

This document provides an overview of some of the major ideas we've covered in PHIL 443 so far. It is not meant to be totally comprehensive, but to complement your review of readings, lecture slides, and notes as you study for the midterm.

## CONTENTS

<b>1</b>	<b>Concepts to Know for the Midterm</b>	<b>1</b>
<b>2</b>	<b>The Standard Model of Rational Choice</b>	<b>2</b>
2.1	The St. Petersburg paradox and expected utility . . . . .	3
2.2	Misers . . . . .	4
2.3	The fair price of a wager and Ramsey's principle of exact credence . . . . .	4
2.4	Objections to the Belief/Desire Thesis . . . . .	4
<b>3</b>	<b>Choice Functions</b>	<b>5</b>
3.1	Principles $\alpha$ and $\beta$ . . . . .	5
3.2	The basic rationality requirements . . . . .	6
3.3	Knockout reasoning . . . . .	7
3.4	Violations of $[\alpha]$ and $[\beta]$ . . . . .	7
3.5	Money pump arguments . . . . .	8
<b>4</b>	<b>The Representation Theorem</b>	<b>8</b>
4.1	Axioms needed to prove the representation theorem . . . . .	8
4.2	Applying the axioms . . . . .	9
4.3	The Allais Paradox . . . . .	10
<b>5</b>	<b>Dutch Books</b>	<b>11</b>
5.1	Identifying and Constructing Dutch books . . . . .	11

## 1 CONCEPTS TO KNOW FOR THE MIDTERM

Here is a full list of concepts you need to know for the midterm. I go over almost all (but not all!!) of these concepts in this guide. Make sure you review all of them.

- Descriptive decision theory vs. rational choice theory (normative)
- Problem of the Points
- Belief/Desire model of practical reasons
- Credences (subjective probabilities)
- Laws of probability
- Conditional Probability
- Utility
- Equivalent utility scales (comparison with temperature)
- The concept of an expectation

- Expected utility
- Expected payoff (and how it differs from expected utility)
- The fair price of a wager
- St. Petersburg gamble
- Choice set
- Choice function
- Choiceworthiness
- Sen's Principles Alpha and Beta
- Choiceworthiness ranking (at-least-as-choiceworthy-as)
- Knockout reasoning
- The decoy effect
- The similarity effect
- Endowment effect
- Negativity bias
- Regret
- MiniMax regret
- Preference rankings
- The "Small Improvements" test
- Money-pump arguments
- The Axioms: Transitivity, Dominance, Sure-thing Principle, Stake Invariance
- Null event
- The Representation Theorem (state it)
- Ramsey's principle (qualitative and quantitative)
- Comparative probability (definition of  $.>.$ )
- Allais paradox
- The Dutch book theorem (be able to state it)

## 2 THE STANDARD MODEL OF RATIONAL CHOICE

The Standard Model is a *normative decision theory* (not a descriptive decision theory) on which agents act rationally when they maximize expected utility. The Standard Model has six tenets:

1. **THE BELIEF/DESIRE THESIS:** an agent's reasons for action are wholly a function about her beliefs about what is probable and desires for outcomes at the time of choice.
  - The fact that an agent desires a prospect, either as an end or as a means to an end, is always a *pro tanto* reason for her to act in ways that cause that prospect to be realized.
  - An agent desires a prospect *all things considered* when she desires it after considering all her competing desires and their intrinsically and instrumentally desirable features.
  - We say that an agent's beliefs and desires *rationalize* their choice of an act  $A$  by showing that, according to their beliefs,  $A$  is the best means they have at their disposal of satisfying their desires.
2. **GRADATION:** beliefs and desires are gradational.
  - We call degrees of belief *credences*.
  - We call degrees of desire *utilities*.
  - We usually describe graded attitudes in comparative terms.

3. REPRESENTATION: graded beliefs and desires are numerically measurable.

- Credences are measured by a credence function  $c$  such that, for any two events  $E$  and  $E^*$ ,  $E > .E^*$  iff  $c(E) > c(E^*)$ .<sup>1</sup>
- Utility is measured by a utility function  $u$  such that, for any two outcomes  $O$  and  $O^*$ ,  $O \succ O^*$  iff  $u(O) > u(O^*)$ .<sup>2</sup>

4. PROBABILISM: a rational agent's beliefs are governed by the laws of probability.

- Normality:  $0 = c(E \& \sim E) \leq c(E) \leq c(E \vee \sim E) = 1$  (i.e., an agent's credence in any event should fall between 0 (her credence in a contradiction) and 1 (her credence in a tautology))
- Additivity:  $c(E) + c(E^*) = c(E \vee E^*) + c(E \& E^*)$
- Bayesian conditioning:  $c(E | F) = c(E \& F) / c(F)$

5. EXPECTED UTILITIES AS ESTIMATES OF DESIRABILITY: the degree to which a rational agent finds a prospect desirable is her expectation of its desirability.

- We calculate the expected utility of an action by summing up the products of the utility of the outcome in each event and the agent's probability that the event occurs. E.g., suppose a miser is considering the following action  $A$ :

	0.2	0.4	0.1	0.3
E <sub>1</sub>		E <sub>2</sub>	E <sub>3</sub>	E <sub>4</sub>
A	\$2	\$3	\$2	\$6

$$EU(A) = (2)(0.2) + (3)(0.4) + (2)(0.1) + (6)(0.3) = 3.6$$

6. MAXIMIZATION: actions are choiceworthy iff they maximize expected utility.

## 2.1 THE ST. PETERSBURG PARADOX AND EXPECTED UTILITY

One problem for the view that an action is choiceworthy iff it maximizes *expected payoff* is the *St. Petersburg Paradox*.<sup>3</sup> Suppose I offer you a bet where I will toss a coin until it comes up heads. I offer you the following wager:

H	TH	TTH	TTTH	TTTTH...
\$0.02	\$0.04	\$0.08	\$0.16	\$0.32...

The expected payoff of this wager is infinite. Therefore, if the value of a bet is its expected payoff, you should be willing to pay any amount of money for this wager. But that's crazy! Your chance of winning less than \$3.00 is 0.996, and your chance of winning more than \$1000 is only 0.0000076.

Bernouilli's solution to the St. Petersburg paradox is to say that the bet has finite expected *utility*, despite its infinite expected *payoff*, because the value of a penny becomes vanishingly small as one accumulates more wealth. He proposed that one's utility of money =  $\log(\text{money})$ . This explains why you should be willing to pay very little money for this wager despite its infinite expected payoff.

<sup>1</sup>The symbol ".>." indicates comparative confidence. " $E > .E^*$ " means "an agent is more confident in  $E$  than in  $E^*$ ."

<sup>2</sup>The symbol  $\succ$  indicates comparative preference.  $O \succ O^*$  means "an agent prefers  $O$  to  $O^*$ ."

<sup>3</sup>For a thorough discussion of the St. Petersburg Paradox, see Lecture 4.

## 2.2 MISERS

A miser is a person who values only money and who values every additional dollar just as much as the last dollar. Most people are not misers—most people value things in addition to money and do not value money linearly—but we often talk about misers because they’re easier to deal with for the purposes of rational choice theory.

A miser’s utility function for money is linear:  $u(\$x) = x$ . Therefore, her expected utility of an bet is equal to the bet’s expected payoff. Non-misers do not have linear utility functions for money. To calculate the expected utility of a bet for a non-miser, you need convert money to utility in accordance with her utility function for money.

E.g., suppose Jim and Angela are offered the following bet on a horse race. They have the same credences in which horse will win. Jim is a miser. Angela is not a miser: her utility function for money is  $u(\$x) = x^{1/2}$ .

0.1 Stewball	0.1 Old Paint	0.6 Secretariat	0.2 Pegasus
\$25	\$64	\$49	\$100

Because Jim is a miser, his expected utility for the bet is equal to his expected payoff:  $(25)(0.1) + (64)(0.1) + (49)(0.6) + (100)(0.2) = 58.3$ . His fair price for the bet is \$58.30.

Here is Angela’s expected utility for the bet:  $(25^{1/2})(0.1) + (64^{1/2})(0.1) + (49^{1/2})(0.6) + (100^{1/2})(0.2) = (5)(0.1) + (8)(0.1) + (7)(0.6) + (10)(0.2) = 7.5$ . To find Angela’s fair price for the bet, we plug her expected utility back into her utility function for money:  $7.5^2 = \$56.25$ .

## 2.3 THE FAIR PRICE OF A WAGER AND RAMSEY’S PRINCIPLE OF EXACT CREDENCE

An agent’s fair price for a wager is the price at which she is indifferent between having the price and having the wager. For a miser, a fair price of a wager is her expected payoff.

Ramsey’s principle of exact credence provides us with a way of calculating an agent’s credence in an event with her fair price for a bet on the event. When  $f$  is a miser’s fair price for a bet that pays  $a$  if  $E$  and  $b$  if  $\sim E$ , then  $c(E) = (f - b)/(a - b)$ . For a non-miser,  $c(E) = (u(f) - u(b))/(u(a) - u(b))$ .

E.g., suppose Jim is a miser and his fair price for the following bet is \$10:

Rain	$\sim$ Rain
\$5	\$12

We can use Ramsey’s principle of exact credence to calculate his credence that it will rain:  $c(\text{Rain}) = (10 - 12)/(5 - 12) = 0.29$ .

## 2.4 OBJECTIONS TO THE BELIEF/DESIRE THESIS

The first tenet of the Standard Model is the Belief-Desire Thesis: the view that an agent’s reasons for actions are wholly a function of her beliefs and desires at the time of action. Here are some ways one might object to the thesis. As an exercise, you should think about how a proponent

of the Standard Model would respond to these objects and defend the Belief/Desire Thesis. As another exercise, you should try to think of other ways one might object to the Belief/Desire Thesis.

1. **THE OBJECTION FROM EGOISM:** Rational egoism is the doctrine that rational agents always promote their own self-interest. The Belief/Desire Thesis entails rational egoism, since it says that we should always strive to choose acts that, in light of our beliefs, will most effectively bring about the satisfaction of our all-things-considered desires. Since rational egoism is implausible, the Belief/Desire Thesis is false.
2. **THE OBJECTION FROM OTHER TYPES OF REASONS FOR ACTION:** According to the Belief/Desire Thesis, only beliefs and desires rationalize actions. But emotions, habits, past plans and intentions, future regrets, and beliefs about the good can also rationalize actions. Therefore, the Belief/Desire Thesis is false.
3. **THE OBJECTION FROM SILENCING:** According to the Belief/Desire Thesis, desires are always *pro tanto* reasons for action. But some desires should be entirely "silenced" by moral considerations so that they don't rise to the level of reasons at all. (e.g., a sadist's desire to torture another person is silenced by moral considerations, so he has no reason to torture at all.) Therefore, the Belief/Desire Thesis is false.
4. **OBJECTION FROM INAUTHENTIC DESIRES:** According to the Belief/Desire Thesis, desires—including unendorsed first-order desires—are always *pro tanto* reasons for action. But unendorsed first-order desires (i.e., first-order desires that conflict with second-order desires) are more like whims, impulses, or compulsions. They don't express our true or authentic attitudes. Since only true or authentic desires give us *pro tanto* reasons for action, the Belief/Desire Thesis is false.

### 3 CHOICE FUNCTIONS

A choice function is a function that, for any belief/desire profile, associates each decision with a choice set  $C(A)$  that contains all and only those acts that an agent may rationally choose. In other words, a choice function takes as input a decision problem with a menu of options and outputs a set of all and only the choiceworthy options, given the chooser's beliefs and desires.

#### 3.1 PRINCIPLES $\alpha$ AND $\beta$

Amartya Sen proposed the choice functions should obey the following constraint: the relative values of acts should not vary with changes to the menu of options. So, for instance, if I prefer fries to salad when those are the only two options on the menu, I should still prefer fries to salad when soup is added to the menu. And if I prefer soup the most when soup, fries, and salad are the three items on the menu, I should still prefer soup over fries when salad is removed from the menu.

Two principles follow from this constraint:

1. Principle- $\alpha$ : choiceworthy acts should remain choiceworthy when other options are *removed* from the menu.

2. Principle- $\beta$ : if some act remains choiceworthy when more options are *added* to the menu, then any act that was choiceworthy before the new options were added remains choiceworthy after.

$[\alpha]$  and  $[\beta]$  jointly entail  $[R]$ : If  $A$  is choiceworthy in some decision in which  $B$  is an option, then  $A$  is choiceworthy in any decision in which  $B$  is choiceworthy and  $A$  is an option.

There are two important consequence of  $[\alpha]$  and  $[\beta]$ . First, they ensure that the comparative choiceworthiness of any pair of options depends entirely on the acts' comparative choiceworthiness in binary decisions where those are the only options. (E.g., if you want to learn whether I find fries or salad choiceworthy, you just need to offer me a binary decision between fries and salad and see what I choose.) Second, they allow us to derive a preference ranking, which allow us to speak about choiceworthiness and preferences more or less interchangeably:

- $A \succeq B$  iff  $A$  is choiceworthy in  $\{A, B\}$
- $A \succ B$  iff  $A$  is choiceworthy in  $\{A, B\}$  and  $B$  is not
- $A \approx B$  iff  $A$  and  $B$  are both choiceworthy in  $\{A, B\}$

### 3.2 THE BASIC RATIONALITY REQUIREMENTS

Below are some basic rationality requirements (BRRs) we might want to impose on agents. The BRRs can be expressed either as constraints on choice functions or on preference rankings.

BRRs for choice functions:

1.  $[\alpha]$ : choiceworthy acts should remain choiceworthy when other options are *removed* from the menu.
2.  $[\beta]$ : if some act remains choiceworthy when more options are *added* to the menu, then any act that was choiceworthy before the new options were added remains choiceworthy after.
3.  $[T]$ : If  $A$  is choiceworthy in some decision in which  $B$  is an option, and if  $B$  is choiceworthy in some decision in which  $C$  is an option, and if the agent's views about the probabilities of states or the desirabilities of outcomes do not vary depending upon which option she is offered, then  $A$  must be choiceworthy in any decision in which  $C$  is choiceworthy and  $A$  is an option.

BRRs for preferences:

1. **CONSISTENCY**: If  $A \succ B$ , then  $A \succeq B$  and  $B \not\succ A$
2. **REFLEXIVITY OF WEAK PREFERENCE**: For all  $A$ ,  $A \succ A$
3. **TRANSITIVITY**: if  $A \succeq B$  and  $B \succeq C$ , then  $A \succeq C$ . If either of the first two preferences is strict, so is the third.

### 3.3 KNOCKOUT REASONING

$[\alpha]$ ,  $[\beta]$ , and  $[T]$  justify *knockout reasoning*. Suppose you are deciding where to move to in Canada. You *decompose* your decision  $D$  into smaller decisions as follows. What makes this a decomposition is that every act in the decision falls into one and exactly one sub-decision.

West ( $D_1$ )	Prairies ( $D_2$ )	Eastern ( $D_3$ )	Maritimes ( $D_4$ )	Territories ( $D_5$ )
BC	Alberta Saskatchewan Manitoba	Ontario Quebec	New Brunswick Nova Scotia Newfoundland Prince Edward Island	NWT Yukon Nunavut

The *knockout principle* says that losers in any of these sub-decisions  $D_n$  can be ignored in the larger decision  $D$ , since they would lose even if they were among the competitors in  $D$ . (For instance, if  $C(\text{Eastern Canada}) = \{\text{Quebec}\}$ , so that Ontario is unchoiceworthy in that binary decision, then Ontario cannot be choiceworthy in the larger decision.) Moreover, winners in  $D$  must be among the winners of one of these sub-decision. (For instance, if Newfoundland is ultimately a choiceworthy option, then it must be the case that Newfoundland was among the choiceworthy options in  $D_4$ .

### 3.4 VIOLATIONS OF $[\alpha]$ AND $[\beta]$

Here are some ways that agents often violate  $[\alpha]$  and  $[\beta]$ . In response to these violations, we must judge either that these agents' behavior is irrational and defend  $[\alpha]$  and  $[\beta]$ , or judge that  $[\alpha]$  and  $[\beta]$  bad rules of rationality and defend these agents' behavior. As an exercise, explain why each of these is a violation of  $[\alpha]$  and  $[\beta]$ . Next, consider which you think is the right argumentative strategy to take. Would you defend  $[\alpha]$  and  $[\beta]$  or defend this behavior?

- **THE DECOY EFFECT:** when agents are offered the following options—(A) \$59 for *The Economist* online access, (B<sup>+</sup>) \$125 for *The Economist* in print and online access, and (C) \$0 for no subscription—they tend to choose (A). But when offered the additional option (B<sup>-</sup>) \$125 for *The Economist* in print only, they change their preference from (A) to (B).
  - Schematically:  $A$  and  $B$  are of the same type, but  $A$  is a little better, so  $A$  would be chosen from  $\{A, B\}$ .  $C$  would be chosen from  $\{A, C\}$ . However,  $A$  would be chosen from  $\{A, B, C\}$ .
- **THE COMPROMISE EFFECT ("Mr. InBetween"):** when agents are offered the following options—(A) a large popcorn and (C) a small popcorn—they tend to choose (C). But when (B) a medium popcorn is added, they change their preference from (A) to (B).
  - Schematically:  $C(A, C) = \{C\}$ , but  $C(A, B, C) = \{B\}$ , where  $A$  is the luxury option,  $C$  is the economy option, and  $B$  is "Mr. InBetween."
- **THE SIMILARITY EFFECT:** suppose you are deciding between buying the *World Book*, *Encyclopedia Britannica*, or no encyclopedia. When you're choosing between *World Book* and not buying an encyclopedia, you prefer the *World Book*. When you're choosing between *Encyclopedia Britannica* or no encyclopedia, you prefer *Encyclopedia Britannica*, but when you're choosing between all three, you prefer no encyclopedia, since the decision makes you think about all the negative features of the different encyclopedias.

- Schematically:  $A$  is similar to  $B$  and both differ from  $C$ .  $A$  is chosen from  $\{A, C\}$ .  $B$  is chosen from  $\{B, C\}$ . But  $C$  is chosen from  $\{A, B, C\}$ .
- The similarity effect is a combination of comparing differentiating features and the negativity bias.

### 3.5 MONEY PUMP ARGUMENTS

A common way of explaining why agents who violate the BRRs are irrational is the money pump argument: we can show that an agent whose preferences violate the BRRs will pay money to make a series of exchanges that result in her ending up right where she started.

To explain the money pump argument, we introduced the *Small Improvements Principle* (SIP): if  $A \succeq B$ , then:

1. You would pay a small fee to trade  $B$  for  $A^+$  (where  $A^+$  is a small improvement of  $A$ ), because  $A^+ \succ B$ .
2. You would pay a small fee to trade  $B^-$  for  $A$  (where  $B^-$  is a small diminution of  $B$ ), because  $A \succ B^-$ .
3. You would trade  $B$  for  $A$  for free, because either  $A \succ B$  or  $A \approx B$ .

We can use the SIP to show that someone who violates the BRRs can be turned into a money pump. Suppose, for instance, that Angela's preferences violate transitivity: Apples  $\succ$  Bananas, Bananas  $\succ$  Clementines, and Clementines  $\succ$  Apples. Angela starts with an apple. Because Clementines  $\succ$  Apples, she will pay a small fee (say, 1¢) to trade her apple for a clementine. Now, because Bananas  $\succ$  Clementines, she will pay a small fee (say, 1¢) to trade her clementine for a banana. However, because Apples  $\succ$  Bananas, she will pay a small fee (say, 1¢) to trade her banana for an apple. In the end, she has paid 3¢ to end up with exactly what she started with! As an exercise, prove that someone whose preferences violate *Consistency* can be turned into a money pump.

## 4 THE REPRESENTATION THEOREM

One of the central results of the Standard Model is a representation theorem: if an agent's preference ranking obeys some very intuitively compelling axioms, then her preference ranking can be represented as the preferences of a person who maximizes expected utility relative to some  $(p, u)$  pair.

### 4.1 AXIOMS NEEDED TO PROVE THE REPRESENTATION THEOREM

There are 8 axioms needed to prove the representation theorem:

1. REFLEXIVITY OF WEAK PREFERENCE: For all  $A$ ,  $A \succeq A$
2. CONSISTENCY: If  $A \succ B$ , then  $A \succeq B$  and  $B \not\succ A$
3. FRAME INVARIANCE: Preferences among acts should depend only on the consequences that the acts might produce in various states on the world, not on the way that these consequences or the acts themselves happen to be described.

4. TRANSITIVITY: If  $A \succeq B$  and  $B \succeq C$ , then  $A \succeq C$ , and if either of the first two preferences is strict, so is the third.
5. DOMINANCE: If an agent weakly prefers  $A$  to  $B$  in every non-null state of the world, then she weakly prefers  $A$  to  $B$ .
6. SURE-THING PRINCIPLE: If two acts are sure to produce the same outcome when  $E$  fails to occur, then an agent's preference between them should depend only on her preference between outcomes on the assumption that  $E$  does occur.
7. STAKE INDEPENDENCE: The stakes of a bet shouldn't change your preferences between the bets.
8. ARCHIMEDEAN AXIOM: no lexical orderings

## 4.2 APPLYING THE AXIOMS

We have done many problems over the last few weeks that involved using the axioms needed to prove the representation theorem to say as much as we can about an agent's preferences and credences. This is the representation theorem in action: when we do these problems, we show that an agent who obeys the axioms has preferences and credences that can be represented as someone who maximizes expected utility.

Here is a practice problem. Suppose that Angela is offered the following bets on who will win the Big 10 championship:

	UM	OSU	Iowa	Other
Bet-A	Pie	Tart	Cake	Pie
Bet-B	Pie	Cake	Tart	Cake
Bet-C	Cake	Tart	Pie	Pie
Bet-D	Cake	Pie	Tart	Pie
Bet-E	Cake	Cake	Cake	Pie

Suppose you are told the following about Angela's preferences:  $\text{Pie} \succ \text{Cake}$ ,  $\text{Bet-A} \approx \text{Bet-C}$ ,  $\text{Bet-D} \approx \text{Bet-B}$ , and  $\text{Bet-E} \approx \text{Cake}$ . Say everything you can about Angela's credences and utilities.

The best way to go about these problems is to go through each of Angela's preferences between bets, apply the axioms, and see what you can infer.

1. Start with  $\text{Bet-A} \approx \text{Bet-C}$ . Isolate just these two bets:

	UM	OSU	Iowa	Other
Bet-A	Pie	Tart	Cake	Pie
Bet-C	Cake	Tart	Pie	Pie

Applying the STP, we can ignore the Iowa and Other columns. Because  $\text{Pie} \succ \text{Cake}$ , we know that Bet-A is a bet on Iowa and Bet-C is a bet on UM. Since  $\text{Bet-A} \approx \text{Bet-C}$ , we can conclude that Angela thinks that  $\text{UM} = \text{Iowa}$  (by the definition of  $=$ ).

2. Now consider  $\text{Bet-D} \approx \text{Bet-B}$ . Isolate just these two bets:

	UM	OSU	Iowa	Other
Bet-B	Pie	Cake	Tart	Cake
Bet-D	Cake	Pie	Tart	Pie

Applying the STP, we can ignore the Iowa column. Because  $\text{Pie} \succ \text{Cake}$ , we know that Bet-B is a bet on OSU or Other and Bet-D is a bet on UM. Since  $\text{Bet-D} \approx \text{Bet-B}$ , we can conclude that Angela thinks that  $\text{UM} = \text{OSU} \cup \text{Other}$  (by the definition of  $=$ ).

- Now consider  $\text{Bet-E} \approx \text{Cake}$ . This means that Angela is indifferent between having Bet-E and having cake for sure. Notice, however, that Bet-E pays out pie if Other. This must mean that Other is a null event—otherwise, Angela would prefer to bet on Other and try to get pie.

We end up with the following credence ranking:  $\text{UM} = \text{Iowa} = \text{OSU} > \text{Other} = 0$ . This means that the probabilities of UM, Iowa, and OSU must be  $1/3$  each.

As an exercise, use this information to say as much as you can about Angela's expected utilities.

#### 4.3 THE ALLAIS PARADOX

The Allais Paradox is sometimes taken to be a counterexample to the Sure Thing Principle (STP). Because the STP is needed to prove the representation theorem, and the representation theorem justifies the Standard Model, the Allais Paradox is a threat to the Standard Model.

Consider the following two bets:

	1%	10%	89%
$L_1$	\$0	\$5 million	\$0
$L_2$	\$1 million	\$1 million	\$0

When presented with these two bets, most people prefer  $L_1$  to  $L_2$ ; they prefer having a 10% chance of winning \$5 million over a 11% chance of winning \$1 million.

But now consider these two next bets:

	1%	10%	89%
$L_3$	\$0	\$5 million	\$1 million
$L_4$	\$1 million	\$1 million	\$1 million

When presented with these two bets, most people prefer  $L_4$  to  $L_3$ ; they prefer having \$1 million for sure rather than having a chance at winning more money but a 1% chance of winning nothing.

These "Allais preferences" ( $L_1 \succ L_2$ ,  $L_4 \succ L_3$ ) violate the Sure Thing Principle. The Sure Thing Principle says that when two actions produce the same outcome in an event  $E$ , then our preference between those actions shouldn't hinge on what happens if  $E$ . But Allais preferences *do* take into account what happens in the event that two actions produce the same outcome.

As an exercise, explain whether you think that the Allais paradox is a convincing counterexample to the Sure Thing Principle.

## 5 DUTCH BOOKS

Dutch books are a way of vindicating the idea people who try to maximize expected utility must have credences that obey the laws of probability. According to the Dutch Book Theorem, a miser who tries to maximize expected utility using beliefs that violate the laws of probability will act in ways that are sure to leave her worse off than some alternative would *in all possible circumstances*.

### 5.1 IDENTIFYING AND CONSTRUCTING DUTCH BOOKS

If a person has credences that do not obey the laws of probability, then they can be suckered into buying/selling bets where they are guaranteed to lose money. So, if you encounter someone with probabilistically incoherent credences, you should get into hustler mode and think about how to Dutch book them.

A few important things to remember when constructing Dutch books:

- A miser's fair price for a bet is her expected payoff for the bet.
- A special case of Ramsey's principle of exact credence: if  $c(E) = x$ , then a miser should be willing to pay  $\$x$  for a bet that pays  $\$1$  if  $E$  and  $\$0$  if  $\sim E$ .
- If a bet is fair, then a person should be willing to take either side of the bet. If you think  $\$0.25$  is a fair price for a bet that pays  $\$1$  if  $E$  and  $\$0$  if  $\sim E$ , then you should be willing to buy the bet for  $\$0.25$  and sell the bet  $\$0.25$ .

Suppose Jim's fair prices for the following bets on who will win the Big 10 championship are as follows:

	UM	Northwestern	Other	Fair price
Bet-A	\$7	\$1	\$7	\$3.40
Bet-B	\$3	\$23	\$3	\$13

It's hard to see immediately whether we can Dutch book Jim. So, the first step is to *normalize* the bets. We do this by first finding a common zero, and then a common unit, for every bet in the book:

	UM	Northwestern	Other	Fair price	Normalization
Bet-A*	\$1	\$0	\$1	\$0.4	$-1, \div 6$
Bet-B*	\$0	\$1	\$0	\$0.5	$-3, \div 20$

This clearly shows that Jim's credences violate the laws of probability. Bet-A is a bet on UM or Other—since his fair price for Bet-A\* is  $\$0.4$ , we can conclude that his credence in (UM or Other) is 0.4. This means that his bet in  $\sim(\text{UM or Other})$  should be 0.6—but it's not! Since his fair price for Bet-B (a bet on Northwestern, which is the same as  $\sim(\text{UM or Other})$ ) is  $\$0.5$ , his credence in  $\sim(\text{UM or Other})$  is 0.5.

Here is how to Dutch book Jim. We buy a bet from him that pays  $\$1$  if UM or Other and  $\$0$  if Northwestern for  $\$0.4$ . Next, we buy a bet from him that pays  $\$1$  if Northwestern and  $\$0$  if UM or Other for  $\$0.5$ . Jim will get  $\$0.9$  from selling these bets to us, but that is 10¢ less than he will be paying out, since he will have to pay out a dollar no matter what! (Poor Jim.)

Here is what things look like from Jim's perspective:

UM	Northwestern	Other	Profit from selling
-\$1	\$0	-\$1	\$0.4
\$0	-\$1	\$0	\$0.5
-\$1	-\$1	-\$1	\$0.9