee g}$.

Suppose for some $c \in \mathcal{F}_c$, $g \succsim c$. Then by the AGL representation and the equivalent representation of (A.8),

$$V(g) = \int_S (U \circ g)\mu_2(ds) + \frac{\lambda_2}{2} \int_S |U(g(s)) - U(e_g^2)|\mu_2(ds)$$

$$= U(e_g^2) + \frac{\lambda_2}{2} \int_S |U(g(s)) - U(e_g^2)|\mu_2(ds) \geq V(c)$$

Since $e_f^1 = e_g^2$, then $U(e_g^2) = U(e_f^1)$, and $e_{f \vee f} \succsim e_{g \vee g}$,

$$\int_S |U(g(s)) - U(e_g^2)|\mu_2(ds) \leq \int_S |U(f(s)) - U(e_f^1)|\mu_1(ds)$$

Then

$$V(g) = U(e_g^2) + \frac{\lambda_2}{2} \int_S |U(g(s)) - U(e_g^2)|\mu_2(ds)$$

$$\leq U(e_f^1) + \frac{\lambda_2}{2} \int_S |U(f(s)) - U(e_f^1)|\mu_1(ds)$$

$$\leq U(e_f^1) + \frac{\lambda_1}{2} \int_S |U(f(s)) - U(e_f^1)|\mu_1(ds) = V(f)$$

Therefore $V(f) \geq U(c)$, showing that $f \succsim_1 c$. 

\qed
A.2 Proofs of Lemmas and Propositions

Proof of Proposition 9.

Consider any $E, E' \in \mathcal{E}$, and single alignment acts $f \in \mathcal{F}^E$, $g \in \mathcal{F}^{E'}$. Let $F$ be a non-overlapping event for both $E$ and $E'$. Consider two distinct $h, h' \in \mathcal{F}$, such that $h(s) \sim \tilde{h}(s)$ and $h'(s) \sim \tilde{h}'(s)$ for all $s \notin F$, which means that their alignment is neutral in $F^c$. Since the range of $U$ is $(-\infty, \infty)$ by unboundedness, without loss of generality let $U(h'_s) = U(h_s) = 0$ for every $s \notin F$. Moreover suppose that for every $s \in F$, $h(s) \succ \tilde{h}(s)$ or $h'(s) \succ \tilde{h}'(s)$, that means that on every state in $F$, $h$ and $h'$ are either strictly considered positive or negative. In addition, let $h \sim h'$. Note that the alignment condition of $s \in F$, and strict monotonicity imply that $U(h(s)) \neq 0$ and $U(h'(s)) \neq 0$ for any $s \in F$, further assume that for all $s \in F$, $h(s) \succ h'(s)$ or $h'(s) \succ h(s)$ (hence on $F$ the acts are always different in terms of preferences).\(^{34}\)

NOEL-mixture consistency guarantees that for $f, g, h$ exists some $\alpha_{fgh}$ such that for any $\alpha \in (\alpha_{fgh}, 1)$, $f \succsim \alpha f + (1 - \alpha) h$ if and only if $g \succsim \alpha g + (1 - \alpha) h$. Then for any $\alpha \in (\min \{\alpha_{fgh}, \alpha_{fgh'}\}, 1)$

$$\alpha f + (1 - \alpha) h \succsim \alpha f + (1 - \alpha) h' \iff \alpha g + (1 - \alpha) h \succsim \alpha g + (1 - \alpha) h'$$ \hspace{1cm} (A.9)

Moreover, since $f$ and $g$ are single-aligned continuity of preferences imply that alignment does not change for small perturbations around $f$ and $g$. Hence, the for $\alpha$ close to one, $\alpha f + (1 - \alpha) h \in \mathcal{F}^E$, and $\alpha g + (1 - \alpha) h \in \mathcal{F}^{E'}$. From the representation of 2, (A.9) implies

$$\int_S U \circ (\alpha f + (1 - \alpha) h) \mu_E(ds) = \int_S U \circ (\alpha f + (1 - \alpha) h') \mu_E(ds) \text{ if and only if}$$

$$\int_S U \circ (\alpha f + (1 - \alpha) h) \mu_{E'}(ds) = \int_S U \circ (\alpha g + (1 - \alpha) h') \mu_{E'}(ds)$$

By linearity of $U$ (from 2) this is equivalent to

$$\alpha \int_S (U \circ f) \mu_E(ds) + (1 - \alpha) \int_S (U \circ h) \mu_E(ds) = \alpha \int_S (U \circ f) \mu_{E'}(ds) + (1 - \alpha) \int_S (U \circ h') \mu_{E'}(ds)$$ \hspace{1cm} (A.10)

$$\iff \alpha \int_S (U \circ g) \mu_{E'}(ds) + (1 - \alpha) \int_S (U \circ h) \mu_{E}(ds) = \alpha \int_S (U \circ g) \mu_{E'}(ds) + (1 - \alpha) \int_S (U \circ h') \mu_{E'}(ds)$$

Since $U(h_s) = U(h'_s) = 0$ for all $s \notin F$, (A.10) reduces to

$$\sum_F (U \circ h - U \circ h') \mu_E(s) = 0 \iff \sum_F (U \circ h - U \circ h') \mu_{E'}(s) = 0$$ \hspace{1cm} (A.11)

\(^{34}\text{By Strong Monotonicity and Continuity there is always possible to find such acts } h \text{ and } h'.\)
Since $S$ is finite, we can consider $(U \circ h)|F$, the restriction of $U \circ h$ to states in $F$, as a vector in $\mathbb{R}^{|F|}$.\footnote{More precisely to the $|F|$-dimensional vector subspace of $\mathbb{R}^{|S|}$ where any coordinate in $S \setminus F$ is 0.} Normalize $\mu_E$ and $\mu_{E'}$ conditional on $F$ to be a probability distribution over $F$, then equation (A.11) becomes

$$\sum_{F'}(U \circ h - U \circ h') \frac{\mu_E(ds)}{\mu_E(F)} = 0 \iff \sum_{F'}(U \circ h - U \circ h') \frac{\mu_{E'}(ds)}{\mu_{E'}(F)} = 0$$

Since all states are non-null, $\mu_E(\cdot|F)$ and $\mu_{E'}(\cdot|F)$ are strictly positive $|F|$-dimensional vectors. These vectors are normal to $(U \circ h - U \circ h') \in \mathbb{R}^{|F|}$, which consists of non-zero elements by the assumption that $h$ and $h'$ are different for all $s \in F$, $(U \circ h - U \circ h')_s \neq 0$ for any $s \in F$. Therefore $\mu_E(\cdot|F)$ and $\mu_{E'}(\cdot|F)$ are colinear as vectors in $\mathbb{R}^{|F|}$, and moreover both are probability distributions which means $\sum_F \mu_E(s|F) = \sum_F \mu_{E'}(s|F) = 1$. Then for all $s \in F$,

$$\frac{\mu_E(s)}{\mu_E(F)} = \frac{\mu_{E'}(s)}{\mu_{E'}(F)} \quad \text{or equivalently} \quad \mu_E(s|F) = \mu_{E'}(s|F) \quad (A.12)$$

For any $E, E' \in \mathcal{E}$, and any $F$ which is non-overlapping with $E$ and $E'$, the conditional distribution $\mu_E(\cdot|F)$ and $\mu_{E'}(\cdot|F)$ which are generated by the set of priors $\{\mu_E\}_{E \in \mathcal{E}}$ from Theorem 1. It follows from (A.12) that for any $E, E'$ where $s, s' \in E \cap E'$,

$$\frac{\mu_E(s)}{\mu_E(s')} = \frac{\mu_{E'}(s)}{\mu_{E'}(s')}$$

It remains to show that if (A.12) holds there exists a unique distribution $\mu$ that generates the conditional distributions, i.e. such that for all $E \in \mathcal{E}$ and non-overlapping $F$,

$$\frac{\mu_E(s)}{\mu_E(F)} = \frac{\mu(s)}{\mu(F)} \quad (A.13)$$

Consider the events $E_i = \{s_i, s_{i+1}\}$ for $i = 1, \ldots, n - 1$.\footnote{This is a set that “spans” $S$, and it is minimal in the sense that there are no overlapping events to $E_i$ which are not singleton states.} If such a $\mu$ exists it follows from (A.13) that for $E_i$,

$$\frac{\mu(s_i)}{\mu(s_{i+1})} = \frac{\mu_{E_i}(s_i)}{\mu_{E_i}(s_{i+1})} \quad (A.14)$$

If such a $\mu$ exists and it is unique for the set $\{E_i\}_{i=1,\ldots,n-1}$, then by (A.2), it will hold for any $E \in \mathcal{E}$. This is because the set $\{E_i\}_{i=1,\ldots,n}$ can be used to find the conditional distributions for any other $E \in \mathcal{E}$ using the results from the first par of the proof. To see this suppose that there is a unique $\mu$ such that for all $\{E_i\}_{i=1,\ldots,n-1}$, (A.13) holds. Now consider some $E \notin \{E_i\}_{i=1,\ldots,n-1}$. $E \subset \bigcup_{i=m}^M E_i := E^M_m$ for some $m < M$, where $m \geq 1$.
and $M \leq n$. To see this note that given a numbering of states, there is a maximal $s_M$, and a minimal $s_m$ states in $E$, hence $E \subseteq \{s_m, s_{m+1}, \ldots, s_{M-1}, s_M\} = E^M$. Then for any $s_t, s_{t+k} \in E$, $t < k$ and $t, k \in \{m, \ldots, M\}$, \[ \frac{\mu_{E^M(m)}(s_t)}{\mu_{E^M(m)}(s_{t+k})} = \frac{\mu_{E^M(m)}(s_{t+1})}{\mu_{E^M(m)}(s_{t+k+1})} \frac{\mu_{E^M(m)}(s_{t+2})}{\mu_{E^M(m)}(s_{t+k+2})} \ldots \frac{\mu_{E^M(m)}(s_{t+k-1})}{\mu_{E^M(m)}(s_{t+k})} \]

Hence it suffices to show that for the set $\{E_i\}_{i=1,\ldots,n}$, a unique distribution exists such that (A.13) holds. Note that (A.14) implies that for any $i = 1, 2, \ldots, n-1$,

\[ \mu(s_i) = \left( \frac{\mu_{E_i}(s_i)}{\mu_{E_{i+1}}(s_{i+1})} \right) \mu(s_{i+1}) \]

These $n-1$ equations, along with the necessary condition to be a probability distribution:

\[ \sum_{i=1}^{n} \mu(s_1) = 1, \]

gives $n$ equations and $n$ unknowns (the $\mu(s_i)$'s), which can be written in the following form:

\[ \begin{bmatrix} 1 - \frac{\mu_{E_1}(s_1)}{\mu_{E_1}(s_2)} & 0 & 0 & \ldots & \ldots & 0 \\ 0 & 1 - \frac{\mu_{E_2}(s_2)}{\mu_{E_2}(s_3)} & 0 & \ldots & \ldots & 0 \\ 0 & 0 & 1 - \frac{\mu_{E_3}(s_3)}{\mu_{E_3}(s_4)} & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & \ldots & 1 - \frac{\mu_{E_{n-1}}(s_{n-1})}{\mu_{E_{n-1}}(s_n)} & 0 \\ 1 & 1 & \ldots & \ldots & \ldots & 1 \end{bmatrix} \begin{bmatrix} \mu \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \] (A.15)

Equation (A.15) has a unique solution if and only if the matrix $A_n$ is invertible. This is proved by induction on $|S|$. Let $|S| = 3$, then need to show that $\det(A_3) \neq 0$.

Let $a_{ij} = \frac{\mu_{E_i}(s_i)}{\mu_{E_j}(s_j)} > 0$. $a_{ij} \in (0, \infty)$ because every state is non-null so $\mu_{E_i}(s_j) > 0$ for all $i, j$.

\[ |S| = 2 \text{ it is vacuously true, since } E_1 = S \text{ by definition of } E_i, \text{ hence } E_1 \notin \mathcal{E}. \]
Prove the stronger condition that \( \det(A_k) > 0 \) instead.

\[
\det \begin{pmatrix}
1 & -a_{12} & 0 \\
0 & 1 & -a_{23} \\
1 & 1 & 1
\end{pmatrix} = 1 (1 + a_{23}) - (-a_{12})(a_{23}) = 1 + a_{23}(1 + a_{12}) > 0
\]

Suppose now that \( \det(A_k) > 0 \) for all \( k < m \). Then

\[
A_m = \begin{bmatrix}
1 & -a_{12} & 0 & \cdots & 0 \\
0 & 1 & -a_{34} & \cdots & 0 \\
0 & \ddots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & 1 & -a_{n(n-1)} \\
0 & \cdots & \cdots & 1 & 1
\end{bmatrix}
\]

Hence \( \det(A_k) = \det(A_{m-1}) + a_{12} \det(B_k) \) where \( B_k \) is defined as

\[
B_k = \begin{bmatrix}
0 & -a_{23} & 0 & \cdots & 0 \\
0 & 1 & -a_{34} & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & 0 & -a_{n(n-1)} \\
1 & 1 & \cdots & \cdots & 1
\end{bmatrix}
\]

Therefore \( \det(A_k) = \det(A_{m-1}) + a_{12}a_{23} \det(A_{m-3}) > 0 \), from the induction hypothesis that for all \( k < m \), \( \det(A_k) > 0 \). Hence the system from equation (A.15) has a unique solution, \( \mu \). From the previous result, for any \( E \in \mathcal{E} \) such that \( |E| > 2 \), \( \mu_E \) is also generated by \( \mu \). Hence there exists a unique \( \mu : 2^S \to [0, 1] \) such that every conditional distribution of \( \mu \) (conditional on event \( F \)), is the same as the conditional distribution of \( \mu_E \) provided that \( F \) and \( E \) are non-overlapping.

\[ \square \]

**Proof of Lemma 10.**

Suppose there exists \( s \in S \) such that \( \mu_E(s) = \mu_{E'}(s) \) for some \( E \neq E' \). Then by Proposition 9, there are two cases:

(i.) \( s \in E \cap E' \) (or \( s \in E^c \cap E'^c \)).

(ii.) \( s \in E \cap (E')^c \) (or \( s \in E' \cap E^c \)).

For case (i) by Proposition 9, for every \( s' \in E \cup E' \), \( \mu_E(s') = \mu_{E'}(s') \) since \( \mu_E(s) = \gamma_E^+ \mu(s) \) and \( \mu_E(s) = \gamma_{E'}^+ \mu(s) \). In addition, for any \( t \in E \cap (E')^c \), \( \mu_E(t) = \mu_{E'}(t) \) by the same argument, which implies that for all \( s \in (E')^c \), \( \mu_E(t) = \gamma_{E'}^+ \mu(t) \), which can only be true if \( \gamma_E^+ = \gamma_{E'}^+ = 1 \). Hence \( \mu_E = \mu_{E'} = \mu \). For the second case the argument is symmetric.
(replacing $\gamma^+_{E'}$ for $\gamma^-_{E'}$).

\[\square\]

**Proof of Proposition 11.**

Consider some single alignment $f \in \mathcal{F}^E$, and $g \in \mathcal{F}^F$ such that $f \sim g$, where $\bar{f} \in \mathcal{F}^{E^c}$ and $\bar{g} \in \mathcal{F}^{F^c}$ are the respective balancing acts. Given $s \in S$, consider some $h \in \mathcal{F}$ such that $U(h_t) = 0$ for all $t \neq s$ and $U(h_s) > 0$.

Suppose $\alpha f + (1 - \alpha) h \succ \alpha g + (1 - \alpha) h$, then by BA-antisymmetry $\alpha \bar{f} + (1 - \alpha) h \prec \alpha \bar{g} + (1 - \alpha) h$. From the definition of alignment and continuity of $\succ$, for $\alpha$ close to 1, then $\alpha f + (1 - \alpha) h \in \mathcal{F}^E$ and $\alpha g + (1 - \alpha) h \in \mathcal{F}^F$; likewise $\alpha \bar{f} + (1 - \alpha) h \in \mathcal{F}^{E^c}$ and $\alpha \bar{g} + (1 - \alpha) h \in \mathcal{F}^{F^c}$. Therefore by the representation result from Proposition 2, from BA-antisymmetry $\mu_E(s) > \mu_F(s)$ if and only if $\mu_{E^c}(s) > \mu_{F^c}(s)$.

Suppose that $\mu_E + \mu_{E^c} \neq \mu_F + \mu_{F^c}$. Then for some $s \in S$, $\mu_E(s) + \mu_{E^c}(s) > \mu_F(s) + \mu_{F^c}(s)$, this in turn implies that there exists some $s' \in S$ where $\mu_E(s') + \mu_{E^c}(s') < \mu_F(s') + \mu_{F^c}(s')$ because all $\mu_E, \mu_{E^c}, \mu_F, \mu_{F^c}$ are probability distributions so $\sum_{t \in S} (\mu_E(t) + \mu_{E^c}(t)) = 2$. Then

$$\mu_E(s) - \mu_F(s) > \mu_{E^c}(s) - \mu_{E^c}(s) \quad \text{and} \quad \mu_E(s') - \mu_F(s') < \mu_{E^c}(s') - \mu_{E^c}(s')$$ (A.16)

It must be the case that the following two conditions hold:

$$\mu_E(s) - \mu_F(s) = \theta (\mu_{E^c}(s) - \mu_{E^c}(s)) \quad \text{for some } \theta < 1$$
$$\mu_E(s') - \mu_F(s') = \theta' (\mu_{E^c}(s') - \mu_{E^c}(s')) \quad \text{for some } \theta' < 1$$ (A.17)

For any $v \in \mathbb{R}^{|S|}$ such that $v_t = 0$ for all $t \neq s, s'$, and $v_s, v_{s'} \neq 0$, there exists some $h \in \mathcal{F}$ associated with $v$, i.e. $v$ is the utility vector associated with $h$. For this $h_v \in \mathcal{F}$, $U(h_t) = 0$ for all $t \neq s, s'$ and $U(h_s) = v_s \neq 0$ and $U(h_{s'}) = v_{s'} \neq 0$, this is guaranteed by strong monotonicity and continuity of $\succ$. By the same argument previously mentioned, for single alignment $f \in \mathcal{F}^E$ and $g \in \mathcal{F}^F$ where $f \sim g$, for $\alpha$ close to 1, then $\alpha f + (1 - \alpha) h \in \mathcal{F}^E$ and $\alpha g + (1 - \alpha) h \in \mathcal{F}^F$; likewise $\alpha \bar{f} + (1 - \alpha) h \in \mathcal{F}^{E^c}$ and $\alpha \bar{g} + (1 - \alpha) h \in \mathcal{F}^{F^c}$. Hence for $h_v$, where $U \circ h_v = v$, representation from proportion 2 and BA-antisymmetry, which requires $\alpha f + (1 - \alpha) h_v \succ \alpha g + (1 - \alpha) h_v$ if and only if $\alpha \bar{f} + (1 - \alpha) h_v \prec \alpha \bar{g} + (1 - \alpha) h_v$. This reduces to

$$\mu_E(s) v_s + \mu_{E^c}(s) v_{s'} > \mu_F(s) v_s + \mu_{F^c}(s') v_{s'}$$
$$\iff \mu_{E^c}(s) v_{s'} + \mu_{E^c}(s') v_s > \mu_{F^c}(s) v_{s'} + \mu_{F^c}(s') v_s$$

In other words, there is no $v_s, v_{s'} \in \mathbb{R}$ that solve the system

$$(\mu_E(s) - \mu_F(s)) v_s + (\mu_{E^c}(s') - \mu_{F^c}(s')) v_{s'} > 0$$
$$(\mu_{E^c}(s) - \mu_{F^c}(s)) v_s + (\mu_{E^c}(s') - \mu_{F^c}(s')) v_{s'} > 0$$

52
Using the observation (A.17), this is equivalent to
\[
\begin{align*}
\theta (\mu_{F^c}(s) - \mu_{E^c}(s)) v_s + \theta' (\mu_{F^c}(s') - \mu_{E^c}(s')) v_{s'} > 0 \\
- (\mu_{F^c}(s) - \mu_{E^c}(s)) v_s + - (\mu_{F^c}(s') - \mu_{E^c}(s')) v_{s'} > 0
\end{align*}
\]
\[
\begin{bmatrix}
\theta (\mu_{F^c}(s) - \mu_{E^c}(s)) & \theta' (\mu_{F^c}(s') - \mu_{E^c}(s')) \\
- (\mu_{F^c}(s) - \mu_{E^c}(s)) & - (\mu_{F^c}(s') - \mu_{E^c}(s'))
\end{bmatrix}
\begin{bmatrix}
v_s \\
v_{s'}
\end{bmatrix}
> \begin{bmatrix} 0 \\ 0 \end{bmatrix}
\tag{A.18}
\]

Since there is no solution to (A.18), there exists some \( p > 0 \) (Stiemke’s Alternative [Stiemke, 1915]) such that
\[
\begin{bmatrix}
p_1 \\ p_2
\end{bmatrix}
\begin{bmatrix}
\theta (\mu_{F^c}(s) - \mu_{E^c}(s)) & \theta' (\mu_{F^c}(s') - \mu_{E^c}(s')) \\
- (\mu_{F^c}(s) - \mu_{E^c}(s)) & - (\mu_{F^c}(s') - \mu_{E^c}(s'))
\end{bmatrix}
\begin{bmatrix}
v_s \\
v_{s'}
\end{bmatrix}
= \begin{bmatrix} 0 \\ 0 \end{bmatrix}
\]

Which implies that \((\mu_{F^c}(s) - \mu_{E^c}(s)) (p_1 \theta - p_2) = 0\) and \((\mu_{F^c}(s') - \mu_{E^c}(s')) (p_1 \theta' - p_2) = 0\), where \( p > 0 \). Since \((\mu_{F^c}(s) - \mu_{E^c}(s)) \neq 0\) and \((\mu_{F^c}(s') - \mu_{E^c}(s')) \neq 0\), it must be the case that \((p_1 \theta' - p_2) = (p_1 \theta - p_2) = 0\), which never holds when at least \( p_1 \) or \( p_2 \) are non-zero, and \( \theta > 1 > \theta' \). A contradiction. Therefore for any \( E, F \in \mathcal{E} \),
\[
\mu_E + \mu_{E^c} = \mu_F + \mu_{F^c}
\]

\[\square\]

**Proof of Proposition 13.**

Recall \( e_f \) is defined as the constant such that for a balanced pair, \((f, \bar{f}), \frac{1}{2}f(s) + \frac{1}{2}\bar{f}(s) = e_f \) for all \( s \in S \). From lemma 12, the representation from Proposition 2, and linearity of \( U \),
\[
\int_S (U \circ f) \mu(ds) = \frac{1}{2} \int_S (U \circ f) \mu_E(ds) + \frac{1}{2} \int_S (U \circ f) \mu_{E^c}(ds)
\]
\[
= \frac{1}{2} \int_S (U \circ \bar{f}) \mu_{E^c}(ds) + \frac{1}{2} \int_S (U \circ f) \mu_{E^c}(ds)
\]
\[
= \int_S \left( \frac{1}{2} (U \circ f) + \frac{1}{2} (U \circ \bar{f}) \right) \mu_{E^c}(ds)
\]
\[
= \int_S (U \circ e_f) \mu_{E^c}(ds) = U \circ e_f
\]

The same holds for \( \int_S (U \circ \bar{f}) \mu(ds) \), since for all \( s, s' \in S, U(f(s)) + U(\bar{f}(s)) = U(f(s')) + U(\bar{f}(s')) = 2U(e_f) \), since \( U \) represents preferences over constant acts, and the definition of \( e_f \). Hence for all \( f, e_f \) is the constant that is equivalent to the expected utility of \( f \in \mathcal{F} \) with the prior \( \mu \).

\[\square\]
Proof of Proposition 14.

Prove by induction on $|E \cap F|$. Let $E \cap F = s$. Then $F \setminus s = E^c$ and $E \setminus s = F^c$.

Hence
\[
\mu_E - \mu_{E \setminus s} = \mu_E - \mu_{F^c} \quad \text{and} \\
\mu_F - \mu_{F \setminus s} = \mu_F - \mu_{E^c}
\]

Hence from Proposition 11, $\mu_E + \mu_{E^c} = \mu_F + \mu_{F^c}$ for all $E, f \in \mathcal{E}$ and the above observation imply $\mu_E - \mu_{E \setminus s} = \mu_F - \mu_{F \setminus s}$.

Suppose (A.1) holds for all $E, F$ with $|E \cap F| = m - 1$ for some $m \in \mathbb{N}$, with $m < n$. Let $|E \cap F| = m$, such that $E \cap F = \{T, s\}$, where $|T| = m - 1$, and $s \notin T$. Consider any $t \in T$, hence $|(E \setminus s) \cap (F \setminus s)| = |T| = m - 1$ then by the induction hypothesis for any $t \in T$
\[
\mu_{E \setminus s} - \mu_{E \setminus \{s, t\}} = \mu_{F \setminus s} - \mu_{F \setminus \{s, t\}}
\]
Likewise, following the same argument and by the induction hypothesis, numbering $T = \{t_1, ..., t_k\}$, we have that
\[
\mu_{E \setminus \{s, t_1, ..., t_i\}} - \mu_{E \setminus \{s, ..., t_{i+1}\}} = \mu_{F \setminus \{s, t_1, ..., t_i\}} - \mu_{F \setminus \{s, ..., t_{i+1}\}}
\]
And finally for $E \setminus \{T, s\}$ and $F \setminus \{T, s\}$, $E \setminus \{T, s\} = F^c$ and $F \setminus \{T, s\} = E^c$ Proposition 11 implies that
\[
\mu_E - \mu_{E \setminus \{T, s\}} = \mu_F - \mu_{F \setminus \{T, s\}}
\]
Hence, subtracting (A.20) for $i = 2$ from (A.19), yields
\[
\mu_{E \setminus s} - \mu_{E \setminus \{s, t_1, t_2\}} = \mu_{F \setminus s} - \mu_{F \setminus \{s, t_1, t_2\}}
\]
Continuing this procedure for $i = 3, ..., k$, along with (A.21) yields that (A.1) must hold, hence proving the result.

\[\square\]

Proof of Proposition 16.

Consider 3 different cases, (i) $E = F^c$, (ii) $E \cap F \neq \emptyset$ and $E^c \cap F \neq \emptyset$, and $E \cap F^c \neq \emptyset$, and (iii) $F \subset E$. It suffices to consider these three conditions since $E \cap F = \emptyset$, and $E^c \cap F^c \neq \emptyset$, and lemma 12 will get the result for $E$ and $F$, from $E^c$ and $F^c$.

First note that the case where $F = E^c$ the result follows straightforwardly from Proposition 11. Now consider the cases where $E \cap F \neq \emptyset$ and $E^c \cap F \neq \emptyset$, and $E \cap F^c \neq \emptyset$. From Proposition 9, $E \in \mathcal{E}$, $\mu_E$ is obtained by distorting $\mu$ by the same amount on all the positive states ($\gamma^+_E$), and by the same amount ($\gamma^-_E$) on the negative states. Hence for
all \( t \in S \),

\[
\frac{\mu_E(t)}{\mu(t)} - \frac{\mu_{E \setminus s}(t)}{\mu(t)} = \frac{\mu_F(t)}{\mu(t)} - \frac{\mu_{F \setminus s}(t)}{\mu(t)} \tag{A.22}
\]

Since it is possible to divide by \( \mu(t) > 0 \), and the equality still holds. By the definition of \( \gamma_E^+ = \frac{\mu_E(s)}{\mu(s)} \) for \( s \in E \), and \( \gamma_E^- = \frac{\mu_E(s)}{\mu(s)} \) for \( s \in E^c \). Suppose \( s \in E \cap F \), then using (A.22) using the definition of states as positive or negative (when viewed from \( E \) or \( F \)),

- for \( s = E \cap F \) : \[
\gamma_E^+ - \gamma_{E \setminus s}^+ = \gamma_F^+ - \gamma_{F \setminus s}^+ \tag{A.23a}
\]
- for \( t \in E \cap F^c \) : \[
\gamma_E^+ - \gamma_{E \setminus s}^+ = \gamma_F^+ - \gamma_{F \setminus s}^+ \tag{A.23b}
\]
- for \( t' \in E^c \cap F \) : \[
\gamma_E^+ - \gamma_{E \setminus s}^+ = \gamma_F^+ - \gamma_{F \setminus s}^+ \tag{A.23c}
\]

(A.23c) \((A.23a)\) : \[
\gamma_E^+ - \gamma_{E \setminus s}^+ = \gamma_{F \setminus s}^+ - \gamma_{F \setminus s}^+ \tag{A.24a}
\]
(A.23b) \((A.23a)\) : \[
\gamma_{E \setminus s}^+ - \gamma_{F \setminus s}^+ = \gamma_F^+ - \gamma_{F \setminus s}^+ \tag{A.24b}
\]
(A.23b) \((A.23c)\) : \[
(\gamma_E^- - \gamma_{E}^-) - (\gamma_{E \setminus s}^- - \gamma_{E \setminus s}^-) = (\gamma_{F \setminus s}^- - \gamma_{F \setminus s}^-) - (\gamma_F^- - \gamma_{F}^-) \tag{A.24c}
\]

It is easy to see that substituting (A.24a) and (A.24b) into (A.24c) gets \( \gamma_E^- - \gamma_E^+ = \gamma_F^- - \gamma_F^+ \).

The other case, where \( F \subset E \), is simpler. There exist \( t \in E^c \), and \( s \in E \cap F \). Applying the same procedure as before yields

- \( s \in E \cap F \) : \[
\gamma_E^+ - \gamma_{E \setminus s}^+ = \gamma_F^+ - \gamma_{F \setminus s}^+ \tag{A.25a}
\]
- \( t \in E \cap F^c \) : \[
\gamma_E^+ - \gamma_{E \setminus s}^+ = \gamma_F^+ - \gamma_{F \setminus s}^+ \tag{A.25b}
\]

Subtracting (A.25b) from (A.25a), yields \( \gamma_E^- - \gamma_E^+ = \gamma_F^- - \gamma_F^+ \).\(^{38}\)

\[\square\]

**Proof of Proposition 17.**

By definition \( \mu_E(s) = \gamma_E^+ \mu(s) \) if \( s \in E \) and \( \mu_E(s) = \gamma_E^- \mu(s) \) if \( s \in E^c \). Since \( \gamma_E^+ \) is the same for all \( s \in E \), it follows that for any \( E' \subseteq E \), then \( \mu_E(E') = \gamma_E^+ \mu(E') \) as well. Then

---

\(^{38}\)Note that if \(|F| = 1\), \( \gamma_{\emptyset} \) would not be defined, but in that case if \(|S| \geq 3\), \(|F^c| = n - 1\) and the result can follow from reversing the roles of \( F \) and \( E \), with \( E^c \) and \( F^c \) and Proposition 11.
\[ \gamma^-_E - \gamma^+_E = \lambda \text{ implies that} \]

\[
\frac{\mu_E(E)}{\mu(E)} - \frac{\mu_E(E^c)}{\mu(E^c)} = \lambda \\
\Rightarrow \frac{\mu_E(E)}{\mu(E)} - \frac{1 - \mu_E(E)}{1 - \mu(E)} = \lambda \\
\Rightarrow \mu_E(E)(1 - \mu_E(E)) - (1 - \mu(E))\mu(E) = \lambda(1 - \mu(E))\mu(E)
\]

Solving for \( \mu_E(E) \) yields \( \mu_E(E) = \mu(E)(1 - \lambda(1 - \mu(E))) \), hence

\[ \gamma^+_E = \frac{\mu_E(E)}{\mu(E)} = (1 - \lambda(1 - \mu(E))) \quad (A.26) \]

Likewise we can solve for \( \gamma^-_E \) to get

\[ \gamma^-_E = \frac{\mu_E(E^c)}{\mu(E^c)} = (1 + \lambda \mu(E)) \quad (A.27) \]

The fact that \( \frac{\mu_E(E)}{\mu(E)} = \frac{\mu_E(s)}{\mu(s)} \) for all \( s \in E \), and \( \frac{\mu_E(E^c)}{\mu(E^c)} = \frac{\mu_E(s)}{\mu(s)} \) for all \( s \in E^c \) follows from Proposition 9. Therefore

\[
\begin{align*}
\mu_E(s) &= \mu(s)(1 - \lambda(1 - \mu(E))) & \text{for } s \in E \\
\mu_E(s) &= \mu(s)(1 + \lambda \mu(E)) & \text{for } s \in E^c
\end{align*}
\]

This proves the result.