

Social Choice and Public Economics

Greg Leo

This textbook was created for my 4420 “Public Economics” course at Loyola Marymount. This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.



This allows others to remix, tweak, and build upon this work non-commercially as long as attribution is given to this work. Derivative works must be licensed under the same terms. You can find more information about the License at the following link: <https://creativecommons.org/licenses/by-nc-sa/4.0/>

Contents

I	Framework	9
1	Binary Relations and Preferences	9
1.1	Definition	9
1.2	Properties of Relations	9
1.3	Preference Relation	10
1.4	Indifference and Strict Preference	11
1.5	Why Complete and Transitive?	11
1.6	Indifference Curves and Other Sets	13
1.7	Key Topics	14
2	Utility	14
2.1	Definition	15
2.2	Ordinal Utility / Cardinal Utility	16
2.3	Key Topics	17
3	Public Decision Models	17
3.1	Public vs Private	17
3.2	Ordinal Models	18
3.3	Cardinal Models	19
3.4	Key Topics	20
4	Pareto	20
4.1	Pareto Dominance	20
4.2	Pareto as a Relation	21
4.3	Strict Pareto Dominance	22
4.4	Pareto Efficiency	22
4.5	Pareto Efficiency Under Cardinal Preferences	23
4.6	Geometry of Pareto Efficiency	23
4.7	Key Topics	25
II	Ordinal Public Choice	26

5	Social Preferences and Preference Aggregation	26
5.1	Definition	27
5.2	Preference Aggregation Rule	27
5.2.1	Dictatorship	28
5.2.2	Unanimity Rule	29
5.2.3	Majority Rule	29
5.2.4	Condorcet Paradox	30
5.2.5	Copeland's Method	30
5.2.6	Borda	31
5.3	Key Topics	32
6	Preference Aggregation Properties	33
6.1	Basic Properties	33
6.2	Classifying Preference Aggregation Rules	33
6.2.1	Dictatorship	33
6.2.2	Unanimity	33
6.2.3	Majority Rule	33
6.2.4	Copeland's Method	34
6.2.5	Borda	34
6.2.6	A Chart of Preference Aggregation Rules	34
6.3	Independence of Irrelevant Alternatives	34
6.4	Why IIA Matters	35
6.4.1	Example: Borda Count	35
6.4.2	Example: Copeland's Method	36
6.5	Arrow's Impossibility	36
6.6	Exercises	37
6.7	Key Topics	37
7	Making Choices	38
7.1	Social Choice from Preference Aggregation	38
7.2	Some Social Choice Functions	38
7.2.1	Dictatorship- Social Choice	38
7.2.2	Unanimity Rule- Social Choice	39
7.2.3	Plurality Vote- Social Choice	39

7.2.4	Borda Count- Social Choice	40
7.3	Key Topics	41
8	Social Choice Function Properties	41
8.1	Basic Properties	41
8.2	Classifying Social Choice Functions	42
8.2.1	Unanimity and Majority	42
8.2.2	Plurality Vote	42
8.2.3	Borda	43
8.2.4	A Chart of Social Choice Functions	43
8.3	Why IIA Matters for Social Choice	44
8.4	Arrow's Impossibility Again	44
8.5	Key Topics	45
9	Strategic Voting	45
9.1	Manipulation	46
9.2	Another Impossibility	46
9.3	Key Topics	46
III	Cardinal Social Choice	48
10	Welfare Functions	48
10.1	Pareto Efficiency Under Cardinal Preferences	48
10.2	Geometry of Pareto Efficiency	48
10.3	Welfare	50
10.4	Social Choice and Preference Aggregation	51
10.5	Common Welfare Functions	51
10.5.1	Utilitarian Welfare	51
10.5.2	Rawlsian Welfare	52
10.5.3	Nash Welfare	53
10.6	Complete, Transitive, Non-Empty, IIA	54
10.7	Monotonicity and Pareto Efficiency	54
10.8	Comparison of Welfare Functions	55
10.9	Welfare Contours	55

10.10	Key Topics	57
11	Randomization and Side-Payments	57
11.1	Randomization	57
11.2	Convex Combinations	59
11.3	Convex Sets	60
11.4	Convex Hulls and Achievable Utilities	61
11.5	The Pareto Frontier	62
11.6	Maximizing Nash Welfare	63
11.7	Optimizing Rawlsian Welfare	66
11.8	Optimizing Utilitarian Welfare	67
11.9	Ex-ante / Ex-Post	68
11.10	Side-Payments	69
11.11	Feasible Utility Combinations using Side Payments	69
11.12	Optimizing Welfare with Side Payments	69
11.13	Key Topics	71
IV	Public Goods	72
12	Introduction to Game Theory	72
12.1	Games	72
12.2	Best Response	74
12.3	Nash Equilibrium	75
12.4	Key Topics	76
13	Public Goods Homogeneous Preferences	76
13.1	Motivating 2x2 Game	76
13.2	Public Goods- Two People	78
13.2.1	Equilibrium	78
13.2.2	Utilitarian Max	80
13.2.3	Tax	81
13.3	Public Goods- Many People	81
13.3.1	Equilibrium	81
13.3.2	Utilitarian Max	82

13.4 Key Topics	82
14 Public Goods Heterogeneous Preferences	83
14.1 Public Goods- Two People, Different Utilities	83
14.1.1 Equilibrium	83
14.1.2 Utilitarian Max	84
14.2 Many People, Different Utilities	85
14.2.1 Equilibrium	85
14.2.2 Utilitarian Max	86
14.3 Voting	87
14.4 Key Topics	89
15 Mechanism Design for Public Goods	90
15.1 Formal Model	90
15.1.1 Efficiency	91
15.2 Mechanism	91
15.2.1 An aside: Incentive-Compatibility and the Revelation Principle	91
15.3 Some Mechanisms	92
15.3.1 Unanimous Agreement with Fixed-Split	92
15.3.2 Median Mechanism with Fixed Split	94
15.3.3 VCG “Pivot” Mechanism	96
15.4 A Naive Approach to Efficiency and Budget Balance	98
15.5 An Impossibility	99
15.6 Key Topics	99
V Exercises	101
16 Exercises	102
16.1 Exercises for Chapter 1	102
16.2 Exercises for Chapter 2	103
16.3 Exercises for Chapter 3	104
16.4 Exercises for Chapter 4	104
16.5 Exercises for Chapter 5	106
16.6 Exercises for Chapter 6	107

16.7 Exercises for Chapter 7	107
16.8 Exercises for Chapter 8	108
16.9 Exercises for Chapter 9	109
16.10Exercises for Chapter 10	109
16.11Exercises for Chapter 11	111
16.12Exercises for Chapter 12	113
16.13Exercises for Chapter 13	114
16.14Exercises for Chapter 14	114
16.15Exercises for Chapter 15	115
17 Solutions	117
17.1 Solutions for Chapter 1	117
17.2 Solutions for Chapter 2	119
17.3 Solutions for Chapter 3	121
17.4 Solutions for Chapter 4	122
17.5 Solutions for Chapter 5	124
17.6 Solutions for Chapter 6	125
17.7 Solutions for Chapter 7	127
17.8 Solutions for Chapter 8	127
17.9 Solutions for Chapter 9	129
17.10Soulutions for Chapter 10	130
17.11Solutions for Chapter 11	132
17.12Solutions for Chapter 12	134
17.13Solutions for Chapter 13	134
17.14Solutions for Chapter 14	135
17.15Solutions for Chapter 15	136
 VI Appendix	 138
 VII Appendix	 139
 A Optimization	 139
A.1 Unconstrained One-Dimensional Optimization	139

A.2	Unconstrained Multi-Dimensional Optimization	140
A.3	Constrained Multi-Dimensional Optimization	142
A.4	Contours	143
A.5	Monotonicity	145
A.6	Three Possibilities for an Optimal Point	145
A.7	Slopes of Contours	147
A.8	Solution - Two Dimensional Constrained Optimization	148
A.9	Gradients	149
A.10	First Order Condition and Solution	149
A.11	Lagrange	150
A.12	Examples	151
A.13	Exercises	152
A	More Preference Aggregation Rules	152
A.1	Plurality Vote	152
A.2	Veto	153

Part I

Framework

In this part, we will look at some of the framework we will use throughout the rest of the course.

1 Binary Relations and Preferences

1.1 Definition

In mathematics, a binary relation is a concept that describes a relationship between things. They allow us to express various kinds of relationships.

Definition 1.1: Relation. Formally, a **binary relation** R on A is a subset of the Cartesian product $A \times A$. That is, $R \subseteq A \times A$. If $(a, b) \in R$, then we say that a is related to b by R , often denoted as aRb .^a

^aTechnically, a relation can be *between* two different sets A and B , but in this course we are usually representing relationships among elements of a single set.

This is a somewhat intense definition, but relations are very familiar. Here are some examples of mathematical binary relations.

Example 1.1: Siblings. Let R be a relation on the set of all people. If Laura l and Mike m are siblings, then $(l, m) \in R$. We can also write lRm . Similarly $(m, l) \in R$ and we can write mRl .

Example 1.2: Friends. Friend of: Let R be a relation on the set of people where aRb means "person a is a friend of person b ". If Michael m and Sarah s are friends, then mRs . Similarly sRm .

Example 1.3: Height. Human height: Let R be the "at least as tall as relation on the set of people" where aRb means "person a is at least as tall as person b ". For example, if John j is taller than Alice a , then jRa . Notice that unlike the previous two examples, we would not say aRj since Alice is not at least as tall as John.

As we can see, binary relations can capture a wide range of relationships.

1.2 Properties of Relations

Notice how in the examples **Example 1.1** (friends) and **Example 1.2** (siblings) there is a symmetry to the relation. If person a is a sibling of person b then person b is also a sibling of a . The same is the case with friends (I think!). In either case, if aRb , then bRa . We say that such a relation is *symmetric*. Can you think of some other relations on the set of humans that are symmetric?

There are many properties such as *symmetry* that we should know about. Here is a list of some properties a relation can have.

Definition 1.2: Reflexive. A relation R on a set A is reflexive if every element is related to itself, i.e., **Formally:** $\forall a \in A, (a, a) \in R$.

Definition 1.3: Complete. A relation R on a set A is total if every pair of elements is related in at least one direction.
Formally: $\forall a, b \in A, aRb$ or bRa or both.

Definition 1.4: Transitive. A relation R on a set A is transitive if a is related to b and b is related to c , then a is related to c .
Formally: $\forall a, b, c \in A, aRb \& bRc \Rightarrow aRc$.

Definition 1.5: Symmetric. A relation R on a set A is symmetric if any time a is related to b , then b is also related to a .
Formally: $\forall a, b \in A, aRb \Rightarrow bRa$.

Definition 1.6: Asymmetric. A relation R on a set A is asymmetric if any time a is related to b , then b is *not* related to a .
Formally: $\forall a, b \in A, aRb \Rightarrow b \not R a$.

1.3 Preference Relation

A preference relation is a set of statements about outcomes, objects, or pairs of bundles. The statement x is at least as good as y is shortened to $x \succsim y$.

Definition 1.7: Rational Preference Relation. A rational preference relation is a *complete* and *transitive* preference relation \succsim where we interpret the statement $a \succsim b$ as “ a is at least as good as b ”.

Info 1.1: What it means to be rational.. There is a lot of misunderstanding about the formal meaning of the word rational in economics, even among economists’ textbook writers. Rationality has little to do with self-interest, being fully informed, or happiness. Though rationality certainly does not preclude these things.
Rational consumers have preferences. Preferences allow the consumer to rank alternatives (ties are allowed). They can have any ranking they want. Rational consumers choose the highest ranked alternative among the set of alternatives they can afford.
Economists represent these rankings with a utility function that gives higher ranked alternatives a higher score. Representing preferences with a utility function allows economists to use the tools of mathematics to study choices.

1.4 Indifference and Strict Preference

Let (x_1, x_2) represent bowls of ice cream x_1 scoops of vanilla and x_2 scoops of chocolate.

Suppose that someone likes a scoop of vanilla more than a scoop of chocolate. Then the following would be true for them: $(1, 0) \succ (0, 1)$. They might also like *any* number of scoops of vanilla more than the same number of chocolate. Then the following would also be true for their preferences: $(2, 0) \succ (0, 2)$ and $(3, 0) \succ (0, 3)$ and $(100, 0) \succ (0, 100)$.

The following is true for a consumer who does not care about flavor at all just the total amount of ice cream: $(1, 0) \succ (0, 1)$, $(0, 1) \succ (1, 0)$. Notice that we have both $(1, 0) \succ (0, 1)$ and $(0, 1) \succ (1, 0)$. That is, a scoop of vanilla is just as good as a scoop of chocolate and a scoop of chocolate is just as good as a scoop of vanilla. When this is the case, we say that the consumer is **indifferent** and write $(1, 0) \sim (0, 1)$.

Definition 1.8: Indifference Relation. $a \sim b$ when $a \succsim b$ and $b \succsim a$. “ a is indifferent to b ”.

If a consumer is not indifferent between two things, we say that they have **strict** preference.

Definition 1.9: Strict Preference Relation. $a \succ b$ when $a \succsim b$ and $b \not\succsim a$. “ a is strictly preferred to b ”

Note that \succsim is symmetric and \succ is asymmetric.

1.5 Why Complete and Transitive?

You might wonder why completeness and transitivity are the two key assumptions we make. Here’s why:

Economics is about choice. We assume people draw on their preferences to make these choices. Suppose we have the following complete and transitive preference relation on the set $\{a, b, c\}$.

$$a \succ b, a \succ c, b \succ c, a \succ a, b \succ b, c \succ c$$

Below, I have created a plot of these preferences where there is an arrow pointing from one letter to another if the first is preferred to the second. For instance, there is an arrow pointing from a to b since $a \succ b$. I have left off the arrows from each letter to themselves since they do not add much to this figure. Just know that technically they should also be there.

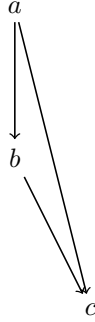


Figure 1.1: A Complete and Transitive Relation on $\{a, b, c\}$

Look how such a complete and transitive relation creates a natural ordering of the objects. Things higher up, like a , are better than everything lower. Even when we get more objects and some indifferences, the same kind of shape appears again. Let's plot the following complete and transitive relation (it is much easier to visualize with the graph).

$$\begin{aligned} a \succsim b, b \succsim a, a \succsim c, a \succsim d, b \succsim c, b \succsim d, c \succsim d, d \succsim c, a \succsim e, \\ b \succsim e, c \succsim e, d \succsim e, a \succsim a, b \succsim b, c \succsim c, d \succsim d, e \succsim e \end{aligned}$$

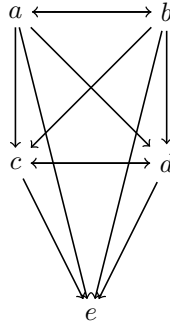


Figure 1.2: A Complete and Transitive Relation on $\{a, b, c, d, e\}$

One nice property of such a shape is that, whatever set of objects a consumer might have to choose from, there is some object in that subset that is at least as good as everything else in that set. In the above example, for instance, a and b are at least as good as everything in $\{a, b, c, d, e\}$. b is at least as good as everything in $\{b, d, e\}$. Whatever set we pick, there is at least one object like that. For whatever set the consumer might be asked to choose from, there is at least one **best** object—something they would be happy to choose.

Definition 1.10: Best. x is **best** from some set B (that includes x) if $x \succsim y$ for every y in B .

We sometimes denote the set of **best** outcomes/options from a set as $C(B)$. This is called the **choice function**. It is the set of best things from a set. Or rather, the things the consumer would

be willing to choose. As an example, for the preferences graphed above, $C(a, b, c, d, e) = \{a, b\}$ and $C(b, c, e) = \{b\}$.

What if we make the relation incomplete by removing $a \succsim c$, what is best from the set of $\{a, c\}$? There is nothing as good as everything else, because the consumer has no idea how to compare a and c . That is $C(\{a, c\}) = \emptyset$. So, somewhat trivially, when a relation is not complete, there are menus of objects that the consumer cannot choose from. In a less trivial way, this also happens when we make the relation intransitive.

Suppose we have the following complete but **intransitive** preference relation on the set $\{a, b, c\}$.

$$a \succsim b, b \succsim c, c \succsim a, a \succ a, b \succ b, c \succ c$$

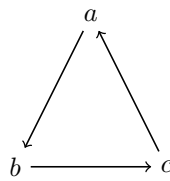


Figure 1.3: A Complete and Intransitive Relation on $\{a, b, c\}$

What is best from $\{a, b, c\}$? That is, what is $C(\{a, b, c\})$. There is nothing at least as good as everything else. a is not at least as good as c , b is not at least as good as a , c is not at least as good as b . What would the consumer choose?!? We have $C(\{a, b, c\}) = \emptyset$. Intransitive preferences create these kinds of cycles (look at the figure again), and when there are cycles, there are sets that the consumer cannot choose from.

1.6 Indifference Curves and Other Sets

For every object, we can use the preference relation to define a few sets. $\succsim(x)$ is the set of objects that is at least as good as x . $\succ(x)$ is the set of objects that is strictly better than x . $\sim(x)$ is the set of objects indifferent to x .

Definition 1.11: Weakly Preferred Set. The set of points weakly preferred to x is:
 $\succsim(x) = \{y | y \in X, y \succsim x\}$

Definition 1.12: Strictly Preferred Set. The set of points strictly preferred to x is:
 $\succ(x) = \{y | y \in X, y \succ x\}$

Definition 1.13: Indifference Set. The set of points indifferent to x is: $\sim(x) = \{y | y \in X, y \sim x\}$

Sets of indifferent bundles are very important in studying preferences. We call such a set of bundles an “indifference curve”. We use indifference curves to visualize preferences. Note: There are many indifference curves. We only sketch a few to get an idea of the “shape” of the preferences.

1.7 Key Topics

- Know what a **Relation** is and how it is used to represent relationships in economics and mathematics.
- Know what it means for a relation to be **Reflexive, Complete, Transitive, Symmetric,** and **Asymmetric**.
- × (Not On Final.) Know what it means for a relation to be **Symmetric**, and **Asymmetric**.
- Be able to determine whether a relation from everyday life is Reflexive, Complete, Transitive, Symmetric, and Asymmetric as in *Exercises 1.1-1.3*.
- Be able to determine whether a formally described relation is Reflexive, Complete, Transitive, Symmetric, and Asymmetric as in *Exercises 1.5-1.6*.
- Be able to **graph a relation** as in *Exercise 1.9*.
- Understand the definition and use of indifference sets, strictly preferred sets, and weakly preferred sets.
- × (Not On Final.) **Draw** indifference sets, strictly preferred sets, and weakly preferred sets for some described preferences as in *Exercises 1.7-1.8*.
- Understand how a preference relation is used and defined and how the weak preference relation \succsim can also describe **strict preference** \succ and **indifference** \sim .
- Given a weak preference relation, write the strict preference relation as in *Exercise 1.10*.
- Given a weak preference relation, write the indifference relation as in *Exercise 1.11*.
- Understand what properties are needed for a preference relation to be called rational and why those properties are important.

2 Utility

A utility function is a way to assign “scores” to bundles, so that better bundles according to \succsim get a higher score. Utility functions allow us to use familiar tools of mathematics to study preferences.

Let’s return to our plot of a complete and transitive preference relation from the last chapter. Recall that here, things higher up are better than anything lower down. It is possible to graph preferences this way as long as preferences are complete and transitive. This time, let’s add some numbers to each level of the graph where things higher up get higher numbers.

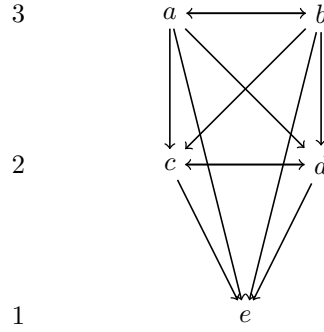


Figure 2.1: A Complete and Transitive Relation on $\{a, b, c, d, e\}$ with Utility

Notice that the number represents "how good" an object is here. a and c get a number 3. They are both indifferent to each other, but strictly better than everything else. c and d get the number 2. They are indifferent to each other but strictly better than e which gets the number 1. We can think of these numbers as "scores" that represent the preferences. In fact, this is precisely what we call a utility function.

2.1 Definition

Definition 2.1: Utility Function. A utility function $U(x)$ represents preferences \succsim when, for every pair of bundles x and y , $U(x) \geq U(y)$ if and only if $x \succsim y$.

Example 2.1: Utility Function Example. Suppose we have the preference relation plotted above.

$$\begin{aligned} a \succsim b, b \succsim a, a \succsim c, a \succsim d, b \succsim c, b \succsim d, c \succsim d, d \succsim c, a \succsim e, \\ b \succsim e, c \succsim e, d \succsim e, a \succ a, b \succ b, c \succ c, d \succ d, e \succ e \end{aligned}$$

Written more succinctly, $a \sim b \succ c \sim d \succ e$.

Some utility functions that represent these preferences are $U(a) = 10, U(b) = 10, U(c) = 5, U(d) = 5, U(e) = 0$ and also $U(a) = 12, U(b) = 12, U(c) = 10, U(d) = 10, U(e) = -100$.

Note that utility functions simply represent the underlying preference relation \succsim of a consumer. When there is a large number of alternatives, the preference relation itself can be cumbersome to work with. However, a utility function can effectively characterize a preference relation in a succinct way.

Example 2.2: Perfect Substitutes Preferences. Suppose a consumer consumes red apples r and green apples g . They like green apples twice as much as red apples, so they would always give up two red apples in exchange for one green apple.

For combinations of red apples and green apples (r, g) this consumer has preference where, for example:

$$\begin{aligned}(2, 0) &\sim (0, 1) \\ (2, 1) &\succ (0, 1) \\ (0, 1) &\succ (1, 0)\end{aligned}$$

We can summarize these preferences with the utility function $u(r, g) = r + 2g$.

Example 2.3: Perfect Complements Preferences. Suppose a consumer consumes only apple pies. An apple pie is made from apples a and crusts c . It takes exactly 2 apples and 1 crust to make a pie.

For combinations of apples and crusts (a, c) this consumer has preference where, for example:

$$\begin{aligned}(2, 0) &\sim (0, 1) \text{ (Since both make zero pies.)} \\ (2, 1) &\sim (2, 2) \text{ (Since both make one pie.)} \\ (4, 2) &\succ (2, 1) \text{ (Since the first makes two and the second makes one pie.)}\end{aligned}$$

We can summarize these preferences with the utility function $u(r, g) = \min \left\{ \frac{1}{2}a, c \right\}$.

2.2 Ordinal Utility / Cardinal Utility

Often, the magnitude of utility is meaningless and only the relationships between scores matter. In this case, we say that the utility is **ordinal**. In [Example 2.1](#), the fact that $U(a) = 10$ and $U(c) = 5$ does not imply that a is twice as good as c . In fact, in the second set of utilities, a gets a utility that is only 1.2 times greater than c .

Sometimes, however, there is meaningful information encoded in a particular representation. Suppose that we have a consumer who consumes two things. t tacos and m money, and their preferences can be represented by $u(t, m) = \sqrt{t} + m$. The utility of the combination $u(4, 10) = 12$ is the same as the utility of $u(0, 12) = 12$. In terms of preferences $(4, 10) \sim (0, 12)$. The utility function directly encodes the amount of money (and no tacos) that some combination is worth to the consumer. **Utility is measured in terms of dollars.** In this sense, the bundle $(36, 18)$ which has utility $u(36, 18) = 24$ is worth "twice" as much as the bundle $u(4, 10) = 12$. The utility is measured in terms of some tangible thing, in this case money. When this is the case, we say that the utility function is **cardinal**.

Definition 2.2: Ordinal Utility. Utility function in which utility numbers have no meaning beyond relative comparisons.

Definition 2.3: Cardinal Utility. Utility function in which utility numbers are measured in terms of something with meaningful magnitude (like money).

One common utility function that we will use in this course is the *quasi-linear* utility function that measures everything in terms of money.

Definition 2.4: Dollar-Denominated Quasi-Linear Utility. A utility function of the form $u(x, m) = f(x) + m$ where m is money.

With this utility function, the utility number $u = u(x, m)$ says "the combination (x, m) is worth the equivalence of $\$u$ to this person."

2.3 Key Topics

- Understand how a utility function can be used to *represent* preferences and be able to discuss this in everyday language as in *Exercise 2.6*.
- Given a preference relation, write down a utility function that represents it as in *Exercise 2.1*.
- Be able to turn a preference relation into a utility function as in *Exercise 2.7*
- × **(Not On Final.)** Given a utility function, determine what is true about the underlying preferences as in *Exercises 2.3 and 2.4*.
- × **(Not On Final.)** Given a simple utility function, sketch indifference curves as in *Exercises 2.7, 2.8 and 2.9*. 2.5 is also relevant but harder to sketch.
- × **(Not On Final.)** Understand what it means for a utility function to be **ordinal** vs. it being **cardinal**.

3 Public Decision Models

3.1 Public vs Private

We now begin our study of public economics in earnest. The way in which public economics differs from the type of economics you might have studied in a course like intermediate economics is that in this course, we focus on situations, outcomes, or choices that affect more than one person. Here, we will differentiate between *private outcomes* and *public outcomes*.

Example 3.1: Private Outcome. Alice is in her studio apartment on Saturday afternoon. She decides to microwave a leftover fish to enjoy for lunch. Besides us, she is the only one who will ever know this happened.

When studying private outcomes, we assume that an individual assesses the available options according to their preferences and chooses their favorite alternative. This is the end of the story with private choice. Choices are either optimal, or they are not.

This course examines the complexities that arise when outcomes are **public** in nature. A **public outcome** impacts multiple individuals. In contrast to private choice, what is optimal for one person may be suboptimal for others.

Example 3.2: Public Outcome. Alice is at work on a Monday afternoon. She decides to microwave a leftover fish to enjoy for lunch. The lingering smell of warm fish reduces office productivity for three days.

3.2 Ordinal Models

The real world is complex. The value of creating theoretical models is that they simplify scenarios down to their core elements. For many scenarios we want to study in this course, I think we can get away with focusing on three elements:

- Who are the people involved?
- What are the potential outcomes?
- What are each person's preferences over those outcomes?

For *ordinal models*, we define only ordinal preference relations over the outcomes.

Definition 3.1: Ordinal Model. An *ordinal public outcome model* is: O : the set of outcomes. P : the set of people. And for every person $i \in P$: \succsim_i their preferences over the set O .

Let's look at an example. In this scenario, Alice and Bob are co-workers. Alice sometimes microwaves fish. Bob hates the smell of microwaved fish. Here we have people Alice a and Bob b and there are two outcomes "Alice microwaves fish" (y) and "Alice does not microwave fish" (n).

Example 3.3: Microwaving Fish.

$$\begin{aligned} P &= \{a, b\} \\ O &= \{\text{yes}, \text{no}\} \\ \text{yes} &\succ_a \text{no} \\ \text{no} &\succ_b \text{yes} \end{aligned}$$

Let's look at a slightly more complex example. Alice and Bob share the office kitchen. Sometimes, it needs to be cleaned. It can be cleaned by only one person, or the work can be shared. However, both prefer that the kitchen be clean, even if that means doing the work alone. Let's formalize this model.

Example 3.4: Cleaning the Kitchen.

$$\begin{aligned}P &= \{a, b\} \\O &= \{\text{both}, \text{Alice}, \text{Bob}, \text{neither}\} \\ \text{Bob} \succ_a \text{both} \succ_a \text{Alice} \succ_a \text{neither} \\ \text{Alice} \succ_b \text{both} \succ_b \text{Bob} \succ_b \text{neither}\end{aligned}$$

3.3 Cardinal Models

In some situations, we need or have a little bit more information– not only relative preferences over outcomes, but also the strength of those preferences measured through some common cardinal utility measure, like the value of each outcome in terms of dollars. To model these scenarios, we use a *cardinal model*. Instead of defining the preference relation \succsim_i for each person, we define their utility function u_i over the outcomes.

Definition 3.2: Cardinal Model. A *cardinal public outcome model* is: O : the set of outcomes. P : the set of people. And for every person $i \in P$: $u_i()$ their cardinal utility function over the set O .

Let's update [Example 3.3](#) and [Example 3.4](#) to cardinal models. For both, let u_a be the utility of Alice and u_b be the utility of Bob. Let's assume that these utilities are measured in terms of dollars so that we have a valid means for assessing the magnitudes of utility. That is, these are cardinal utilities.

Example 3.5: Microwaving Fish: Utility Version.

$$\begin{aligned}P &= \{a, b\} \\O &= \{\text{yes}, \text{no}\} \\ u_a(o) &= \begin{cases} 10 & o = \text{yes} \\ 9 & o = \text{no} \end{cases} \\ u_b(o) &= \begin{cases} 1 & o = \text{yes} \\ 10 & o = \text{no} \end{cases}\end{aligned}$$

One way for us to interpret these utilities is that while Alice would pay up to \$1 to be able to microwave fish, Bob would pay up to \$9 to prevent it.

Example 3.6: Cleaning the Kitchen: Utility Version.

$$\begin{aligned}P &= \{a, b\} \\O &= \{\text{both, Alice, Bob, Neither}\} \\u_a(o) &= \begin{cases} 12 & \text{if } o = \text{both} \\ 10 & \text{if } o = \text{Alice} \\ 25 & \text{if } o = \text{Bob} \\ 5 & \text{if } o = \text{neither} \end{cases} \\u_b(o) &= \begin{cases} 12 & \text{if } o = \text{both} \\ 25 & \text{if } o = \text{Alice} \\ 10 & \text{if } o = \text{Bob} \\ 5 & \text{if } o = \text{neither} \end{cases}\end{aligned}$$

3.4 Key Topics

- × **(Not On Final.)** Understand the elements of an **Ordinal Model**.
- × **(Not On Final.)** Understand the elements of a **Cardinal Model**.
- × **(Not On Final.)** Be able to modify a ordinal model to change the preferences, add people, or change the outcomes as in *Exercises 3.1 and 3.2*.
- × **(Not On Final.)** Be able to modify a cardinal model to change the preferences, add people, or change the outcomes as in *Exercises 3.3 and 3.4*.

4 Pareto

4.1 Pareto Dominance

Imagine yourself as a benevolent ruler, making decisions for society. How would you choose outcomes? One thing that I think should be true of every *benevolent* ruler's preferences is that if there were two outcomes and one was at least as good for *everyone* as the other, the ruler should prefer that better outcome.

Think about the cleaning example in [Example 3.4](#). How would you choose what to implement? Having both clean seems fair, but, perhaps, maybe only having one clean is more efficient. I don't think there is a clear best outcome, but I think essentially everyone would agree that having no one clean is not a desirable outcome. **Any** of the other outcomes are better for both Alice and Bob.

This notion of being *at least as good for everyone* is what we call "Pareto domination." An outcome o' Pareto dominates o if o' is at least as good for everyone. Formally:

Definition 4.1: Pareto Dominates. An outcome o' Pareto dominates o if $o' \succsim_i o$ for all $i \in P$.

4.2 Pareto as a Relation

Let's define Pareto dominance as a relation. We will say aPb if outcome a Pareto dominates b according to the definition above.

Example 4.1: Pareto Dominance in Cleaning the Kitchen. In [Example 3.4](#). For convenience, let's simplify the outcome names from $O = \{both, Alice, Bob, neither\}$ to $O = \{ab, a, b, n\}$. We can now define the Pareto relation P on this set: $abPn$, aPn , bPn , $abPab$, aPa , bPb , nPn .

Is P a complete and transitive relation? It is **always transitive** as long as all of the individuals in the model have transitive preferences. This is because, for example, if *everyone* likes a over b and everyone likes b over c then since everyone has transitive preferences, everyone will like a over c and hence a will Pareto dominate c .

However, Pareto dominance is **not always complete**. In fact, in the example above, it is not complete. There is no relationship between ab and a or between ab and b or between a and b . Pareto dominance cannot compare these outcomes.

Info 4.1: Pareto Dominance is Incomplete. The **Pareto dominates** relation is always transitive but not always complete.

Since Pareto dominance is incomplete, for any given set of options, there may not be an outcome that Pareto dominates all the others. If there were, it would be pretty clearly the best outcome.

Let's look at an example of Pareto dominance.

Example 4.2: Example of Pareto Dominance for Two People.

$$a \succ_1 c \succ_1 b \succ_1 d \succ_1 e$$

$$b \succ_2 d \succ_2 a \succ_2 c \succ_2 e$$

Everything Pareto dominates e . b Pareto dominates d . a Pareto dominates c . But otherwise, the outcomes are not comparable in Pareto terms. The Pareto dominance relation P is:

$$aPc, aPe, bPd, bPe, cPe, dPe, \\ aPa, bPb, cPc, dPd, ePe$$

Let's plot this relation using the same type of graph from the chapter on preferences.

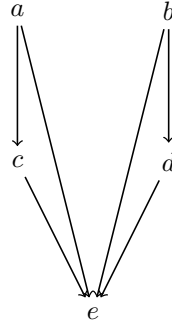


Figure 4.1: A graph Pareto dominance relation from **Example 4.2**

4.3 Strict Pareto Dominance

Just like with preferences, where we defined the strict preference relation such that $a \succ b$ any time $a \succsim b$ but $b \not\succsim a$, we can do the same with Pareto dominance.

Definition 4.2: Strictly Pareto Dominates. An outcome o' Strictly Pareto dominates o if $o'Po$ but $o \not P o'$.

What will this mean in terms of preferences of the individuals? Well, first, $o'Po$ says everyone should like o' at least as well as o . But at the same time, $o \not P o'$ says that it is not the case that everyone likes o at least as well as o' . This means at least one person must like o' strictly better than o . (Take a moment to convince yourself of this.) Combining these, o' will strictly Pareto dominate o if everyone likes o' at least as well **and** at least one person likes it strictly more.

Definition 4.3: Strictly Pareto Dominates. An outcome o' Strictly Pareto dominates o if $o' \succsim_i o$ for all $i \in P$ and there is some $i \in P$ such that $o' \succ_i o$.

In the example above, everything strictly dominates e . a strictly dominates c and b strictly dominates d . Notice that in each case, on the graph, an outcome that is strictly dominates by some other outcome has at least one one-way arrow leading into it.

4.4 Pareto Efficiency

As we have seen, Pareto dominance is not complete. In that sense, for any given set of options, there may not be an outcome that Pareto dominates all the others. If there were, it would be pretty clearly the best outcome. However, just like in our case of the cleaning example, we can at least use Pareto dominance to eliminate the clearly undesirable options. What are those clearly undesirable options? **The ones that are strictly Pareto dominated.** This is because for any outcome that is strictly Pareto dominated there must be a way to make everyone at least as well off and at least one person strictly better off. That's a win/win.

One great thing about Pareto efficiency is that, even though there will not always be some outcome that Pareto dominates all others, **there will always be at least one Pareto**

efficient outcome.

Definition 4.4: Pareto Efficiency. A outcome o is Pareto efficient if there is no other outcome o' that strictly Pareto dominates o .

If we graph preferences, such as in [Figure 4.1](#), a Pareto efficient outcome has no one-way arrow leading into it. We can also define Pareto efficiency in terms of the preferences of individuals since we have defined strict Pareto dominance in terms of individual preference above.

Definition 4.5: Pareto Efficiency in Terms of Individual Preferences. A outcome o is Pareto efficient if there is no other outcome o' such that for all people $o' \succsim_i o$ and for at least one person $o' \succ_i o$.

In [Example 4.2](#), the Pareto efficient outcomes are a and b . From a there is no way to make someone better off without making someone strictly worse off. The same goes for b . Looking at the graph, they are also the only outcomes that have no one-way arrow that points to them.

In our cleaning example, a , b and ab are all Pareto efficient since, starting from any of these, there is no way to make someone strictly better off without also making someone strictly worse off.

4.5 Pareto Efficiency Under Cardinal Preferences

When preferences are Cardinal, that is, the preferences are given a magnitude through the utility function that is measured in terms of some natural “measuring stick” (like money) we can still use Pareto efficiency. We just need to rewrite the definition in terms of utilities rather than the preference relation.

Definition 4.6: Pareto Efficiency. A outcome o is Pareto efficient if there is no other outcome o' such that strictly Pareto dominates o . That is, there is no o' such that $U_i(o') \geq U_i(o)$ for all $i \in P$ and $U_j(o') > U_j(o)$ for at least one $j \in P$.

It is important to note that Pareto efficiency does not imply fairness or equity. An allocation can be Pareto efficient even if it is highly unequal. For example, look at the cardinal version of the “Microwaving Fish” model in [3.5](#). Alice likes to microwave fish, but Bob *really* hates the smell of warm fish. Like, he throws up. Both outcomes are Pareto efficient, but letting Alice cook the fish seems a little unfair.

Thus, while achieving Pareto efficiency is often a goal of policy interventions. Policymakers will often need consider other criteria, such as equity and fairness, when designing policies to social outcomes. We will return to this later. For now, let’s look at a slightly more complex model to solidify understanding about Pareto efficiency.

4.6 Geometry of Pareto Efficiency

When we are using cardinal utilities, we can visualize Pareto dominance and Pareto efficiency. We begin with a plot of the utility pairs from our running example [Example 3.6](#) shown in [Figure 4.2](#).

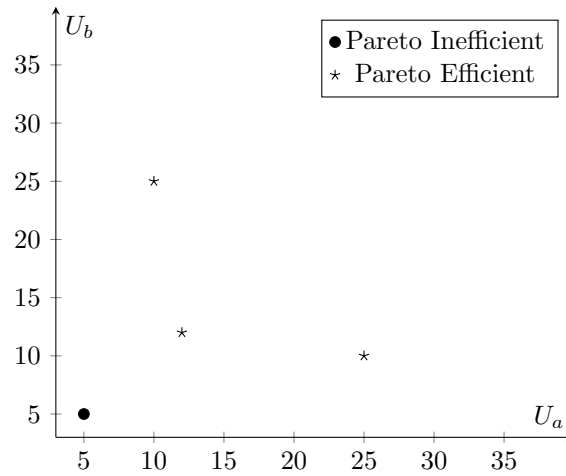


Figure 4.2: Cleaning Utility Pairs

What does it look like when a choice combination is not Pareto efficient? Have a look at [Figure 4.3](#) which shows the possible points that Strictly Pareto dominate $(5, 5)$ from [Example 3.6](#). Notice that $(12, 12)$, $(25, 10)$ and $(10, 25)$ are all in the blue region. They all strictly Pareto dominate $(5, 5)$.

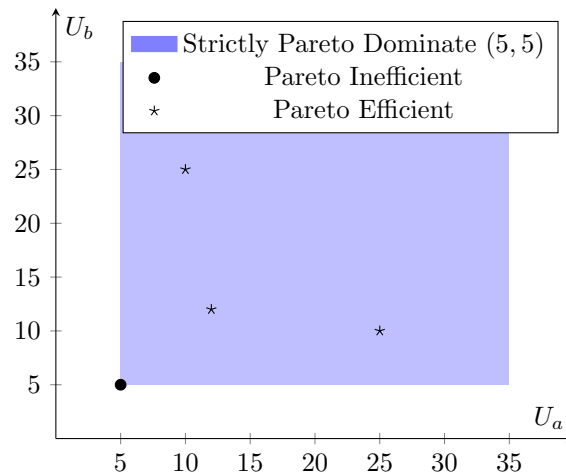


Figure 4.3: Outcomes that Strictly Pareto Dominate $(5, 5)$ from [Example 3.6](#).

Here, $(5, 5)$ is not Pareto efficient because there are outcomes that strictly Pareto dominate it.

On the other hand, if we repeat this exercise with the point $(12, 12)$, we see there are no outcomes in the blue region. It is Pareto efficient! This shown in [Figure 4.4](#).

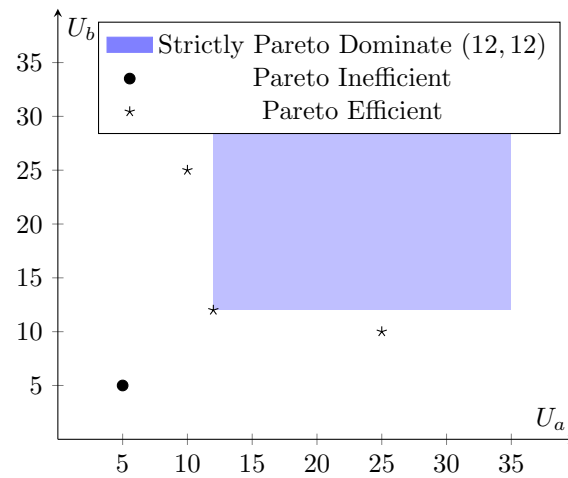


Figure 4.4: Outcomes that Strictly Pareto Dominate $(12, 12)$ from **Example 3.6**.

4.7 Key Topics

- Understand what it means for an outcome to **Pareto dominate** another in terms of individual preferences and be able to determine, given preferences, when one outcome Pareto dominates another.
- × (Not On Final.) Be able to **graph the Pareto dominance relation** as in exercise 4.1.
- Know what it means for an outcome to *strictly* **Pareto dominate** another.
- Be able to determine which outcomes Pareto dominate and strictly Pareto dominate others as in exercises 4.9 and 4.10.
- Know what it means for an outcome to be **Pareto efficient**.
- Determine which outcomes are Pareto efficient as in exercises 4.2, 4.3, 4.4, 4.7, 4.11.
- × Know how to determine when outcomes are Pareto efficient in a cardinal model and plot the outcomes as in exercise 4.8.

Part II

Ordinal Public Choice

In this part, we will look at how societies can aggregate their preferences to make decisions.

For convenience and throughout *Part II*, unless otherwise specified, we will assume all individuals in the models have **strict preferences** over all distinct outcomes. This simplifies some of the work in these sections. However, nothing about what we will study here requires to make this assumption. Essentially everything could be extended to include the possibility of indifference in individual preferences.

To be a bit more formal about this. For the rest of this part, in every model, we will assume that for each individual $i \in P$, there are no two *distinct* outcomes $x \neq y$ such that $x \sim_i y$. This means that everyone's preferences will look like this: $x \succ_i y \succ_i z \dots$ with no indifference. This will make things a little easier, but know that much of what we will discuss will apply to situations where indifference is allowed.

When there is no indifference, then our definition of Pareto efficiency and Pareto dominance can be updated. Recall that x Pareto dominates y if everyone likes it at least as well. But if x and y are distinct and everyone has strict preferences, then the only way this can happen is if everyone likes x **strictly more** than y .

Definition 4.7: Pareto Dominance Under Strict Preferences. If everyone's preferences are strict, then x Pareto dominates y if $x \succ_i y$ for all $i \in P$.

We can update Pareto efficiency accordingly as well.

Definition 4.8: Pareto Efficiency Under Strict Preferences. If everyone's preferences are strict, then x is Pareto efficient if there is no y that everyone likes strictly better. That is, not y such that $y \succ_i x$ for all $i \in P$.

5 Social Preferences and Preference Aggregation

Pareto efficiency is a lovely property. I think it is somewhat indisputably desirable as a property to strive for. Unfortunately, as we have seen, Pareto efficiency is not always enough to make a choice among outcomes. It is a property that is transitive but not complete. It helps us choose outcomes, but it often fails to provide complete guidance on what to choose. In a sense, it is not enough to help us construct a preference relation on how *we* might assess the outcomes.

We might want to try to extend Pareto efficiency in a way that offers a complete preference relation over outcomes that we can use to assess options on behalf of society. We call such preferences **Social Preferences** since they are preferences about the outcomes that affect society.

5.1 Definition

Definition 5.1: Social Preference Relation. A social preference relation \succsim_* is a complete and transitive relation on the set of outcomes O . It is used to evaluate outcomes by an administrator to evaluate outcomes on behalf of society.

Think about your preferences about the outcomes in [Example 3.3](#) and [Example 3.4](#). What social preferences do you have over these outcomes?

Example 5.1: Social Preferences for [Example 3.4](#). In [Example 3.4](#) “Cleaning the Kitchen” a social preference relation might be the following:

$$\text{both} \succ_* \text{Alice} \sim_* \text{Bob} \succ_* \text{neither}$$

Notice that this is a preference relation that does not match the preferences of any of the individuals in the model.

5.2 Preference Aggregation Rule

Technically, a social preference relation can be *any* complete and transitive relation on the outcomes. But really, an administrator should consider how their constituents care about the outcomes, that is, the social preference relation should somehow be constructed by referencing the individual preferences.

Definition 5.2: Preference Aggregation Rule. A **preference aggregation rule** (also known as a social welfare function) is a way to turn individual preferences into the social preference relation. Formally, it is a *mapping* from the set of possible individual preferences over the outcomes into a social preference relation over the outcomes.

These can be somewhat hard to define using formal notation, so I will use intuitive descriptions of various rules when necessary. Let’s look at a few preference aggregation rules in the context of the following two models:

In this section, I present many preference aggregation rules. These are meant to show you *some* of the possibilities and demonstrate how different rules can prioritize different kinds of goals. These should get your mind working on different ways to construct interesting rules. By the end of the section, I hope you will be able to come up with your own rules and think about how they might work in practice.

I will use the following two running examples throughout this section. Before you begin reading through each rule, think about what you would choose as the social preferences for each example.

Example 5.2: Example 1.

There are three people. They have these preferences:

- 1: $a \succ b \succ c$
- 2: $a \succ c \succ b$
- 3: $c \succ a \succ b$

Example 5.3: Example 2.

There are five people. They have these preferences:

- 1: $a \succ c \succ b$
- 2: $a \succ c \succ b$
- 3: $b \succ c \succ a$
- 4: $b \succ a \succ c$
- 5: $c \succ a \succ b$

5.2.1 Dictatorship

A dictatorship where there is one person whose preferences completely determine the social preferences. The preferences of everyone else are ignored.

This may seem like a strange rule, but I think it is surprisingly common. For example, the faculty of the LMU Economics department occasionally meets for lunch. The faculty rotates, being the *one* who gets to choose the restaurant. We could vote every time, or submit our rankings, or do something else. But instead, we rotate being the "dictator". It is simple and decisive, and for scenarios like this that happen over and over, rotating through "dictators" makes what would usually be a very unfair rule fair in the long-run.

Definition 5.3: Dictatorship. Pick a person $i \in P$. The social preferences are that person's preferences. $\succ^* = \succ_i$.

Example 5.4: Example 1.

Let's assume person 1 is the dictator.

- 1: $a \succ b \succ c$
- 2: $a \succ c \succ b$
- 3: $c \succ a \succ b$

$$a \succ^* b \succ^* c$$

Example 5.5: Example 2.

For the examples, let's assume person 1 is the dictator.

- 1: $a \succ c \succ b$
- 2: $a \succ c \succ b$
- 3: $b \succ c \succ a$
- 4: $b \succ a \succ c$
- 5: $c \succ a \succ b$

$$a \succ^* c \succ^* b$$

Another nice thing we can say about dictatorships is that at least they always give us a complete and transitive social preference ordering. Let's not undervalue that. There are reasonable rules that do not return a complete and transitive ordering.

5.2.2 Unanimity Rule

The unanimity rule is the rule that results from using **Pareto dominance** to create social preferences. Here, we rank one option over another if it Pareto dominates: everyone agrees it is better. Since preferences are strict here, that means everyone likes it strictly more.

Definition 5.4: Unanimity Rule. $x \succ^* y$ if $x \succ_i y$ for everyone. That is, one outcome is ordered above another if *everyone* thinks it is better.

Example 5.6: Example 1.

- 1: $a \succ b \succ c$
- 2: $a \succ c \succ b$
- 3: $c \succ a \succ b$

$$a \succ^* b$$

This is an incomplete social preference. It says nothing about the preferences between a, c and b, c .

Example 5.7: Example 2.

- 1: $a \succ c \succ b$
- 2: $a \succ c \succ b$
- 3: $b \succ c \succ a$
- 4: $b \succ a \succ c$
- 5: $c \succ a \succ b$

Unanimity gives us nothing here!

As we can see, and probably as you expected, the unanimity rule is almost never complete! However, it is transitive. If everyone likes x over y and everyone likes y over z they will certainly like x over z (assuming they all have transitive preferences). Now let's look at a rule that is complete, but (perhaps surprisingly) may be intransitive.

5.2.3 Majority Rule

Aka. Pairwise Voting.

The pairwise majority social welfare function is a rule where the preference of each pair of alternatives is determined by the majority of voters. In the late 18th century, this was considered the sort of definition of social preferences. If most of society likes x over y then we can say that $x \succ^* y$ truly represents "society's" preferences between x and y . There's a problem though...

Definition 5.5: Majority Rule. $x \succ^* y$ if more than half of the people prefer x to y . In other words, x is better than y in the social preferences if it wins a pairwise vote between those outcomes.

Example 5.8: Example 1.

- 1: $a \succ b \succ c$
- 2: $a \succ c \succ b$
- 3: $c \succ a \succ b$

$$a \succ^* c \succ^* b$$

Example 5.9: Example 2.

- 1: $a \succ c \succ b$
- 2: $a \succ c \succ b$
- 3: $b \succ c \succ a$
- 4: $b \succ a \succ c$
- 5: $c \succ a \succ b$

$$a \succ^* c \succ^* b$$

In these two cases, we get a complete and transitive social preference. Since some outcome will always win each pairwise vote, the rule will always produce a complete social preference relation, but now we will see that it will not always be transitive.

5.2.4 Condorcet Paradox

The Condorcet paradox, named after the French mathematician and philosopher Marquis de Condorcet, who discovered in the late 18th century, that social preferences can be cyclic, even if the individual preferences are not.

Example 5.10: Condorcet Cycle.

- 1: $a \succ b \succ c$
- 2: $b \succ c \succ a$
- 3: $c \succ a \succ b$

$$a \succ^* b, b \succ^* c, c \succ^* a$$

Intransitive Social Preference Relation

5.2.5 Copeland's Method

Copeland's method, is named after Arthur Copeland who popularized it in the 1950s, though it appears to date back to work by Ramon Llull in the early 1300s! It handles Condorcet's paradox by assigning 1 point for each pairwise win and 0.5 points for each pairwise tie (though in all the examples below there will be an odd number of people and so there will be no ties). The candidate with the highest total points is the winner.

This is also how many sports tournaments are conducted. In sports we can easily get intransitivity. Team a beats b , b beats c , but c beats a . You might see how that would be a problem with picking a winner. For instance, a similar method is used to determine which teams advance from the "group" round to the "knockout" round in the World Cup.

Definition 5.6: Copeland's Method. Conduct a pairwise vote for every pair. If an outcome wins a vote, add one to its score. (If there is a tie add $\frac{2}{2}$.) The social preferences are ranked by score. So if x gets a higher score than y it is ranked higher.

Example 5.11: Example 1.

1: $a \succ b \succ c$
 2: $a \succ c \succ b$
 3: $c \succ a \succ b$
 a beats b . a beats c . c beats b .
 a wins 2 votes. c wins 1 vote. b wins 0 votes.

$$a \succ^* c \succ^* b$$

$$a :, b :, c :$$

Example 5.12: Example 2.

1: $a \succ c \succ b$
 2: $a \succ c \succ b$
 3: $b \succ c \succ a$
 4: $b \succ a \succ c$
 5: $c \succ a \succ b$
 a wins 2 votes. c wins 1 vote. b wins 0 votes.

$$a \succ^* c \succ^* b$$

Example 5.13: Condorcet Cycle.

1: $a \succ b \succ c$
 2: $b \succ c \succ a$
 3: $c \succ a \succ b$
 All outcomes win one of their pairwise votes.

$$a \sim^* b \sim^* c$$

5.2.6 Borda

Plurality vote focuses on maximizing the number of people who get their favorite outcome and veto attempts to minimize the number of people who get their least favorite outcome. What if we want to create a rule that balances both of these goals? We can think of the Borda count as a rule that attempts to balance both goals.

Definition 5.7: Borda Count. Each rank is assigned a certain number of points, with higher ranks receiving more points. The option outcome with the highest total points wins. If there are 3 outcomes, we might assign 3 for a first-rank, 2 for a second, and 1 for a third.

With Borda count, it is sort of “traditional” to assign a score of 1 for the last rank and work up from there. For example, using scores of 3, 2, 1 for three outcomes. However, the outcome will be the same regardless of what numbers we use as long as they are all one-apart. For example, we could use 1, 0, −1.

Example 5.14: Example 1.

1: $a \succ b \succ c$
 2: $a \succ c \succ b$
 3: $c \succ a \succ b$
 $a : 3 + 3 + 2 = 8$
 $b : 2 + 1 + 1 = 4$
 $c : 1 + 2 + 3 = 6$

$$a \succ^* c \succ^* b$$

Example 5.15: Example 2.

1: $a \succ c \succ b$
 2: $a \succ c \succ b$
 3: $b \succ c \succ a$
 4: $b \succ a \succ c$
 5: $c \succ a \succ b$
 $a : 11, b : 9, c : 10$

$$a \succ^* c \succ^* b$$

5.3 Key Topics

- Understand what a **Social Preference Relation** is and how it might be used.
- Understand what a **Preference Aggregation Rule** is.
- Know the following preference aggregation rules and how they work:
 - Dictatorship
 - Majority Rule
 - Borda Count
- Be able to apply each of the preference aggregation rules above to determine social preferences for a given set of individual preferences as in *Exercises 5.3 and 5.4*.
- Know what the **Condorcet Paradox** is and be able to determine if Majority Rule leads to an intransitive social preference as in *Exercise 5.2*.

6 Preference Aggregation Properties

6.1 Basic Properties

There are several properties that we might want in a rule. Let's begin with two properties that ensure that the resulting social preference relation is indeed a rational preference relation:

Definition 6.1: Complete. A preference aggregation rule is **complete** if \succsim_* is complete for all profiles of individual preferences: $(\succsim_1, \dots, \succsim_n)$.

Definition 6.2: Transitive. A preference aggregation rule is **transitive** if \succsim_* is transitive for all profiles of individual preferences: $(\succsim_1, \dots, \succsim_n)$.

I hope I have also argued that Pareto efficiency is a very desirable property in previous chapters. Let's add that to the mix. Here, we will write a property that says that the rule needs to respect Pareto dominance. This definition will be given in the context of models where *everyone has strict preferences*.

Definition 6.3: Pareto Efficient. A preference aggregation rule is **Pareto Efficient** if for every x and y such that for every person $i \in P$, $x \succ_i y$ then $x \succ_* y$. That is, if everyone likes x strictly better than y , then the social preference also strictly prefers x to y .

6.2 Classifying Preference Aggregation Rules

6.2.1 Dictatorship

Since a **Dictatorship** just uses an individual's preferences as the social preference, it is always complete and transitive. It is also Pareto efficient. If everyone likes x over y , so will the dictator, and so $x \succ_* y$. Thus, it is also Pareto efficient.

6.2.2 Unanimity

We have seen that **Unanimity Rule** is transitive, Pareto efficient (by definition), but it is **not complete**.

6.2.3 Majority Rule

We have seen that **Majority Rule** is **complete**. It is also **Pareto efficient**. If everyone prefers x to y then a more than a majority will vote for x , thus $x \succ_* y$. However, due to the Condorcet paradox it is **not transitive**.

6.2.4 Copeland's Method

Copeland's method is **non-dictatorial** since it uses the preferences of everyone to determine the social preferences.

Info 6.1: Methods that use a Score are Complete and Transitive. Any rule that uses a score is complete and transitive.

For these rules, if the score of an outcome is at least as high, then $x \succsim^* y$. Since every pair of outcomes gets a score, we can compare every outcome. Similarly, if the score of x is higher than y , and the score of y is higher than z , then the score of x is higher than z . Thus, these Since Copeland's method assigns scores based on how many pair-wise votes an outcome wins, is **complete** and **transitive**.

It is also **Pareto efficient**. To see this, suppose everyone prefers a to b . We need $a \succ^* b$ which is true if a beats strictly more other outcomes in a pair-wise vote than b . But, since preferences are transitive, anyone who would vote for b over some other outcome would also vote for a over that outcome. Thus, a beats everything b does, plus it also beats b . Thus, the score of a is larger than b .

6.2.5 Borda

Borda count is **non-dictatorial** since it uses the preferences of everyone to determine the social preferences.

Since the Borda count also assigns scores, it is **complete** and **transitive**.

Finally, it is **Pareto efficient** since if everyone agrees that $a \succ_i b$ then a gets a higher score in everyone's preference than b . Thus, the sum of the scores for a must be strictly higher than b and so $a \succ^* b$.

6.2.6 A Chart of Preference Aggregation Rules

Rule	Complete	Transitive	Pareto
Dictatorship	✓	✓	✓
Unanimity Rule	×	✓	✓
Majority Rule	✓	×	✓
Copeland's Method	✓	✓	✓
Borda Count	✓	✓	✓

Table 1: Comparison of Preference Aggregation Rules

6.3 Independence of Irrelevant Alternatives

There's something a little weird about Copeland's Method and Borda count...

Let's start with some preferences:

- Person 1: $a \succ b \succ c$
- Person 2: $b \succ a \succ c$
- Person 3: $c \succ a \succ b$

In both Borda and Copeland's method, the social preferences are $a \succ^* b \succ^* c$. Let's focus on the fact that $a \succ^* b$ here. Let's change Person 2's preference over a and c to be $c \succ a$ instead of $a \succ c$. We get:

- Person 1: $a \succ b \succ c$
- Person 2: $b \succ c \succ a$
- Person 3: $c \succ a \succ b$

In both Borda and Copeland's method, the social preferences are now $a \sim^* b \sim^* c$ and $a \sim^* b$. But, we did not change anything about anyone's preferences over a and b , and yet, the social preference changed. When this is possible for a preference aggregation rule, we say it fails **Independence of Irrelevant Alternatives**.

Definition 6.4: Independence of Irrelevant Alternatives. A preference aggregation rule obeys **Independence of Irrelevant Alternatives** [IIA] if for any two sets of preferences where the preference for x and y is the same between the two sets, they should have the same social preference between x and y .

6.4 Why IIA Matters

6.4.1 Example: Borda Count

Suppose there are 100 people in a society and there are three types of preferences:

- Type 1 (25 People): $a \succ b \succ c$
- Type 2 (40 People): $b \succ c \succ a$
- Type 3 (35 People): $c \succ a \succ b$

In Borda Count, the social preferences are $c \succ^* b \succ^* a$ since the scores are:

- a : $25 * 3 + 40 * 1 + 35 * 2 = 185$
- b : $25 * 2 + 40 * 3 + 35 * 1 = 205$
- c : $25 * 1 + 40 * 2 + 35 * 3 = 210$

But, if we remove a , preferences become:

- Type 1 (25 People): $b \succ c$

- Type 2 (40 People): $b \succ c$
- Type 3 (35 People): $c \succ b$

In Borda Count, the social preferences are $b \succ^* c$, since the scores are:

- b : $25 * 2 + 40 * 2 + 35 * 1 = 165$
- c : $25 * 1 + 40 * 1 + 35 * 2 = 135$

6.4.2 Example: Copeland's Method

In Copeland's method, we can get similar oddities.

- Type 1 (45 People): $a \succ b \succ c$
- Type 2 (15 People): $b \succ c \succ a$
- Type 3 (40 People): $c \succ b \succ a$

In Copeland's Method, the social preferences are $c \succ^* b \succ^* a$, since b beats a , b beats c , and c beats a .

Let's remove a :

- Type 1 (45 People): $b \succ c$
- Type 2 (15 People): $b \succ c$
- Type 3 (40 People): $c \succ b$

In Copeland's Method, the social preferences are $b \succ^* c$. Again, we get a reversal!

6.5 Arrow's Impossibility

Rule	Complete	Transitive	Pareto	IIA
Dictatorship	✓	✓	✓	✓
Unanimity Rule	×	✓	✓	✓
Majority Rule	✓	×	✓	✓
Copeland's Method	✓	✓	✓	×
Borda Count	✓	✓	✓	×

Table 2: Comparison of Preference Aggregation Rules

Our goal in this whole process was to look for rule that built on Pareto efficiency and also filled in the gaps to create a complete and transitive social preference relation. We found two options in Borda and Copeland's method. Unfortunately, we ran into a new problem: IIA.

So far, the only thing we have seen that meets all of our assumptions is a dictatorship! As it turns out, that is the only preference aggregation rule that is complete and transitive, Pareto efficient, and does not violate IIA.

Info 6.2: Arrow's Impossibility Theorem. If there are at least three outcomes, the only preference aggregation rule that is **complete**, **transitive**, **Pareto efficient**, and does not violate **IIA** is a **dictatorship**!

6.6 Exercises

In **Plurality Vote** the preference aggregation rule, the **score** of an outcome is the number of people who rank that outcome highest. Social preferences are determined by score. See [subsection A.1](#) for more info.

In **Veto**, the score of an outcome is the negative of the number of people who rank it last. Social preferences are determined by the score as in the other scoring methods above with a higher score being ranked higher. See [subsection A.2](#) for more info.

Rule	Complete	Transitive	Pareto	IIA
Dictatorship	✓	✓	✓	✓
Unanimity Rule	×	✓	✓	✓
Majority Rule	✓	×	✓	✓
Copeland's Method	✓	✓	✓	×
Borda Count	✓	✓	✓	×
Plurality Vote	✓	✓	×	×
Veto	✓	✓	×	×

Table 3: Comparison of Preference Aggregation Rules

6.7 Key Topics

- Know what it means for a preference aggregation rule to:
 - be **Complete**
 - be **Transitive**
 - be **Pareto Efficient**
 - × (Not On Final.) respect **Independence of Irrelevant Alternatives** [IIA]
- Know that any rule that assigns social preferences based on some score (as in Borda count) will always be complete and transitive.
- Given a preference aggregation rule that is not Pareto efficient, come up with a counter-example demonstrating this as in *Exercises 6.2 and 6.3*.
- × (Not On Final.) Given a preference aggregation rule that is not IIA, come up with a counter-example demonstrating this as in *Exercises 6.2 and 6.3*.
- Know what Arrows impossibility theorem for Preference Aggregation rules says, what properties can only be met by a dictatorship, and what those properties mean.

7 Making Choices

Preference aggregation rules produce a preference relation. In most real-world situations, we do not need a preference relation, we just need to make a choice. There is a concept related to preference aggregation rules that has the focus not of making a preference relation, but rather just making a choice. We call these social choice functions.

Definition 7.1: Social Choice Function. A **social choice function** (also known as a social welfare function) is a way to turn individual preferences into a choice. Formally, it is a *mapping* from the set of possible individual preferences over the outcomes into an outcome or subset of the outcomes- the choice/choices. That is, from the set of possibilities, it picks a winner or winners (if there are ties).

7.1 Social Choice from Preference Aggregation

We learned in [section 1](#) that when a preference relation is complete and transitive, we can use it to make a choice from any subset of the outcomes through the notion of best outcomes. Recall from [Definition 1.10](#) that an outcome x is **best** from some set B according \succsim if, for every other y in the set B , $x \succsim y$. That is, x is best from a set if it is preferred to all other outcomes in the set.

In this sense, if a preference aggregation rule results in a complete and transitive social preference relation \succsim^* , we can use that relation to make a choice from any subset of the outcomes. In this sense, every preference aggregation rule is a social choice function.

Info 7.1: Social Choice Functions from Preference Aggregation Rule. Any preference aggregation rule can be converted into a social choice function by taking the **best** outcomes from the resulting social preference relation.

However, some social choice functions are more suited towards picking a winner than creating an entire preference ordering, and the properties we look for in a social choice function are not the same as those of a preference aggregation rule.

7.2 Some Social Choice Functions

7.2.1 Dictatorship- Social Choice

Definition 7.2: Dictatorship. Pick a person $i \in P$. The social choice is that person's favorite outcome.

Example 7.1: Example 1. Let's assume person 1 is the dictator.

1: $a \succ b \succ c$

2: $a \succ c \succ b$

3: $c \succ a \succ b$

The social choice is a .

Example 7.2: Example 2. Assume person 1 is the dictator.

1: $a \succ c \succ b$

2: $a \succ c \succ b$

3: $b \succ c \succ a$

4: $b \succ a \succ c$

5: $c \succ a \succ b$

The social choice is a .

7.2.2 Unanimity Rule- Social Choice

The unanimity rule is a social choice function that selects an outcome if it is unanimously preferred over another by all individuals.

Definition 7.3: Unanimity Rule- Social Choice. Choose x if for every person and for all other outcomes y $x \succ_i y$ for everyone. That is, an outcome is chosen if *everyone* thinks it is better than every other outcome.

Example 7.3: Example 1.

1: $a \succ b \succ c$

2: $a \succ c \succ b$

3: $c \succ a \succ b$

Unanimity gives no choice here.

Example 7.4: Example 2.

1: $a \succ c \succ b$

2: $a \succ c \succ b$

3: $b \succ c \succ a$

4: $b \succ a \succ c$

5: $c \succ a \succ b$

Unanimity gives no choice here.

7.2.3 Plurality Vote- Social Choice

Plurality vote focuses on the goal of giving as many people as possible their top-ranked outcome.

Definition 7.4: Plurality Vote- Social Choice. The social choice is the outcome which the most number of people rank first.

Example 7.5: Example 1.

- 1: $a \succ b \succ c$
- 2: $a \succ c \succ b$
- 3: $c \succ a \succ b$

a is the social choice.

Example 7.6: Example 2.

- 1: $a \succ c \succ b$
- 2: $a \succ c \succ b$
- 3: $b \succ c \succ a$
- 4: $b \succ a \succ c$
- 5: $c \succ a \succ b$

a and b are social choices.

7.2.4 Borda Count- Social Choice

The Borda count is a social choice function that balances the preferences of individuals by assigning points to ranks. Higher ranks receive more points, and the outcome with the highest total points is chosen.

Definition 7.5: Borda Count- Social Choice. Each rank is assigned a certain number of points, with higher ranks receiving more points. The outcome with the highest total points is the social choice. If there are 3 outcomes, for example, we might assign 3 points for a first-rank, 2 points for a second, and 1 point for a third.

Example 7.7: Example 1.

- 1: $a \succ b \succ c$
- 2: $a \succ c \succ b$
- 3: $c \succ a \succ b$

Scores:

$$a : 3 + 3 + 2 = 8$$

$$b : 2 + 1 + 1 = 4$$

$$c : 1 + 2 + 3 = 6$$

a is the social choice.

Example 7.8: Example 2.

- 1: $a \succ c \succ b$
- 2: $a \succ c \succ b$
- 3: $b \succ c \succ a$
- 4: $b \succ a \succ c$
- 5: $c \succ a \succ b$

Scores:

$$a : 3 + 3 + 1 + 2 = 9$$

$$b : 2 + 2 + 3 + 1 = 8$$

$$c : 1 + 1 + 2 + 3 = 7$$

a is the social choice.

7.3 Key Topics

- Understand what a **Social Choice Function** is and how it differs from an preference aggregation rule.
- Know the following social choice functions and how they work:
 - **Dictatorship**
 - **Plurality Vote**
 - **Borda Count**
- Be able to apply each of the social choice functions above to determine social choices a given set of individual preferences as in *Exercises 7.1 and 7.2*.
- Given a description of a simple social choice function not in the list above determine social choices a given set of individual preferences as in *Exercises 7.3*.

8 Social Choice Function Properties

8.1 Basic Properties

A preference aggregation rule results in a preference relation. For such a relation to be capable of making choices from *any* subset of outcomes, it needs to be complete and transitive. However, a social choice function only needs to pick a winner or winners from the entire set of alternatives.

We can replace *completeness* and *transitivity* with the following property:

Definition 8.1: Nonempty. A social choice function is **nonempty** if the set of choices is nonempty for all profiles of individual preferences: $(\succsim_1, \dots, \succsim_n)$.

In preference aggregation rules, if there is an outcome y such that there is another x that everyone strictly prefers to y , then y could never be ranked highest, since *Pareto efficiency* of the social welfare function will require $x \succ^* y$. In that sense, y should also never *win*— it should never be

the choice. That leads to the following extension of Pareto efficiency to social choice functions. This says that if y is strictly Pareto dominated, it cannot be chosen.

Definition 8.2: Pareto Efficient. A social choice function is **Pareto Efficient** if for every y where there is another outcome x such that every person $i \in P$, $x \succ_i y$ then y cannot be in the set of choices. That is, if everyone likes x strictly better than y , then the social preference also strictly prefers x to y .

Lastly, we can extend independence of irrelevant alternatives to social choice functions:

Definition 8.3: Independence of Irrelevant Alternatives. A social choice function obeys **Independence of Irrelevant Alternatives** [IIA] if for any two sets of preferences where the preference for x and y is the same between the two sets, if x is chosen in the first set and y is not, then y cannot be chosen in the second set.

8.2 Classifying Social Choice Functions

8.2.1 Unanimity and Majority

Among the social choice functions we looked at above, **Unanimity** and **Majority Vote** are **not nonempty**. For instance, neither pick a winner under the Condorcet paradox preferences below:

1: $a \succ b \succ c$

2: $b \succ c \succ a$

3: $c \succ a \succ b$

There is no outcome that is unanimously better than all others, so unanimity does not pick a choice. Similarly, majority vote results in an intransitive cycle where $a \succ^* b$, $b \succ^* c$, $c \succ^* a$, thus there is no outcome at least as good as all others.

They are, however, **Pareto efficient** and **IIA**.

8.2.2 Plurality Vote

Since some outcome always wins the plurality vote, it is **nonempty**.

Plurality vote is Pareto efficient because if some outcome x is preferred by everyone to y , then y cannot be anyone's favorite. Thus, it cannot win.

Plurality vote is **not IIA**. Consider the following sets of preferences:

Set 1:

1: $a \succ b \succ c$

2: $a \succ b \succ c$

3: $b \succ a \succ c$

Set 2:

1: $a \succ b \succ c$

2: $c \succ a \succ b$

3: $b \succ a \succ c$

a is the choice in the first set, and a and b are the choices in the second set, even though everyone has the same preferences over a and b . This violates IIA.

8.2.3 Borda

Since some outcome always wins the Borda count, it is **nonempty**.

It is Pareto efficient since if x is strictly preferred by everyone over y , x must get a strictly higher score, thus, y cannot get the highest score and be a choice.

It is **not IIA**. Consider the following sets of preferences:

Set 1:

- 1: $a \succ b \succ c$
- 2: $b \succ a \succ c$
- 3: $c \succ a \succ b$

Set 2:

- 1: $a \succ b \succ c$
- 2: $b \succ c \succ a$
- 3: $c \succ a \succ b$

a is the choice in set 1, but a , b and c are all choices in set 2. This violates IIA.

8.2.4 A Chart of Social Choice Functions

Rule	Nonempty	Pareto	IIA
Dictatorship	✓	✓	✓
Unanimity Rule	×	✓	✓
Majority Rule	×	✓	✓
Plurality Vote	✓	✓	×
Borda	✓	✓	×

Table 4: Comparison of Social Choice Functions

8.3 Why IIA Matters for Social Choice

This example is identical to the example used to show why IIA matters for preference aggregation. I have duplicated it here for convenience.

Suppose there are 100 in a society and there are three types of preferences:

- Type 1 (25 People): $a \succ b \succ c$
- Type 2 (40 People): $b \succ c \succ a$
- Type 3 (35 People): $c \succ a \succ b$

In Borda Count, the choice is c since the scores are:

- a : $25 * 3 + 40 * 1 + 35 * 2 = 185$
- b : $25 * 2 + 40 * 3 + 35 * 1 = 205$
- c : $25 * 1 + 40 * 2 + 35 * 3 = 210$

But if we remove a . Preferences are:

- Type 1 (25 People): $b \succ c$
- Type 2 (40 People): $b \succ c$
- Type 3 (35 People): $c \succ b$

In Borda Count, the choice is b since the scores are:

- b : $25 * 2 + 40 * 2 + 35 * 1 = 165$
- c : $25 * 1 + 40 * 1 + 35 * 2 = 135$

8.4 Arrow's Impossibility Again

Arrow's impossibility theorem says only a dictatorship can aggregating preferences in a way that is complete, transitive, Pareto efficient and respects IIA. A preference aggregation rule creates a social preference that lets a decision-maker make a choice from *any* subset of the outcomes. That's what a preference relation is good for. It might seem like if we were not worried about being able to make a choice from every subset, but just wanted to make a choice from the whole set of outcomes, it might be easier to find a suitable rule.

Notice in the chart above that even though all a social choice function has to do is pick a winner, we cannot seem to get one that has the three properties we might want. In fact, it is still impossible.

Info 8.1: Arrow's Impossibility Theorem for Social Choice. If there are at least three outcomes, the only social choice function that is **nonempty**, **Pareto efficient**, and does not violate **IIA** is a **dictatorship**!

As far as I can tell, this result was formalized by Denicolò Vincenzo in [2].

8.5 Key Topics

- Know what it means for a social choice function to:
 - be **Nonempty**
 - be **Pareto Efficient**
 - × (Not On Final.) respect **Independence of Irrelevant Alternatives** [IIA]
- Given a social choice function from the list in the key topics of Chapter 7 or based on a simple description that is **not nonempty**, come up with a counter-example demonstrating this as in *Exercises 8.2 and 8.3*.
- Given a social choice function from the list in the key topics of Chapter 7 or based on a simple description that is **not Pareto efficient**, come up with a counter-example demonstrating this as in *Exercises 8.6*.
- × (Not On Final.) Given a social choice function from the list in the key topics of Chapter 7 or based on a simple description that does **not respect IIA**, come up with a counter-example demonstrating this as in *Exercises 8.4 and 8.5*.
- Know what Arrow's impossibility theorem for Social Choice Functions says, what properties can be met only by a dictatorship, and what those properties are.

9 Strategic Voting

So far we have we have looked at instances where preferences are *known*. But what happens if preferences are *not known*? We have to collect them from constituents.

Let's add one more property to the mix. Arrow's impossibility theorem says that even if we allow ties for the "choice" then there is no social choice function that always makes choice that is IIA and Pareto efficient.

However, in the real world, *some* outcome actually needs to get chosen. A real-world social choice function needs to have a way of breaking ties. Any social choice function that breaks ties and just picks one outcome is called **decisive**:

Definition 9.1: Decisive. A social choice function is **decisive** if there is always a single choice for all profiles of individual preferences: $(\succsim_1, \dots, \succsim_n)$.

Note that decisiveness is **stronger** than **nonempty**. For a rule to be decisive, it has to be nonempty **and** there always has to be just one choice. Most real-world social choice functions have some built-in tie-breaking rule. However, we can also easily convert any non-empty social choice function into a decisive one by adding a simple tie-breaking rule. Below, I will use the rule that breaks ties by choosing the outcome lowest in the alphabet.

For example suppose we have the following preferences:

- 1: $a \succ b \succ c$
- 2: $b \succ c \succ a$
- 3: $c \succ a \succ b$

Plurality Vote is **not decisive** since there is a tie for number of first place votes between a and b and c . However, if we add a tie-breaking rule that the outcome lowest in the alphabet wins any tie then the rule is decisive. The choice will be a .

Definition 9.2: Plurality Vote with Alphabetical Tie-Breaker. The social choice is the outcome which the most number of people rank first. If there is a tie, the outcome lowest in the alphabet wins.

9.1 Manipulation

In the example above, notice that if person 2 changed their vote to c instead of b by pretending to have preferences $c \succ b \succ a$ then c would win instead, and person 3 likes this better!

This is an example of **manipulation**.

Definition 9.3: Non-manipulable. A social choice function is **Non-manipulable** if no individual can ever achieve a more preferred outcome by misrepresenting their preferences.

Manipulation can undermine the fairness and accuracy of voting outcomes. It can lead to scenarios where the chosen outcome does not reflect the true preferences of the voters.

9.2 Another Impossibility

Arrow's impossibility (See [Info Box 6.2](#) and [Info Box 8.1](#)) says that we have to be OK with either relaxing Pareto efficiency or IIA. Suppose we were OK with relaxing IIA— I think it is the more natural option to give up. There are many social choice rules that are **decisive** and **Pareto efficient**.

For instance, **Plurality Vote with Alphabetical Tie-Breaker** or **Borda with Alphabetical Tie-Breaker** are both Pareto efficient and decisive. However, they are also both manipulable. We have seen that in the example above for Plurality vote with Alphabetical Tie-Breaker, but the same example works for Borda count. Person 2 would still like to pretend to have preferences $c \succ b \succ a$ giving the win to outcome c instead of outcome a .

As it turns out, as long as a social choice rule is **Pareto efficient** and **decisive** we will always be able to find an example where the rule can be manipulated.

Info 9.1: Gibbard-Satterthwaite. If there are at least three outcomes, the only social choice function that is **decisive**, **Pareto efficient**, and **non-Manipulable** is a **dictatorship**!

9.3 Key Topics

- Know what it means for a social choice function to:
 - be **Decisive**

– be **Non-Manipulable**

- Come up with an example of manipulation given individual preferences and a social choice function with tie-breaking rule as in *Exercises 9.2 and 9.3*.
- Know what the Gibbard-Satterthwaite theorem says, what properties can be met only by a dictatorship, and what those properties are.

Part III

Cardinal Social Choice

In this section we assume that people in our models have cardinal preferences (dollar-denominated utilities) over the outcomes. This gives us lots of new possibilities.

10 Welfare Functions

10.1 Pareto Efficiency Under Cardinal Preferences

When preferences are Cardinal, that is, the preferences are given a magnitude through the utility function that is measured in terms of some natural “measuring stick” like money, we can still use Pareto efficiency. We just need to rewrite the definition in terms of utilities rather than the preference relation.

Definition 10.1: Pareto Efficiency. A outcome o is Pareto efficient if there is no other outcome o' such that strictly Pareto dominates o . That is, there is no o' such that $U_i(o') \geq U_i(o)$ for all $i \in P$ and $U_j(o') > U_j(o)$ for at least one $j \in P$.

It is important to note that Pareto efficiency does not imply fairness or equity. An allocation can be Pareto efficient even if it is highly unequal. For example, look at the cardinal version of the “Microwaving Fish” model in 3.5. Alice likes to microwave fish, but Bob *really* hates the smell of warm fish. Like, he throws up. Both outcomes are Pareto efficient, but letting Alice cook the fish seems a little unfair.

Thus, while achieving Pareto efficiency is often a goal of policy interventions. Policymakers will often need consider other criteria, such as equity and fairness, when designing policies to social outcomes. We will return to this later. For now, let’s look at a slightly more complex model to solidify understanding about Pareto efficiency.

10.2 Geometry of Pareto Efficiency

When we are using cardinal utilities, we can visualize Pareto dominance and Pareto efficiency. We begin with a plot of the utility pairs from our running example [Example 3.6](#) shown in [Figure 10.1](#).

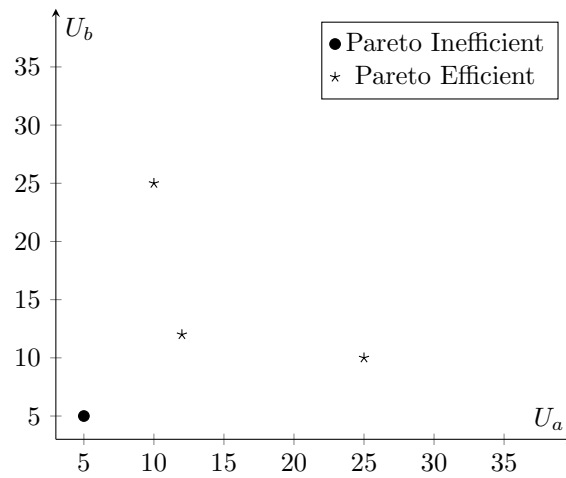


Figure 10.1: Cleaning Utility Pairs

What does it look like when a choice combination is not Pareto efficient? Have a look at **Figure 10.2** which shows the possible points that Strictly Pareto dominate $(5, 5)$ from **Example 3.6**. Notice that $(12, 12)$, $(25, 10)$ and $(10, 25)$ are all in the blue region. They all strictly Pareto dominate $(5, 5)$.

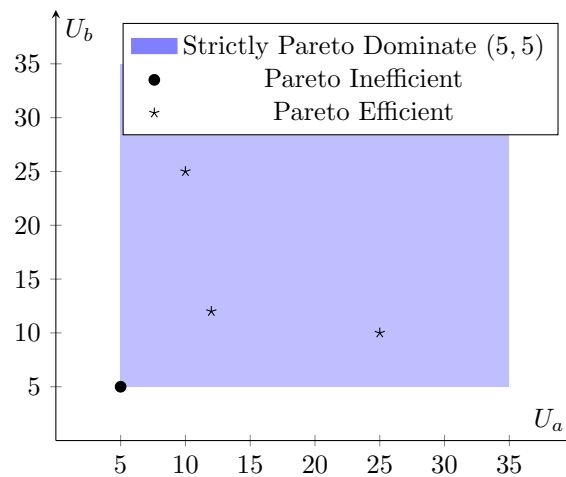


Figure 10.2: Outcomes that Strictly Pareto Dominate $(5, 5)$ from **Example 3.6.**

Here, $(5, 5)$ is not Pareto efficient because there are outcomes that strictly Pareto dominate it.

On the other hand, if we repeat this exercise with the point $(12, 12)$, we see there are no outcomes in the blue region. It is Pareto efficient! This shown in **Figure 10.3**.

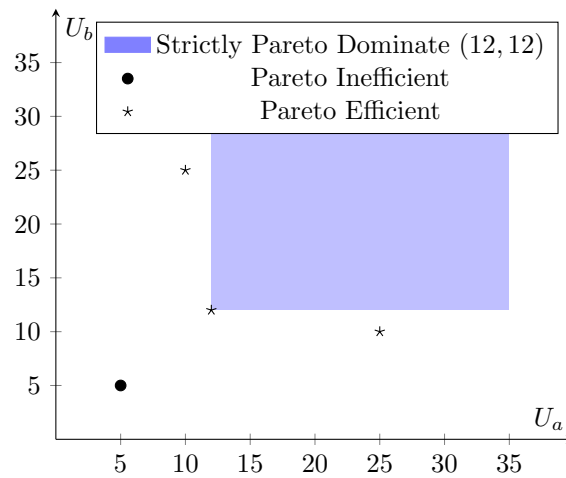


Figure 10.3: Outcomes that Strictly Pareto Dominate (12, 12) from Example 3.6.

10.3 Welfare

As in ordinal models, Pareto Efficiency quickly reaches the end of its usefulness. For instance, in Example 3.4, two of the four outcomes are Pareto efficient: Alice makes the fish and Bob leaves, Alice does not make the fish and Bob stays. How should we decide between these points? Pareto efficiency will not help us.

What do you think about the two options? If you had to pick one to impose, which would you pick and why?

One way to decide among the options is to be formal about our preferences over the possible pairs of utilities that can arise in such a model. In ordinal models, we did this by aggregating individual preferences into a social preference relation.

For a cardinal model, we can do something analogous, aggregate the *utility* preferences of everyone into a sort of *social utility* or what we call **welfare**. The rule or function that converts the combination of individual utilities for an outcome into welfare is called a **welfare function**.

A welfare function is simply a way of "scoring" utility combinations. It is really no different than a utility function for "bundles" you may have learned about in intermediate microeconomics. Here, however, the "bundles" are combinations of utilities for each person.

Think of welfare functions as utility functions for the "planner" or whoever is assessing the outcomes for society. Of course, since any utility function represents some underlying preference, a welfare function creates social preferences over the possible outcomes.

Just like there are many preference aggregation rules, there are *many* possible welfare functions that we can use.

Definition 10.2: Welfare Function. A welfare function W is a function W that maps the cardinal utilities of each person (U_1, U_2, \dots, U_n) into a number— the *welfare*.

Technically, a welfare function is a function of the utilities. The utilitarian welfare function defined below, for example, gets the welfare by taking the average of the utilities. If there are two people and some outcome x provides the utilities $U_1(x) = 10$ and $U_2(x) = 10$ then $W(10, 10) = \frac{10+10}{2} = 10$. So we can write the welfare function as function of the utilities like above $W(10, 10)$ or, to be more succinct, we can write it as a function of the outcomes. Here, $W(x) = W(U_1(x), U_2(x)) = 10$.

10.4 Social Choice and Preference Aggregation

We can apply a welfare function as a preference aggregation rule. To apply it as a preference aggregation rule, simply calculate the welfare of every outcome and rank the outcomes by their welfare. That is $x \succsim y$ when $W(x) \geq W(y)$ and $x \succ y$ when $W(x) > W(y)$.

Suppose $W(a) = 10, W(b) = 5, W(c) = 5$ then the resulting social preferences are $a \succ^* b \sim^* c$.

We can also use a welfare function as a social choice function by make the social choice any of the outcomes for which receive the highest welfare. Formally, an outcome x is a social choice if $W(x) \geq W(y)$ for every other outcome y .

Suppose $W(a) = 10, W(b) = 5, W(c) = 5$ then the resulting social choice is a .

10.5 Common Welfare Functions

Here are some common welfare functions. For each welfare function, we will look at how it ranks the outcomes of the model in 3.6. I have reprinted that example here for convenience.

Example 10.1: Cleaning the Kitchen: Utility Version.

$$\begin{aligned}
 P &= \{a, b\} \\
 O &= \{\text{both, Alice, Bob, Neither}\} \\
 u_a(o) &= \begin{cases} 12 & \text{if } o = \text{both} \\ 10 & \text{if } o = \text{Alice} \\ 25 & \text{if } o = \text{Bob} \\ 5 & \text{if } o = \text{neither} \end{cases} \\
 u_b(o) &= \begin{cases} 12 & \text{if } o = \text{both} \\ 25 & \text{if } o = \text{Alice} \\ 10 & \text{if } o = \text{Bob} \\ 5 & \text{if } o = \text{neither} \end{cases}
 \end{aligned}$$

10.5.1 Utilitarian Welfare

The utilitarian welfare function is rooted in the philosophy of utilitarianism, which was developed by Jeremy Bentham in the 18th century and John Stuart Mill of the 19th century [1, 3]. Utilitarianism aims to maximize overall utility. This function has been widely used in economics to evaluate social welfare outcomes under different policies.

Definition 10.3: Utilitarian Welfare. $W(U_1, \dots, U_n) = \sum_{i=1}^n \frac{1}{n} U_i = \frac{\sum_{i=1}^n U_i}{n}$.

The utilitarian welfare function aims to pick the outcome that maximizes the total amount of value to society. Since the average utility in society is just the total divided by the number of people, if the total goes up, the average will as well. Thus, maximizing average utility is the same as maximizing total utility.

Example 10.2: Utilitarian Welfare **Example 10.1.** Let's use the utilitarian welfare function $W(U_a, U_b) = \frac{1}{2}U_a + \frac{1}{2}U_b$ for **Example 10.1**. Recall that the utilities of each choice combination are:

$$U_a(o) = \begin{cases} 12 & \text{if } o = ab \\ 10 & \text{if } o = a \\ 25 & \text{if } o = b \\ 5 & \text{if } o = n \end{cases}$$

$$U_b(o) = \begin{cases} 12 & \text{if } o = ab \\ 25 & \text{if } o = a \\ 10 & \text{if } o = b \\ 5 & \text{if } o = n \end{cases}$$

To maximize the (average) utilitarian welfare function $W(U_a, U_b) = \frac{1}{2}U_a + \frac{1}{2}U_b$, we need to calculate the total welfare for each combination c :

- For ab :

$$W(ab) = \frac{1}{2}(12) + \frac{1}{2}(12) = 6 + 6 = 12$$

- For a :

$$W(a) = \frac{1}{2}(10) + \frac{1}{2}(25) = 5 + 12.5 = 17.5$$

- For b :

$$W(b) = \frac{1}{2}(25) + \frac{1}{2}(10) = 12.5 + 5 = 17.5$$

- For n :

$$W(n) = \frac{1}{2}(5) + \frac{1}{2}(5) = 5 + 5 = 5$$

Therefore, the social preferences are $a \sim^* b \succ^* ab \succ^* n$ the social choices are a, b .

10.5.2 Rawlsian Welfare

The Rawlsian welfare function has its origins in the philosophical work of John Rawls. In [4], Rawls introduced the concept of the 'veil of ignorance': the idea that, when making decisions about the rules of a society, one should do so without knowing their "position" within it to ensure fairness. The result of this thought experiment is this welfare function that prioritizes fairness above all.

Definition 10.4: Rawlsian (Max-Min) Welfare. $W(U_1, \dots, U_n) = \min\{U_1, U_2, \dots, U_n\}$

Example 10.3: Maximizing Rawlsian Welfare **Example 10.1.** Let's use the Rawlsian welfare function $W(U_a, U_b) = \min\{U_a, U_b\}$ for **Example 10.1**. To maximize the Rawlsian welfare function $W(U_a, U_b) = \min\{U_a, U_b\}$, we need to calculate the minimum utility for each combination c :

- For ab :

$$W(ab) = \min\{12, 12\} = 12$$

- For a :

$$W(a) = \min\{10, 25\} = 10$$

- For b :

$$W(b) = \min\{25, 10\} = 10$$

- For n :

$$W(n) = \min\{5, 5\} = 5$$

Therefore, the social preferences are $ab \succ^* a \sim^* b \succ^* n$ the social choice is ab .

10.5.3 Nash Welfare

Definition 10.5: Nash Welfare. $W(U_1, \dots, U_n) = \prod_{i=1}^n U_i^{\frac{1}{n}}$.

This should only be used when all $U_i > 0$!

Example 10.4: Maximizing Nash Welfare **Example 10.1.** Let's use the Nash welfare function for **Example 10.1**. To maximize the Nash welfare function, we need to calculate the product of utilities for each outcome:

- For ab :

$$W(ab) = (12^{0.5})(12^{0.5}) = 12$$

- For a :

$$W(a) = (10^{0.5})(25^{0.5}) = \sqrt{250}$$

- For b :

$$W(b) = (25^{0.5})(10^{0.5}) = \sqrt{250}$$

- For n :

$$W(n) = (5^{0.5})(5^{0.5}) = 5$$

Therefore, the social preferences are $a \sim^* b \succ^* ab \succ^* n$ the social choices are a, b .

10.6 Complete, Transitive, Non-Empty, IIA

Since any welfare function assigns a score, the welfare, to any outcome, the social preferences that result from applying a welfare function to a set of outcomes is always complete and transitive. Thus, when applied as a preference aggregation rule, a welfare function is **complete** and **transitive**. Furthermore, some outcome or outcomes will always get the highest welfare, if we apply a welfare function as a social choice function, it will always be **nonempty**.

Furthermore, when applied as a preference aggregation rule, welfare functions are also **IIA**. Suppose we have two outcomes x and y where the welfare of x is larger than y , $W(u_1(x), u_2(x)) > W(u_1(y), u_2(y))$. Whatever we do to the utilities of other outcomes, we do not change anyone's utility for x and y then nothing will change about the welfare of x and y . Thus, the resulting social preferences between x and y will not change. Furthermore, if x was a social choice and y was not then $W(x) > W(y)$. As long as utilities for x and y remain the same then it will still be the case that $W(x) > W(y)$ even if we change utilities for other outcomes. Thus, no matter how we change the utilities of those other outcomes, y cannot become a social choice. Thus, applying a welfare function as a social choice function is also **IIA**.

10.7 Monotonicity and Pareto Efficiency

How can we ensure a welfare function is Pareto efficient? That is, it respects Pareto dominance? We can do this by assuming the function is monotonic. That is, it is increasing in the utilities of the individuals.

Definition 10.6: Monotonic. A welfare function is **monotonic** if whenever $u_i(x) \geq u_i(y)$ for all people, then $W(x', y') \geq W(x, y)$ and if, in addition, $u_i(x) > u_i(y)$ for anyone then $W(x) > W(y)$.

Put another way, a welfare function is **monotonic** if whenever x Pareto dominates y then $W(x) \geq W(y)$ and whenever x strictly Pareto dominates y then $W(x) > W(y)$.

Recall that for cardinal models, some outcome a Pareto dominates another b if $U_i(a) \geq U_i(b)$ for all people: $i \in P$. Furthermore, it strictly Pareto dominates if the above condition as true and there is at least one person for whom $U_i(a) > U_i(b)$. Thus, another way to think of monotonicity is that if a Pareto dominates b then $W(a) \geq W(b)$. If it strictly Pareto dominates then $W(a) > W(b)$.

Notice that if everyone strictly prefers x to y then $U_i(x) > U_i(y)$ for everyone. By if the welfare function is monotonic, then $W(x) > W(y)$. Thus, applying if we apply a monotonic welfare function as a preference aggregation rule, then $x \succ^* y$. In this sense, any monotonic welfare function will operate as a Pareto efficient preference aggregation rule. Furthermore, if we apply a monotonic welfare function as a social choice function then any outcome that is strictly Pareto dominated cannot be a social choice. Suppose y is strictly Pareto dominated by x then $W(x) > W(y)$ if the welfare function is monotonic. Thus, y cannot possibly be the social choice. In this sense, any monotonic welfare function will operate as a Pareto efficient social choice function.

Info 10.1: Welfare Functions Avoid Arrows Impossibility Theorem. When a welfare function W is monotonic, it can be used as a complete, transitive, IIA, and Pareto efficient preference aggregation rule and a nonempty, IIA, and Pareto efficient social choice function!

As a note, **Utilitarian Welfare**, **Rawlsian Welfare**, and **Nash Welfare** are all monotonic.

10.8 Comparison of Welfare Functions

Notice that Rawlsian welfare picks (y, y) with utility $(12, 12)$ as the optimal outcome, while utilitarian and Nash welfare choose either (y, n) or (n, y) with utilities $(10, 25)$ and $(25, 10)$. Why does this happen?

In rough terms, our three welfare functions put different priorities on “efficiency” (in the sense of total utility) and inequality. Utilitarian welfare completely prioritizes efficiency and ignores inequality. For example, the unweighted utilitarian welfare function ranks $(10, 0)$ and $(5, 5)$ equally. On the other hand, Rawlsian welfare prioritizes equality. For example, it ranks $(5, 5)$ and $(10, 5)$ equally. The Nash welfare function falls somewhere in between these, striking a balance between efficiency and equality.

In **Example 10.1**, $(12, 12)$ is perfect equal despite having lower efficiency than $(10, 25)$ and $(25, 10)$.

10.9 Welfare Contours

Suppose we have a utility function $U(x, y)$ that represents preferences over sets of bundles (x, y) . Any equation $U(x, y) = c$ defines an indifference curve, that is a set of bundles (x, y) that all give constant utility c . For instance, if utility is perfect substitutes: $U(x, y) = x + y$ then the set of bundles (x, y) that all give utility of 10 is given by the equation $x + y = 10$. This is the line $y = 10 - x$.

We can do the same with welfare functions. A **welfare contour** is a set of combinations of utilities (u_1, u_2) that all give the same welfare. Let’s look at the contours of our welfare functions. Below, I have plotted the welfare contours through each of the outcomes in our running example of cleaning the work kitchen.

The utilitarian welfare contours are straight lines. This represents the fact that the utilitarian welfare function only cares about total utility, but not how that utility is distributed. If we take away one point/dollar of utility from person 2 and give it to person 1, the welfare remains constant.

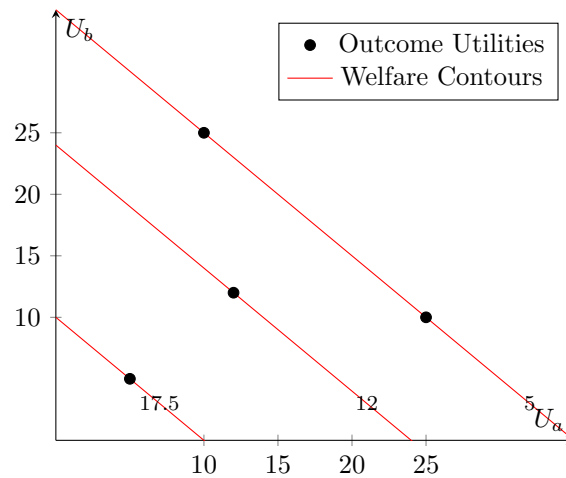
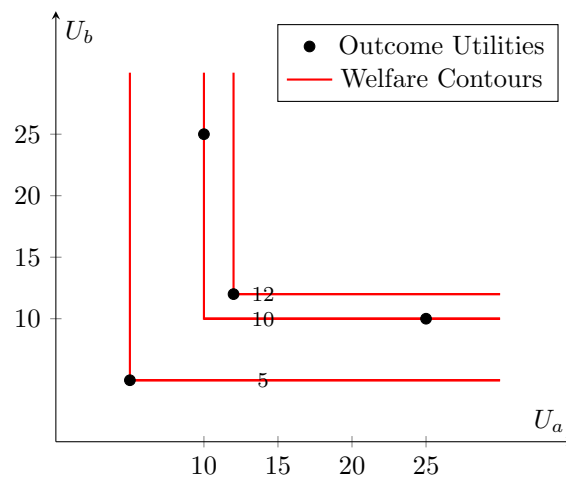


Figure 10.4: Utilitarian Welfare Contours for Example 10.1

The Rawlsian welfare contours are L-shaped. This represents the fact that starting at a point where both people have the same utility, the welfare remains the same if you only increase the utility of one person.



The Nash welfare contours bend away from the origin. They are steep when person 2 has a lot of utility relative to person 1. This is because to keep the same welfare, the function is willing to take away a lot of utility from person 2 to give a little to person one. The opposite happens when person 1 has a lot of utility relative to person 2.

Figure 10.5: Rawlsian Welfare Contours for Example 10.1

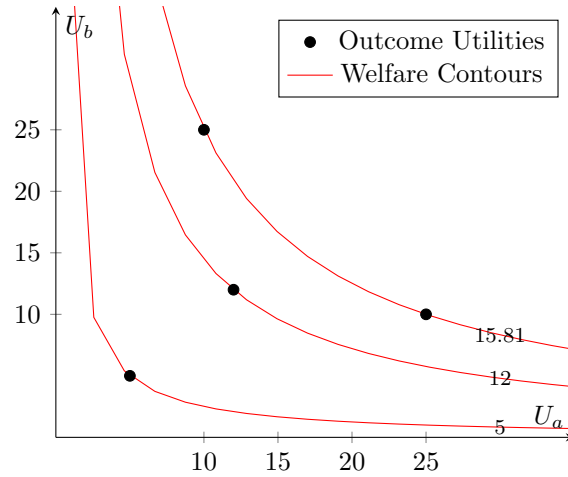


Figure 10.6: Nash Welfare Contours for **Example 10.1**

10.10 Key Topics

- Plot the utility pairs of outcomes in a cardinal model on a Cartesian plane as in *Exercises 10.1 and 10.6*.
- Calculate the Utilitarian, Rawlsian, and Nash welfare of outcomes and determine which outcome maximizes the relevant welfare functions as in *Exercises 10.2, 10.3, 10.5, 10.7, 10.9, 10.10*.

11 Randomization and Side-Payments

11.1 Randomization

Suppose that Fiona (Alice and Bob's boss) comes up with a brilliant compromise. Alice and Bob will flip a coin. If the coin is heads, Alice cleans and if it is tails, Bob cleans. We call this a **Randomized Choice**. The coin-flip is just one scheme that Fiona could have come up with involving **randomization**. Let's define this concept formally:

Definition 11.1: Randomized Choice. A randomized choice s is a probability distribution over the set of outcomes. Formally, $s \in \Delta O$

Example 11.1: Coin-Flip Scheme. Recall that for [Example 10.1](#), $O = \{ab, a, b, n\}$. The coin-flip scheme is the following distribution over $O = \{ab, a, b, n\}$:

$$\begin{aligned}s(ab) &= 0 \\ s(a) &= 0.5 \\ s(b) &= 0.5 \\ s(n) &= 0\end{aligned}$$

We can calculate the expected utility for both Alice and Bob of the coin-flip scheme.

The possible outcomes for Alice are:

- Heads: Alice cleans. The utility for Alice is $U_a(a) = 10$.
- Tails: Bob cleans. The utility for Alice is $U_a(b) = 25$.

The expected utility for Alice is:

$$\begin{aligned}E[U_a] &= 0.5 \times U_a(a) + 0.5 \times U_a(b) \\ &= 0.5 \times 10 + 0.5 \times 25 \\ &= 17.5\end{aligned}$$

Similarly, the expected utility for Bob is given by:

$$\begin{aligned}E[U_b] &= 0.5 \times U_b(a) + 0.5 \times U_b(b) \\ &= 0.5 \times 25 + 0.5 \times 10 \\ &= 17.5\end{aligned}$$

Notice that this pair Pareto dominates the (12, 12) pair when they both clean! This is demonstrated in [Figure 11.1](#)

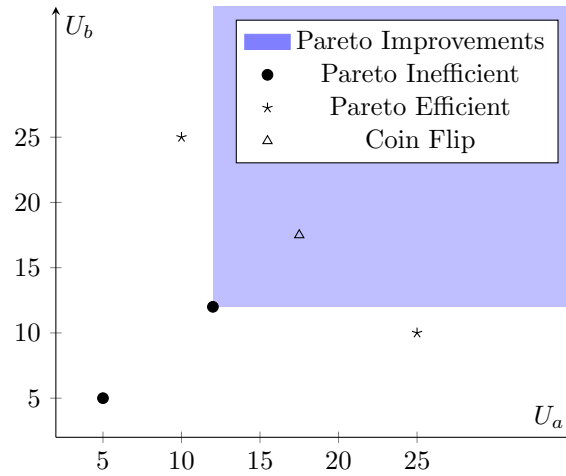


Figure 11.1: (12, 12) is Pareto dominated by Coin-flip Scheme

11.2 Convex Combinations

Notice how the point representing the expected utility of the coin-flip scheme is half-way between the utilities of the two outcomes in the scheme. In fact, if Fiona had chosen another probability, say $\frac{1}{4}$ chance that Alice cleans and $\frac{3}{4}$ chance that Bob cleans, the expected utility pair would be $\frac{1}{4}$ of the way between the points (10, 25) and (25, 10) on a straight line. In fact, the line between these two points gives all of the expected utility pairs that can be achieved by randomizing between these outcomes. This is known as the **convex combination**.

Definition 11.2: Convex Combination: Two Points. A *convex combination* of two points x and y is a point z of the form:

$$z = tx + (1 - t)y$$

with $0 \leq t \leq 1$.

Informally, the set of points z that meets this condition creates a straight line between the two points.

Notice that Fiona could have also randomized between other outcomes. In fact, she could have randomized between more than two outcomes. Consider the following randomized choice:

$$\begin{aligned} s(ab) &= 0.25 \\ s(a) &= 0.5 \\ s(b) &= 0.1 \\ s(n) &= 0.15 \end{aligned}$$

The expected utilities of this are

$$E(U_a) = 0.25(12) + 0.5(10) + 0.1(25) + 0.15(5) = 11.25$$

$$E(U_b) = 0.25(12) + 0.5(25) + 0.1(10) + 0.15(5) = 17.25$$

This is the point (11.25, 17.25) shown on the plot below.

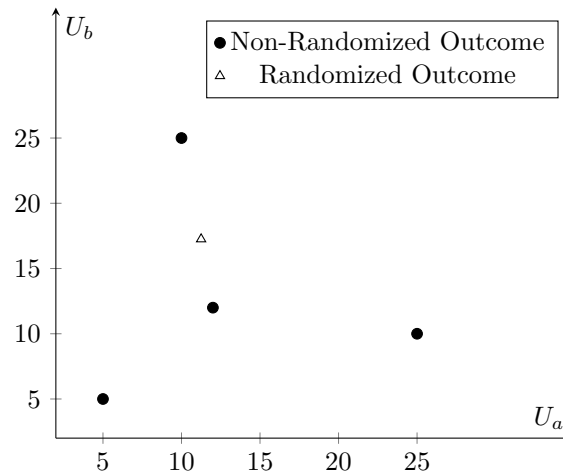


Figure 11.2: (11.25, 17.25) Outcome Together with Non-randomized Outcomes.

This is still called a convex combination. Here, however, it is the convex combination of several points.

Definition 11.3: Convex Combinations. A *convex combination* of multiple points x_1, x_2, \dots, x_n is a point z of the form:

$$z = \sum_{i=1}^n t_i x_i$$

with $t_i \geq 0$ and $\sum_{i=1}^n t_i = 1$.

11.3 Convex Sets

A convex set is a set of points that contains all of its convex combinations. While convex combinations can technically be combinations of multiple points, if a set contains all its convex combinations of two points, it will also contain all of its convex combinations. Because of this, we can check if a set is convex by checking whether, when we draw a straight line through any two points, that line remains in the set.

Definition 11.4: Convex Set. A set S is called a convex set if, for any two points x and y in S , every convex combination of x and y is entirely contained within S .

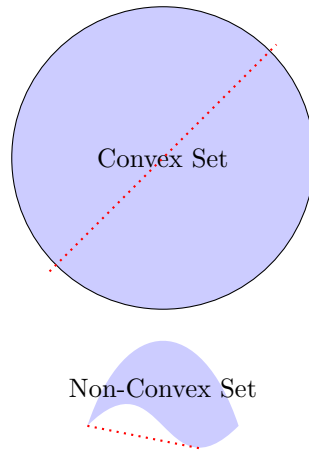
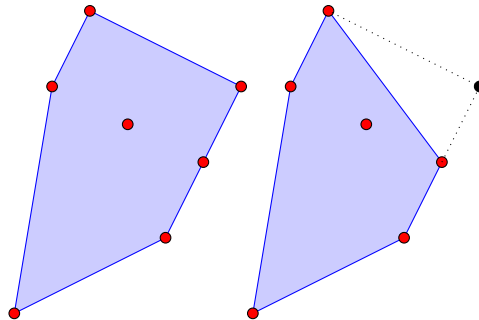


Figure 11.3: Convex and Non-Convex Sets

11.4 Convex Hulls and Achievable Utilities

Definition 11.5: Convex Hull. The *convex hull* of a set of points X is the set of all convex combinations of points in X . It is also the smallest convex set that contains X .

The convex hull of a set of points can be intuitively understood by imagining a rubber band stretched around the outermost points of the set. When the rubber band is released, it will contract and fit snugly around these points, forming the smallest convex shape that contains all the points. This shape is the convex hull. Two examples of convex hulls are shown in Figure 11.4



Convex hull of red points before and after black point is removed.

Figure 11.4: Convex Hulls

We might wonder what the set of possible utility pairs is under *some* correlated choice scheme. Quite simply, it is the convex hull of the utilities of the choices themselves.

Info 11.1: Utilities of Randomized Outcomes. The set of expected utilities achievable under some randomized choice is the convex hull of the utilities of the feasible choices. If we let U be the set of utilities $(u_1(c), u_2(c), \dots, u_n(c))$ for some $c \in C$ then this set of achievable expected utilities is $\text{Conv}(U)$.

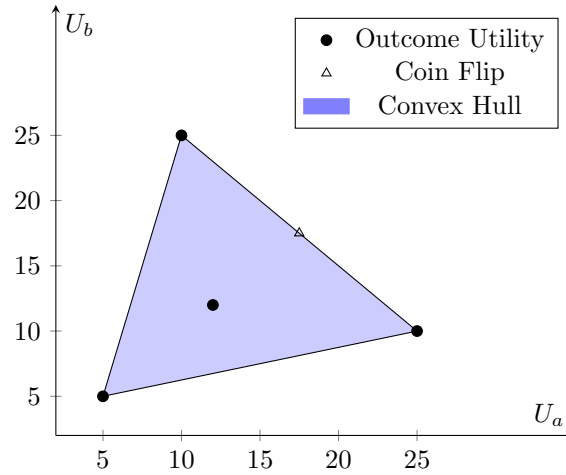


Figure 11.5: Achievable Utility Combinations from **Example 10.1**

11.5 The Pareto Frontier

It is also useful to formalize the set of all achievable Pareto efficient points that can be achieved with correlated choice schemes. This is known as the **Pareto frontier**.

Definition 11.6: Pareto Frontier. The Pareto frontier of some set of points U is the set of all Pareto efficient points in $\text{Conv}(U)$.

Let's look at the Pareto frontier of the utility combinations that are feasible in example **Example 10.1**. Here we will see it is the line that joins the points $(10, 25)$ and $(25, 10)$.

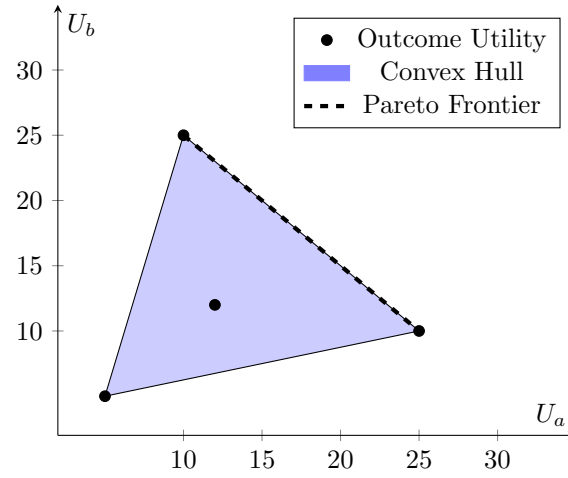


Figure 11.6: Pareto Frontier **Example 10.1**

Info 11.2: Welfare Max is on Pareto Frontier. When a welfare function W is monotonic (see **Definition A.1**), the correlated choice scheme that maximizes W will always be on the Pareto frontier.

11.6 Maximizing Nash Welfare

Let's return to this example:

Example 11.2: Cleaning the Kitchen: Utility Version.

$$P = \{a, b\}$$

$$O = \{\text{both, Alice, Bob, Neither}\}$$

$$u_a(o) = \begin{cases} 12 & \text{if } o = \text{both} \\ 10 & \text{if } o = \text{Alice} \\ 25 & \text{if } o = \text{Bob} \\ 5 & \text{if } o = \text{neither} \end{cases}$$

$$u_b(o) = \begin{cases} 12 & \text{if } o = \text{both} \\ 25 & \text{if } o = \text{Alice} \\ 10 & \text{if } o = \text{Bob} \\ 5 & \text{if } o = \text{neither} \end{cases}$$

Using what we have learned so far, let's look at the possible expected utilities that can be achieved by some correlated choice scheme (see **Definition 11.1**).

First, using the result in **Info Box 11.2**, we know the optimal pair must be on the Pareto frontier. As we can see in **Figure 11.6**, this is the line that joins (10, 25) and (25, 10). In fact it is a convex

combination. The pairs of utilities that can be achieved are $(t10 + (1 - t)25, (t(25) + (1 - t)10))$. Let's calculate the welfare of one of these points for a generic t :

$$(t10 + (1 - t)25)^{\frac{1}{2}}(t(25) + (1 - t)10)^{\frac{1}{2}}$$

Here's a little maximization trick. If we take a strictly increasing transformation of a function, the resulting function has the same maximum as the original. What if we take the log of the function above?

$$(t10 + (1 - t)25)(t(25) + (1 - t)10)$$

This simplifies to:

$$-225t^2 + 225t + 250$$

This is maximized where it's derivative is zero: $450t = 225$

$$t = \frac{1}{2}$$

Notice how the coin-flip scheme is on a higher welfare contour than either of the non-randomized outcomes.

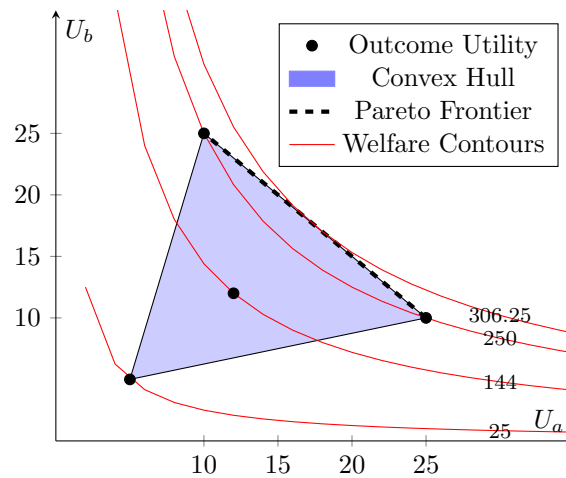


Figure 11.7: Coin-Flip is on the Highest Welfare Contour

Let's look at another example:

Example 11.3: Microwaving Fish with Different Utilities.

$$P = \{a, b\}$$

$$O = \{yes/stay, yes/leave, no/stay, no/leave\}$$

$$U_a(o) = \begin{cases} 30 & \text{if } o = (yes/stay) \\ 30 & \text{if } o = (yes/leave) \\ 25 & \text{if } o = (no/stay) \\ 25 & \text{if } o = (no/leave) \end{cases}$$

$$U_b(o) = \begin{cases} 0 & \text{if } o = (yes/stay) \\ 10 & \text{if } o = (yes/leave) \\ 20 & \text{if } o = (no/stay) \\ 10 & \text{if } o = (no/leave) \end{cases}$$

First we need to plot this to find the Pareto frontier.

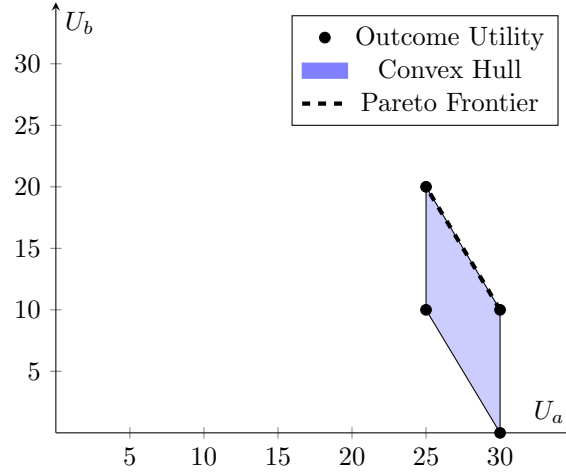


Figure 11.8: Pareto Frontier

The Pareto frontier is the set of points on the line between $(30, 0)$ and $(25, 20)$. Taking t to be the probability of choosing $(30, 0)$, these points can be written as $(t(30) + (1 - t)25, t(0) + (1 - t)20)$ for any t from 0 to 1. The Nash welfare of any of these points is:

$$(t(30) + (1 - t)25)^{\frac{1}{2}}(t(0) + (1 - t)20)^{\frac{1}{2}}$$

This is maximized at the same t as:

$$(t(30) + (1 - t)25)(t(0) + (1 - t)20)$$

This simplifies to:

$$-100t^2 - 400t + 500$$

Setting the derivative of this with respect to t equal to zero and solving, we get $t = -2$. But wait, that's not between 0 and 1! Nash welfare wants us to pick as low a t as possible, but the low is 0 not -2 . So the optimal point is $t = 0$ which is simply the outcome that gives utility (25, 20) or the outcome no/stay (not randomized).

11.7 Optimizing Rawlsian Welfare

Rawlsian welfare is a little easier to optimize. The trick is to look for the randomized outcome that is as close as possible to the 45-degree line where both expected utilities are equal.

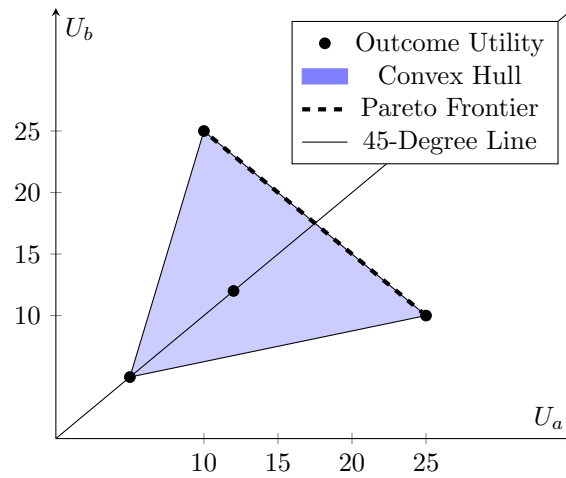


Figure 11.9: Coin-Flip is the Randomized Outcome Maximizing Rawlsian Welfare

The 45-degree line is where the utility of both people is equal. If there is an outcome on the Pareto frontier where both utilities are equal, the outcome maximizes Rawlsian welfare. Here, we can see there is one. When that is the case, we can find the t that makes both people's utility the same on the Pareto frontier, by taking the formula for points on the Pareto frontier (we found this above):

$$(t10 + (1 - t)25, (t(25) + (1 - t)10))$$

Now, set the two utilities equal and solve for t . Again, the outcome that maximizes Rawlsian welfare is the coin-flip scheme.

$$t10 + (1 - t)25 = (t(25) + (1 - t)10)$$

$$t = \frac{1}{2}$$

What about our modified fish-cooking example?

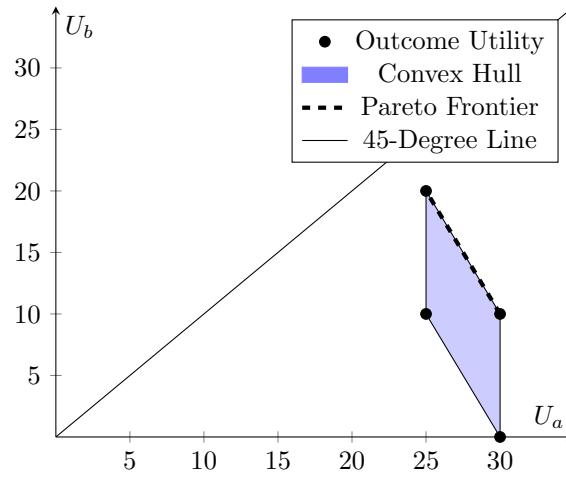


Figure 11.10: Pareto Frontier

Here, there is no outcome on the Pareto frontier that is on the 45-degree line, but $(25, 20)$ is the closest we can get. This is the pair of utilities that maximizes Rawlsian welfare. We don't need to randomize here.

11.8 Optimizing Utilitarian Welfare

Utilitarian welfare does not require much visualization. We are just looking for the point on the Pareto frontier that has the highest total utility. We can take the equation for points on the Pareto frontier and find the average utility. For the cleaning example, the point on the frontier are:

$$(t10 + (1 - t)25, (t(25) + (1 - t)10))$$

Taking the average of these:

$$\frac{t10 + (1 - t)25 + (t(25) + (1 - t)10)}{2}$$

This simplifies to: $\frac{35}{2} = 17.5$ which tells us that all the points on the frontier have the same utilitarian welfare. We do not need to go any further. Any of these points maximizes utilitarian welfare.

What about the fish example? The points on the Pareto frontier are:

$$(t(30) + (1 - t)25, t(0) + (1 - t)20)$$

Taking the average:

$$\frac{t(30) + (1-t)25 + t(0) + (1-t)20}{2}$$

$$\frac{45 - 15t}{2}$$

This is decreasing in t . Thus, utilitarian welfare is maximized when t is as low as possible, or $t = 0$. This is the point $(25, 20)$.

Notice that in both cases, the optimal points included at least one of the endpoints. **This will always be the case.** Because of that you can simply calculate the utilitarian welfare of the endpoints of the Pareto frontier. If one is higher, that is the utilitarian welfare maximizing outcome. If they are the same, any randomization between them has the same utilitarian welfare and any of them maximize utilitarian welfare.

11.9 Ex-ante / Ex-Post

In [Example 10.1](#) we have seen how a higher expected utility can be achieved for both Alice and Bob by using the coin-flip scheme ([Example 11.1](#)). Each person's expected utility is 17.5. Without randomization, the most fair outcome is (y, y) where they both get 12.

The coin-flip scheme seems like a clear winner. In fact, we have seen that from an expected utility standpoint $(17.5, 17.5)$ Pareto dominates $(12, 12)$. But this hides some subtlety. $(17.5, 17.5)$ cannot be an outcome. Once the coin is flipped, someone has to do the work. The actual utility outcomes will either be $(10, 25)$ or $(25, 10)$.

Should we use the *ex-ante* utility $(17.5, 17.5)$ or the *ex-post* utilities $(10, 25)$ or $(25, 10)$? It can make a big difference.

Suppose we use a **Rawlsian Welfare** $W(U_a, U_b) = \min\{U_a, U_b\}$ to assess the outcomes. From an ex-ante perspective, the coin-flip scheme with Rawlsian welfare 17.5 clearly dominates (y, y) with Rawlsian welfare 12. However, from an ex-post perspective, the welfare of the actual outcomes of the coin-flip scheme have Rawlsian welfare 10 while (y, y) has Rawlsian welfare 12. Thus, if we take an ex-post approach, (y, y) dominates.

The debate of whether to use ex-ante and ex-post welfare measures to evaluate outcomes or policies is very important in policy-making. In fact, this borders on the ongoing debates about fairness in society regarding whether individuals should have fairness of opportunity or fairness of outcomes.

When it comes to choosing whether to prioritize fairness or efficiency as discussed in [subsection 10.8](#), welfare functions are able to balance how much priority is put on each. For instance, **Rawlsian Welfare** prioritizes equality, **Utilitarian Welfare** prioritizes efficiency, and **Nash Welfare** balances the two. Recently, [\[5\]](#) and related work has shown how ex-post and ex-ante welfare analysis can be incorporated into a single welfare function. In this sense, the debate about ex-ante and ex-post welfare does not need to be about whether to use one or the other, but rather how much to prioritize each.

11.10 Side-Payments

In [Example 10.1](#), having both Alice and Bob clean is very fair. They both get utility 12. But it is more efficient if only one person cleans since the utilities are $(10, 25)$ or $(25, 10)$

Suppose that Fiona (Alice and Bob's boss) comes up with a brilliant compromise. Instead of both cleaning and getting utility 12, Bob will clean and Alice will pay him \$7.50. Recall that utility is measured in dollars. Thus, the result of this is the outcome $(10 + 7.5, 25 - 7.5) = (17.5, 17.5)$

This demonstrates that with *side payments* or the so-called *transferable utility*, new outcomes can be unlocked that are "between" the outcomes associated with each combination of choices.

11.11 Feasible Utility Combinations using Side Payments

With side payments, the set of feasible utility outcomes is quite large. Any pair of utilities can be implemented as long as the total amount of utility is less than or equal to the total amount of utility in the outcome that has the highest utilitarian welfare.

We can write this set in the following way:

Info 11.3: Feasible Utilities with Side Payments. The set of feasible utility combinations with side payments is the set of all utilities combinations that can be achieved by transfers from any of the outcomes that maximize utilitarian welfare. This set can be represented by all of the utility pairs below the triangle created by drawing a line with slope of -1 through the utilitarian welfare maximizing outcome.

11.12 Optimizing Welfare with Side Payments

When side payments are possible. It is possible to both maximize efficiency and fairness at the same time. Here is a procedure.

First, find the outcome(s) that maximizes utilitarian welfare. Remember that utilitarian welfare measures the average welfare in an outcome. The point that maximizes average welfare also maximizes total welfare. Let's suppose the highest achieve utilitarian welfare is some number w . It is possible to find an outcome and a set of side-payments that ensures everyone gets w in utility. This point has the same utilitarian welfare as the best possible outcome, but it also gives everyone the same utility, thus it also simultaneously maximizes Rawlsian and Nash welfare among all the feasible utility combinations with side-payments.

To achieve this, pick one of the utilitarian welfare maximizing outcomes, calculate the utilitarian welfare w in that outcome. The payments needed to give everyone utility w is simply the difference between w and what they get in the utilitarian welfare maximizing outcome. Let's look at an example:

In the cleaning example, either a or b give a utilitarian welfare of 17.5. The set of utilities achievable with side-payments for this model is shown below.

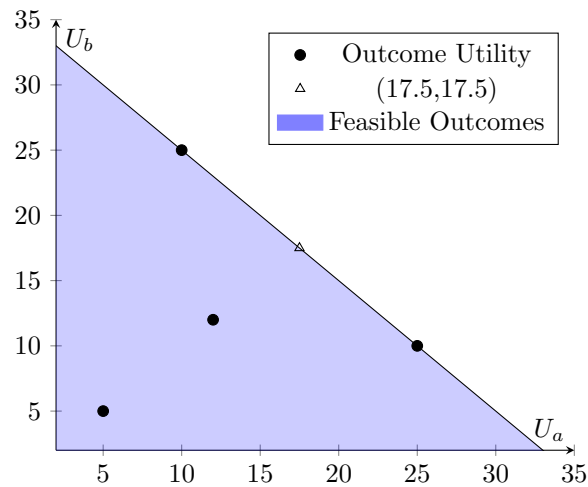


Figure 11.11: Achievable Utility Combinations from Example 10.1

We can achieve $(17.5, 17.5)$ by either choosing outcome a or outcome b and then using side payments. If we choose a the utilities are $(10, 25)$. We need Alice to get 17.5 and bob to get 17.5. Thus, we need Alice to get 7.50 extra and bob to get 7.50 less. So, we just have bob give Alice 7.50. Notice this is also shown by taking the difference between the two pair of utilities $(17.5, 17.5) - (10, 25) = (7.5, -7.5)$. Alice gets 7.5 Bob gives up 7.5.

Let's try this for the our other running example, the modified microwaving fish model. First, we need to find the outcome with highest utilitarian welfare. This is n/s since it has the highest average utility. A line with slope of -1 through this point represents the possible points we can achieve by using side-payments from the outcome. Since the total amount of utility is 45 in this outcome, it is a line connecting $(0, 45)$ and $(45, 0)$.

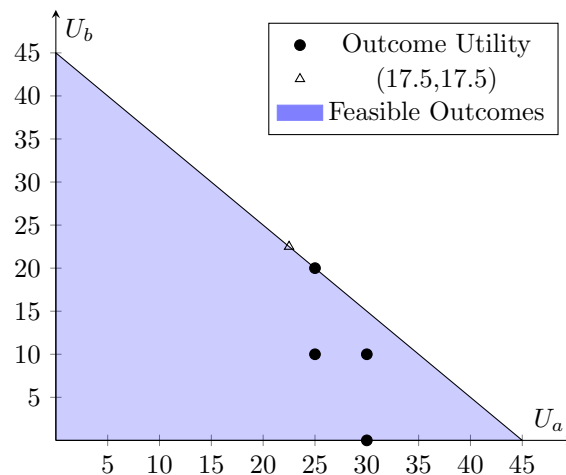


Figure 11.12: Pareto Frontier

The point simultaneously maximizing Nash, Rawlsian, and Utilitarian welfare is the point where the utilities are equal on this line. That is the point (22.5, 22.5). To achieve this point, we pick the outcome n/s which gives utility 25, 20 then have person 1 give person 2 a payment of 2.25.

11.13 Key Topics

- Plot the set of utility pairs achievable with randomization by finding the convex hull of the points achievable in the non-randomized outcomes as in *Exercise 11.1*.
- Understand what the Pareto frontier is and be able to find it as in *Exercise 11.1*.
- Be able to find the point on the Pareto frontier that maximizes Utilitarian, Rawlsian, and Nash welfare as in *Exercises 11.3, 11.4, 11.5, 11.8, 11.9, 11.10, 11.12, 11.13, 11.14*.
- Be able to find the set of points achievable with side-payments as in *Exercise 11.6*.
- Be able to find the points achievable with side-payments that simultaneously maximizes Utilitarian, Nash, and Rawlsian welfare as in *Exercises 11.7, 11.11, 11.15*.

Part IV

Public Goods

12 Introduction to Game Theory

12.1 Games

In every model so far, we have treated the outcomes as if they are chosen by the administrator. Now we want to look at what happens if the choice is left to individuals.

We will modify our models so that outcomes are determined by the choices (we will call them strategies) of the individuals in the model.

Remember that a model is defined this way:

Definition 12.1: Cardinal Model. A *cardinal public outcome model* is: O : the set of outcomes. P : the set of people. And for every person $i \in P$: $u_i()$ their cardinal utility function over the set O .

A **game** is very similar, except instead of outcomes, the people have choices (called **strategies** that lead to outcomes.

Definition 12.2: Game.

A game is: P : the set of people. S_i the set of strategies for for each person. And for every person $i \in P$: u_i the cardinal utility for every combination of choices s_1, s_2, \dots, s_n .

For the rest of this subsection, I will describe a game by modifying [Example 3.6](#) to be a game.

As before $P = \{\text{Alice}, \text{Bob}\}$, though, it's probably easier to shorten this to $\{a, b\}$.

In games, we can also use numbers to reference people. If there were ten people in some model, we might use $N = \{1, \dots, 10\}$. If we wanted to create a model that works for some unspecified number of people, we could use $N = \{1, \dots, n\}$, letting n be a variable.

We now move on to the strategies: S_i . The possible strategies need to enumerate the *options available to each person*. For a particular person i , we will denote their possible choices as S_i . Then, S is all of the combinations of choices that can be made, one for each agent. Formally, S is the Cartesian product of each S_i . This is written as:

$$S = \bigtimes_{i \in N} S_i \tag{1}$$

In [Example 3.4](#), we need a set of choices for Alice and a set for Bob. We will denote Alice's possible choices as S_a and Bob's possible choices as S_b . Alice can choose to microwave some fish and Bob can choose to stay in the office or leave. Then we can write: $S = S_a \times S_b$.

We can use some shorthand here to denote the possible choices.

- y : "clean"

- n : “don’t clean”

... and we can write: $S_a = \{y, n\}$ and $S_b = \{y, n\}$. The possible elements of S are:

- (y, y) : both clean
- (y, n) : Alice cleans
- (n, y) : Bob cleans
- (n, n) : neither cleans

We can refer to a “generic” element of an agents choice set as s_i such that $s_i \in S_i$. Similarly, we can refer to a particular combination of choices as s . If we have $N = 1, 2, 3$ (three players), then s looks like $s = (s_1, s_2, s_3)$.¹

Finally, we come to preferences U . This is a utility function that maps the set of possible strategies into a number for each person. Because utility of one agent can depend on the strategies of another, **we must define utility over the possible combinations of strategies**. We also need to define utility for each agent.

In **Example 3.4**, we have U_a and U_b for Alice and Bob’s utility functions, respectively. Each of these needs to provide an amount of utility for the four possible scenarios in S (outlined above).

Here is a possible for these utility functions that captures what we want in this model:

$$U_a(s) = \begin{cases} 12 & \text{if } s = (y, y) \\ 10 & \text{if } s = (y, n) \\ 25 & \text{if } s = (n, y) \\ 5 & \text{if } s = (n, n) \end{cases}$$

$$U_b(s) = \begin{cases} 12 & \text{if } s = (y, y) \\ 25 & \text{if } s = (y, n) \\ 10 & \text{if } s = (n, y) \\ 5 & \text{if } s = (n, n) \end{cases}$$

It is often useful to display the utility combinations in a table form. In fact, this method is used extensively in game theory. Here is a 2x2 table of the possibilities in this game. The row indicates Alice’s choice s_a and the column indicates Bob’s choice s_b . In fact, this table gives us everything we need to know about the game in one place. A table like this puts all of the information about a game into one convenient place.

¹Sometimes we want to have an agent that doesn’t make any decisions, but is impacted by the decisions of others. We can do that by making the choice set “null” (denoted with \emptyset , the empty set).

Example 12.1: Kitchen Cleaning Game.

$s_a \backslash s_b$	y	n
y	(12,12)	(10,25)
n	(25,10)	(5,5)

Let's make a game out of the microwaving fish. Here are the people, strategies, and preferences:

$$\begin{aligned}
 P &= \{a, b\} \\
 S_a &= \{y, n\} \\
 S_b &= \{s, l\} \\
 U_a(s) &= \begin{cases} 25 & \text{if } c = (y, s) \\ 25 & \text{if } c = (y, l) \\ 20 & \text{if } c = (n, s) \\ 20 & \text{if } c = (n, l) \end{cases} \\
 U_b(s) &= \begin{cases} 5 & \text{if } c = (y, s) \\ 10 & \text{if } c = (y, l) \\ 20 & \text{if } c = (n, s) \\ 10 & \text{if } c = (n, l) \end{cases}
 \end{aligned}$$

Let's put this in table form:

Example 12.2: Microwaving Fish, Game.

$s_a \backslash s_b$	s	l
y	(25,5)	(25,10)
n	(20,20)	(20,10)

12.2 Best Response

How should someone play a game? We assume they will pick the strategy that is best for them. In the microwaving fish game, what should Bob do if Alice is going to microwave fish? Since the utility of leaving is bigger than staying conditional on Alice microwaving fish, $u_b(y, l) > u_b(y, s)$ it is in Bob's best interest to choose strategy l in this case. We say that l is a **best response** to y . Similarly s is a best response to n .

To formally define a best response, we need some notation. Just like s is a combination of strategies for all players and $s \in S$, and s_i is a strategy for player i and $s_i \in S_i$, we can define s_{-i} to be a combination of choices for all players who are not i . Similarly, S_{-i} is the set of all such combinations of choices for players who are not i such that $s_{-i} \in S_{-i}$.

Definition 12.3: A. Best Response for person i , s_i is a **Best Response** to s_{-i} if $U_i(s_i, s_{-i}) \geq U_i(s'_i, s_{-i})$ for every other strategy $s'_i \in S_i$.

This says that s_i is a Best response for person i to some combination of strategies from the others s_{-i} if choosing s_i is at least as good for person i as anything else they could choose. We collect all of a player's best responses into a function called the best response function. Think of it as a guide for a player on what to choose.

Definition 12.4: A. Best Response Function for person i is a function that gives the **best response** from S_i for person i for every combination of strategies of the other player(s) S_{-i} . It is written as follows: $B_i(s_{-i}) = s_i$. This says that, against the combination of strategies, s_{-i} , s_i is the best response for i .

For Bob, the best response function is $B_b(y) = l, B_b(n) = s$. What about for Alice. Her utility is not affected by what Bob does and she is always better off making the fish. So her best response function is $B_a(s) = y, B_a(l) = y$.

Let's look at the cleaning example. What is the best response for Alice if Bob will clean y ? She is better off not cleaning. We write: $B_a(y) = n$. Similarly if Bob will not clean n , she is better off cleaning: $B_a(n) = y$. For Bob, we have $B_b(y) = n$ and $B_b(n) = y$.

12.3 Nash Equilibrium

Equilibrium is the concept we use to "solve" a model. That is, make a prediction about the choices that will prevail in the model. The informal idea of equilibrium is that it is a combination of choices such that no person has incentive to change their strategy holding fixed the choices of others.

In other words, a combination of strategies s is a Nash equilibrium if each s_i in the combination is a best response to the strategies of the others s_{-i} .

Definition 12.5: Nash Equilibrium. The **equilibrium** or more formally **Nash equilibrium** of a game is a $s = (s_1, \dots, s_n) \in S$ such that for all $i \in P$ and s_i is a best response to s_{-i} .

Example 12.3: Cleaning Game, Equilibrium. Example 12.1 has two equilibria: (y, n) and (n, y) . Alice cleaning is a best response to Bob not cleaning and Bob not cleaning is a best response to Alice cleaning. Neither have incentive to change their strategy, holding fixed the strategy of the other players. Thus, (y, n) is an equilibrium. Similarly Alice not cleaning is a best response to Bob cleaning and Bob cleaning is a best response to Alice not cleaning. Thus (n, y) is also an equilibrium.

Here is another convenient way to find equilibrium when you can put the utilities in a grid

Using the table form of the utility combinations seen in Example 12.1, we can highlight the best responses. Here, I have put in bold the utility of a player if it is the highest utility achievable given the other's choice. For instance, 10 is in bold for Alice in the y, n pair, since 10 is the highest utility Alice can achieve when Bob chooses n . She does this by choosing y . Similarly 35 is in bold for Bob in the (y, n) strategy combination since 35 is the highest utility Bob can achieve when Alice chooses y . In this way, we can find equilibrium simply by looking for pairs where both numbers are in bold. Again, that occurs both at (y, n) and (n, y) .

$s_a \backslash s_b$	y	n
y	(12,12)	(10 , 35)
n	(35 , 10)	(5,5)

Let's look at the microwaving Fish example:

Example 12.4: Microwaving Fish Game, Equilibrium. Example 12.2 has one equilibria: (y, l) .

$s_a \backslash s_b$	s	l
y	(25 ,5)	(25 , 10)
n	(20, 20)	(20,10)

12.4 Key Topics

- Be able to find the Pareto efficient outcomes in a game as in *Exercises 12.2, 12.5, 12.8*
- Be able to find the best responses in a 2x2 game as in *Exercises 12.3, 12.6, 12.9*
- Be able to find the Nash equilibrium of a 2x2 game as in *Exercises 12.4, 12.7, 12.10*

13 Public Goods Homogeneous Preferences

13.1 Motivating 2x2 Game

In every office I have ever worked in, the coffee machine provides a hot drink that is coffee in only the most strict theoretical sense. Usually, the “coffee” it makes tastes like hot chemicals—

something that comes straight out of a car engine. I suspect it is because these things just never get maintained.

Let's introduce a new scenario for Bob and Alice. Suppose that Alice and Bob are the only ones who use the coffee machine. It needs a little cleaning to stay in good shape. It costs about \$15 worth of effort to maintain the machine. During a particular week, if both choose to maintain the machine, it is worth \$30 to them both (they can avoid going to the local Coffee shop and spending \$6 a day). If only one of them maintains the machine, it is worth \$20 to them both. If no one maintains the machine, it is worth \$10 to them both.

Thus, if both Alice and Bob maintain the machine (m), they each receive a utility of $30 - 15 = 15$. If Alice contributes but Bob does not, Alice receives a utility of $20 - 15 = 5$, while Bob receives a utility of $20 - 0 = 20$. Bob benefits from the coffee machine without contributing to its maintenance. Conversely, if Bob contributes and Alice does not, Alice receives a utility of 20, and Bob receives a utility of 5. If neither Alice nor Bob contribute (n), both receive a utility of 10.

Example 13.1: The Coffee Machine, Game.

$c_a \backslash c_b$	m	n
m	(15,15)	(5,20)
n	(20,5)	(10,10)

Notice that the Pareto efficient outcomes are $(m,m), (m,n), (n,m)$. That is, the outcome is Pareto efficient if at least one person maintains the coffee machine. The utilitarian welfare maximizing outcome is (m,m) . It also happens to Rawlsian welfare. Let's find the Nash equilibrium of this game:

$c_a \backslash c_b$	m	n
m	(15,15)	(5, 20)
n	(20 ,5)	(10 ,10)

Notice that n (not maintaining the coffee machine) is a best response to either n or m ! That is, regardless of what the other person is doing, it is better to shirk the responsibility and not maintain the coffee machine. The only Nash equilibrium is (n,n) . That is, when left to decide for themselves, Alice and Bob will end up choosing an outcome that is not even Pareto efficient!

The reason for this is that, the utility cost of maintenance in dollar terms is \$15, but maintaining the machine only adds about \$10 of personal value to it. However, since each person that maintains the machine adds \$20 of *total* value to it, it is efficient from a utilitarian perspective for both people to maintain the machine. The issue is, when someone decides whether or not to maintain the machine, they only take into account how it helps their utility, not how it helps the other person's utility.

The fact that the action of maintaining the machine helps the other person makes maintenance create what we call a **positive externality**. An externality is a situation in which a choice by one person affects the utility of others. Here, maintenance is a positive externality since maintenance increases the utility of the other person. The fact that the Nash equilibrium of this game is not even Pareto efficient is due to the fact that neither person considers the positive externality they create by maintaining the machine, they only focus on their individual utility. This results in under-maintenance of the machine relative to the utilitarian maximizing outcome.

13.2 Public Goods- Two People

Suppose now Alice and Bob must decide how much contribution to put into maintaining the coffee machine. The contribution of Alice is g_a and for Bob is g_b . Contribution is measured in dollars so that if a player puts in contribution 1, their utility is reduced by 1 point. The value of the coffee machine to each person is $10\sqrt{g_a + g_b}$. The game can be written like this: the players are $P = \{a, b\}$, the strategies are the contributions chosen by each player: $g_a \in [0, \infty]$ and $g_b \in [0, \infty]$ (that is they can put in as much contribution as they want. The utilities are $u_a(g_a, g_b) = 10\sqrt{g_a + g_b} - g_a$ and $u_b(g_a, g_b) = 10\sqrt{g_a + g_b} - g_b$.

This game is much more complex than the simple 2x2 games we have looked at so far. Here, each player has an infinite number of strategies. The utility function for Alice is shown below. Notice at $(g_a, g_b) = (0, 0)$, Alice's utility is 0. On the other hand, at $(0, 100)$ where Bob contributes a lot and Alice contributes nothing, Alice's utility is 100. On the other hand at $(100, 0)$ where Alice chooses a 100 contribution and Bob contributes nothing, Alice's utility is 0.

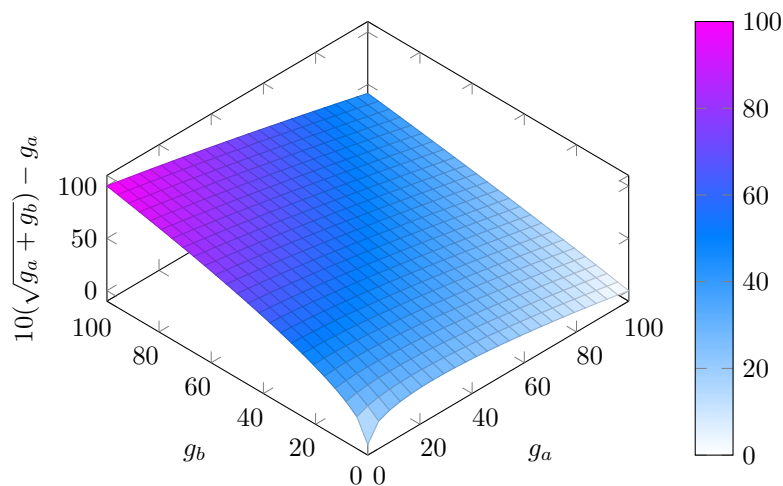


Figure 13.1: A plot of $U_a(g_a, g_b) = 10(\sqrt{g_a + g_b}) - g_a$

13.2.1 Equilibrium

What is the optimal strategy for Alice given Bob's choice of g_b . We can find this by solving for the maximum of Alice's utility with respect to g_a , leaving g_b as a variable.

This occurs where the derivative of her utility is zero with respect to g_a . That is, for every g_b , she needs to find a point where she cannot increase her utility by changing g_a . This can only occur where the slope of her utility function is zero with respect to g_a .

$$\frac{\partial (10\sqrt{g_a + g_b} - g_a)}{\partial g_a} = \frac{5}{\sqrt{g_a + g_b}} - 1$$

Setting this to zero:

$$\frac{5}{\sqrt{g_a + g_b}} - 1 = 0$$

This simplifies to:

$$g_a = 25 - g_b$$

This says, Alice best responds to Bob's choice of g_b by choosing whatever is needed to get to a total contribution of 25. Notice that $25 - g_a$ could be negative if $g_b > 25$. However, 0 is the smallest contribution. If it is the case that $g_b > 25$, while Alice would *like* to respond with a negative contribution, her best option is zero. Thus, her full best response function is:

$$g_a = \begin{cases} 25 - g_b, & \text{if } g_b \leq 25 \\ 0, & \text{if } g_b > 25 \end{cases} \quad (2)$$

Definition 13.1: Individually Ideal Total Contribution. The individually ideal total contributions for a player i is amount that

Bob best-responds to Alice in the same way:

$$g_b = \begin{cases} 25 - g_a, & \text{if } g_a \leq 25 \\ 0, & \text{if } g_a > 25 \end{cases} \quad (3)$$

Definition 13.2: Total Contributions g . Total contributions in a public goods game is the sum of the individual contributions. Recall that P is the set of people in the game.
 $g = \sum_{i \in P} g_i$.

Definition 13.3: Individually Ideal Total Contribution. Each player's best response has them contributing to bring the total contributions up to some g^* if the total contributions would not otherwise exceed that value even if they chose $g_i = 0$. We refer to g^* as the individually ideal total contributions.

In this game, the individual ideal total contributions are both 25. The individual ideal total contributions will always be the same for any players that share the same utility function. The chosen contributions of the players in this game are an equilibrium as long as their contributions sum to the individual ideal level of 25. For example, $g_a = 12.5$ with $g_b = 12.5$ or $(12.5, 12.5)$ is an equilibrium since 12.5 is a best response to 12.5: $12.5 = 25 - 12.5$. Similarly $g_a = 25$ with $g_b = 0$ or $(25, 0)$ is an equilibrium since 25 for Alice is a best response to 0 from Bob and 0 from Bob is a best response to 25 from Alice.

Since 25 is the total contribution in any equilibrium, we say that the equilibrium total contributions in this game is $g = 25$.

Definition 13.4: Equilibrium Total Contribution. The **equilibrium total contributions** in a public goods model is the total contributions g that occurs in any Nash equilibrium.

13.2.2 Utilitarian Max

Every equilibrium has utilitarian welfare (average utility) of 37.5. Of these, (12.5, 12.5) maximizes Rawlsian welfare. Let's use it for comparison. In this equilibrium, both players get a utility of 37.5.

However, notice that if, instead Alice and Bob both chose 50, their utilities would be $10\sqrt{50+50} - 50 = 50$. They can do better than equilibrium. Here, by increasing their total contributions above the equilibrium level. In this case, their total contributions are 100. In fact, this is the total contribution level that maximizes their utilitarian welfare. To see this, let's write the utilitarian welfare of any pair of strategies:

$$\frac{10\sqrt{g_a + g_b} - g_a + 10\sqrt{g_a + g_b} - g_b}{2}$$

$$\frac{20\sqrt{g_a + g_b} - (g_a + g_b)}{2}$$

Let's define g to be the total contribution and rewrite this as follows:

$$\frac{20\sqrt{g} - (g)}{2}$$

Now we can find the total contributions that maximize utilitarian welfare. The derivative of this is $\frac{10}{\sqrt{g}} - 1$. Solving for the g that makes this zero gives us $g = 100$. Thus, utilitarian welfare is maximized when $g = g_a + g_b = 100$. Thus, (50, 50) is the best the players can do. It is a fair outcome that also maximizes utilitarian welfare.

Definition 13.5: Utilitarian Ideal Total Contribution. The **utilitarian ideal total contributions** in a public goods model is the total contributions g that maximizes utilitarian welfare.

As in the coffee machine example at the beginning of this chapter, the equilibrium is strictly Pareto dominated! This is a hallmark of the public goods game. In equilibrium, players contribute too little relative to what would make them all most well-off.

The problem with utilitarian maximizing outcomes is that they are not equilibrium. Suppose Alice were to choose 50. Bob's best response would be 0! In fact, if we plug 50 into Bob's best response function above we get $g_b = 25 - 50 = -25$. Really, Bob would like to have a negative contribution (-25), but this is nonsense. Since the lowest his contribution can be is 0, that is his best response.

So ultimately, while both players are better off when they choose (50, 50), neither has incentive to actually stick with that plan. Either would have incentive to free-ride and choose zero contribution.

13.2.3 Tax

One solution to getting the outcome to be the utilitarian max is to have a **tax**. Here, the tax is an amount t everyone has to pay and that amount is contributed to the public good. If we Bob and Alice's utility with tax t , we get:

$$\begin{aligned}u_a(t, t) &= 10\sqrt{t+t} - t \\u_b(t, t) &= 10\sqrt{t+t} - t\end{aligned}$$

Thus, imposing a tax t is the same as forcing each person to contribute t . In this sense, the utilitarian maximum can be achieved by imposing a tax on everyone that adds up to the utilitarian ideal total contributions of 100. Here, this can be done by taxing both 50. Let's refer to this as the **utilitarian ideal tax**.

Definition 13.6: Utilitarian Ideal Tax. In a public goods model, the **utilitarian ideal tax** is the utilitarian welfare maximizing total contributions divided by the number of people. Let g^* be the utilitarian welfare maximizing total contributions and let n be the number of players in P . The utilitarian ideal tax is $t = \frac{g^*}{n}$.

In this model, each person also considers the ideal tax. If we calculate the t that maximizes either person's utility, we will find that $t = 50$ is their favorite tax. We refer to this as the **individually ideal tax**.

Definition 13.7: Individually Ideal Tax. In a public goods model, there **individually ideal tax** for person $i \in P$ is the t that maximizes their utility when all players contribute t .

Info 13.1: Individual and Utilitarian Ideal Taxes. When everyone has the same utility, the individual ideal tax and the utilitarian ideal tax coincide.

13.3 Public Goods- Many People

Let's try this same model with many people, not just two. There are n people $P = \{1, 2, \dots, n\}$. Let's define some notation. Let g_i be the contribution chosen by person i and let $g = \sum_{i=1}^n g_i$ (the total contribution by all n people). Now we define a new expression, the total contribution put in by all people except i is $g_{-i} = g - g_i$. We can use this to define the utility functions:

$$u_i(g_i, g_{-i}) = 10\sqrt{g_i + g_{-i}} - g_i$$

13.3.1 Equilibrium

We first need to find the best response. For player, i this occurs where their utility function has zero slope with respect to g_i .

We solve for where the derivative of the utility function with respect to g_i is zero:

$$\frac{5}{\sqrt{g_i + g_{-i}}} - 1 = 0$$

$$g_i = 25 - g_{-i}$$

In this game, we find that the individually ideal total contributions and the equilibrium total contributions are 25. As long as $g = \sum_{i=1}^n g_i = 25$, the strategies are an equilibrium.

13.3.2 Utilitarian Max

Let's take the same approach as above to find the strategies that maximize utilitarian welfare. Summing the n individuals utilities, we get:

$$\sum_{i=1}^n 10\sqrt{g_i + g_{-i}} - g_i$$

Recall that $g = \sum_{i=1}^n g_i$

$$\begin{aligned} & \sum_{i=1}^n 10\sqrt{g} - g_i \\ & \sum_{i=1}^n 10\sqrt{g} - \sum_{i=1}^n g_i \\ & 10n\sqrt{g} - g \end{aligned}$$

Let's maximize this with respect to g . The derivative with respect to g is:

$$\frac{5n}{\sqrt{g}} - 1$$

Solving for when this is zero, we get $g = 25n^2$. For $n = 2$, we get $g = 100$ as above. For $n = 100$, we get $g = 250000$! Recall that the equilibrium total contributions were 25 regardless of the number of people in the model. Since the utilitarian ideal total contributions are $25n^2$, the utilitarian ideal is n^2 times larger than the equilibrium contributions. That 10000 times when $n = 100$!

We can achieve this utilitarian maximum by using the utilitarian ideal tax of $\frac{25n^2}{n} = 25n$. For $n = 100$, we get $t = 2500$.

13.4 Key Topics

For questions from this chapter on an exam, I will only use the log utility function.

- Be able to find the best response function of a player in a public goods game with homogeneous preferences as in *Exercise 13.1*
- Be able to find the individually ideal total contributions for the players in a public goods game with homogeneous preferences as in *Exercise 13.2*
- Be able to find the equilibrium total contributions and understand that any individual contributions that sum to that equilibrium total constitutes an equilibrium of the game as in *Exercise 13.2*
- Be able to find the utilitarian ideal total contributions for a public goods model with homogeneous preferences as in *Exercises 13.3 and 13.4*

14 Public Goods Heterogeneous Preferences

14.1 Public Goods- Two People, Different Utilities

What if Bob and Alice had different amounts of utility for the coffee machine? May Alice cares more about coffee than Bob does. Suppose their utility functions are:

$$\begin{aligned}u_a(g_a, g_b) &= 20\sqrt{g_a + g_b} - g_a \\u_b(g_a, g_b) &= 10\sqrt{g_a + g_b} - g_b\end{aligned}$$

14.1.1 Equilibrium

To find Alice's best response function, find where her utility has zero slope with respect to g_a :

$$\frac{\partial (20\sqrt{g_a + g_b} - g_a)}{\partial g_a} = \frac{10}{\sqrt{g_a + g_b}} - 1$$

$$\frac{10}{\sqrt{g_a + g_b}} - 1 = 0$$

$$g_a = 100 - g_b$$

Alice's individually ideal total contributions are 100. Her full best response function is:

$$g_a = \begin{cases} 100 - g_b, & \text{if } g_b \leq 100 \\ 0, & \text{if } g_b > 100 \end{cases} \quad (4)$$

Now for Bob:

$$\frac{\partial (10\sqrt{g_a + g_b} - g_b)}{\partial g_a} = \frac{5}{\sqrt{g_a + g_b}} - 1$$

$$\frac{5}{\sqrt{g_a + g_b}} - 1 = 0$$

$$g_b = 25 - g_a$$

Bob's individually ideal total contributions are 25. Her full best response function is:

$$g_b = \begin{cases} 25 - g_a, & \text{if } g_a \leq 25 \\ 0, & \text{if } g_a > 25 \end{cases} \quad (5)$$

What is the equilibrium of this model? Well, we know Bob will never contribute more than 25, that means Alice will always contribute *something*. Can both possible be contributing? If so, then it would need to be that $g_a = 100 - g_a$ and $g_b = 25 - g_a$. Or, put another way, $g = 100$ and $g = 25$. This is impossible!

Info 14.1: Equilibrium with Different Utilities. In a public goods model, let g^* be the highest individually ideal total contributions. In any equilibrium, the total contributions are g^* and the only people who contribute are those who's individually ideal total contributions is g^* .

In this game, in equilibrium, the only person who contributes is Alice and she contributes $g_a = 100$.

14.1.2 Utilitarian Max

Now things get even more different than the model where everyone has the same preferences. Let's write the average of Alice and Bob's utilities. The utilitarian welfare of them choosing strategies g_a and g_b .

$$\frac{(20\sqrt{g_a + g_b} - g_a) + (10\sqrt{g_a + g_b} - g_b)}{2}$$

$$\frac{(30\sqrt{g} - g)}{2}$$

This is maximized where $g = 225$, the utilitarian ideal contributions. This can be achieved with a tax of $t = \frac{225}{2} = 112.5$, the utilitarian ideal tax.

What are the individually ideal taxes? Let's set up the individual utilities under a tax t .

Alice's utility under a tax t is:

$$u_a(t, t) = 20\sqrt{t + t} - t$$

This is maximized where $t = 200$. This would result in a total contribution of 400.

Bob's utility under a tax t is:

$$u_b(t, t) = 10\sqrt{t+t} - t$$

This is maximized where $t = 50$. Which would in a total contribution of 100.

Notice that $t = 112.5$ is no one's favorite tax. But what if we imagined an "average" of Alice and Bob. Alice's utility can be written like $20\sqrt{g} - g_i$. Bob's can be written $10\sqrt{g} - g_i$. The average of these is $15\sqrt{g} - g_i$. What if someone with those preferences were in a game like this with one other person. Their ideal tax would maximize:

$$15\sqrt{t+t} - t$$

This is maximized at 112.5. Thus, the utilitarian ideal tax can be thought of as the individually ideal tax for someone with preferences that are the *average* of those in the game.

14.2 Many People, Different Utilities

Now let's move to our most complex scenario. We have n people $P = \{1, 2, \dots, n\}$. Recall that g_{-i} is the total contributions made by people other than i . That is $g_{-i} = g - g_i$. Each has utility function:

$$a_i\sqrt{g_i + g_{-i}} - g_i$$

14.2.1 Equilibrium

Let's find the best response of person i :

$$\frac{\partial(a_i\sqrt{g_i + g_{-i}} - g_i)}{\partial g_i} = \frac{a_i}{2\sqrt{g_{-i} + g_i}} - 1$$

$$\frac{a_i}{2\sqrt{g_{-i} + g_i}} - 1 = 0$$

$$g_i = \frac{1}{4}a_i^2 - g_{-i}$$

The individually ideal total contributions is $\frac{1}{4}a_i^2$. Recall that in any public goods model where players have different preferences, the total contributions will always be equal to the maximum of the individually ideal total contributions and the only people who contribute anything are those who have individually ideal total contributions equal to that maximum. Suppose we have three people with $a_1 = 10, a_2 = 20, a_3 = 30$. Then, the individually ideal total contributions are 25, 100, 225. Thus, in equilibrium, the total contributions will be 225 and person 3 contributes all of it. The only equilibrium is $(0, 0, 225)$.

Suppose we have five people with $a_1 = 5, a_2 = 10, a_3 = 15, a_4 = 20, a_5 = 20$. Then, the individually ideal total contributions are 6.25, 25, 56.25, 100, 100. Thus, in equilibrium, the total contributions will be 100 made up of individual contributions from players 4 and 5. For instance, $(0, 0, 0, 50, 50)$, $(0, 0, 0, 25, 75)$, and $(0, 0, 0, 100, 0)$ are all equilibrium.

14.2.2 Utilitarian Max

Just as with two people, we can write the utilitarian welfare of any set of strategies as follows:

$$\frac{1}{n} \sum_{i=1}^n (a_i \sqrt{g} - g_i)$$

This can be written in terms of total contributions alone (g):

$$\frac{\sum_{i=1}^n a_i}{n} \sqrt{g} - \frac{1}{n} \sum_{i=1}^n g_i$$

$$\frac{\sum_{i=1}^n a_i}{n} \sqrt{g} - \frac{g}{n}$$

Note that $\frac{\sum_{i=1}^n a_i}{n}$ is simply the average value of a_i . Let's call it \bar{a} :

$$\bar{a} \sqrt{g} - \frac{g}{n}$$

This is maximized where the derivative of this function with respect to g is zero:

$$\frac{\partial (\bar{a} \sqrt{g} - \frac{g}{n})}{\partial g} = \frac{\bar{a}}{2\sqrt{g}} - \frac{1}{n}$$

$$\frac{\bar{a}}{2\sqrt{g}} - \frac{1}{n} = 0$$

$$g = \frac{\bar{a}^2 n^2}{4}$$

Notice that the utilitarian optimal total contributions only depends on the average preference \bar{a} . The utilitarian ideal total contributions for a model where people have different preferences a_i is the same as one where everyone has the same preferences, equal to the average \bar{a} .

We can get the utilitarian ideal tax by splitting up the total contributions above equally people the n people. We get:

$$t = \frac{\bar{a}^2 n}{4}$$

However, when people have different preferences, the individually idea taxes will not all be equal to the utilitarian ideal. Some will want more tax, and some will want less. Under tax t , an individual's utility is:

$$u_i(t) = a_i \sqrt{nt} - t$$

The ideal tax for any individual can be found by maximizing this utility with respect to t :

$$\begin{aligned}\frac{\partial (a_i \sqrt{nt} - t)}{\partial t} &= \frac{na_i}{2\sqrt{nt}} - 1 \\ \frac{na_i}{2\sqrt{nt}} - 1 &= 0 \\ t &= \frac{na_i^2}{4}\end{aligned}$$

How does an individual's ideal tax compare to the utilitarian ideal? If $a_i < \bar{a}$ they prefer a smaller tax than the utilitarian ideal tax.

How does an individual's ideal tax compare to the utilitarian ideal? If $a_i > \bar{a}$ they prefer a larger tax than the utilitarian ideal tax.

Essentially, people who have an above-average strength of preference for the public good prefer to have a higher tax to create a larger public good. Those with below average strength of preference prefer a lower tax and a smaller public good.

14.3 Voting

Suppose we have three people with $a_1 = 10, a_2 = 20, a_3 = 60$. Then, the individually ideal total contributions are 25, 100, 900. Thus, in equilibrium, the total contributions will be 900 and person 3 contributes all of it. The only equilibrium is $(0, 0, 900)$.

Since the average a_i is $\bar{a} = 30$, the utilitarian ideal contributions are $\frac{30^2(3^2)}{4} = 2025$. This can be achieved with the utilitarian ideal tax of $t = 675$ each. However, the "favorite" taxes of each person are respectively, 75, 300, and 2700.

What if they voted on these values? Here are their utilities of each tax:

	$i = 1$	$i = 2$	$i = 3$
$t = 75$	75	225	825
$t = 300$	0	300	1500
$t = 2700$	-1800	-900	2700

Table 5: Utilities of Each Tax Policy

In a pairwise vote between these options:

$t = 300$ beats $t = 75$ since both 2, 3 prefer it.
 $t = 300$ beats $t = 2700$ since both 1, 2 prefer it.
 $t = 75$ beats $t = 2700$ since both 1, 2 prefer it.

Thus, if we were to use the majority rule on these taxes, we get the following social preferences $300 \succ^* 75 \succ^* 2700$. Notice, that in this setting, the social preferences created by majority rule

is transitive! In this case, 300 is clearly the best. In fact, even if we gave players other taxes to vote on, the result would still be transitive.

The reason that the majority rule is transitive here is because individuals have **single peaked** preferences over the outcomes. Let's look at person 2's utility over possible taxes:

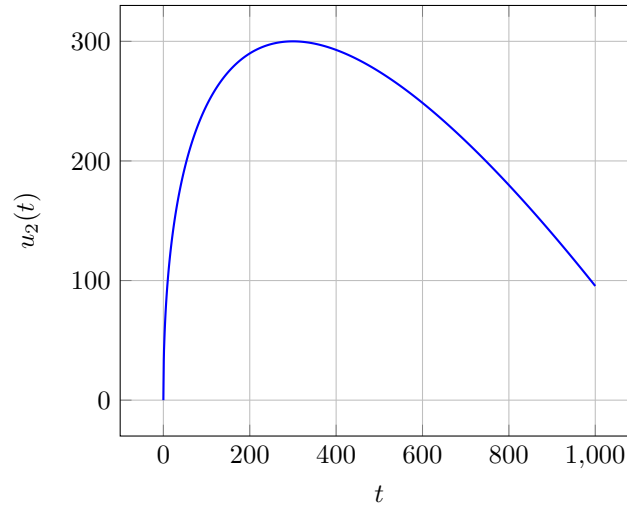


Figure 14.1: Person 2's Utility over Possible Taxes

Notice how there is a single peak. In fact, the preferences over taxes for everyone in a public goods model will always be single peaked. When preferences are single peaked, there is a nice exception to arrows impossibility theorem. This happens because certain preferences are impossible. Here, for instance, $2700 \succ_i 75 \succ_i 300$ could not possibly be someone's preference over the taxes. Arrows impossibility theorem only applies when every possible ordinal preferences over the outcomes are allowed.

Info 14.2: Single Peaked Preferences. Because preferences over taxes are single peaked in a public goods model, majority rule always results in a complete and transitive social preference ordering. Thus, majority rule will act as a preference aggregation rule that is complete, transitive, IIA, and Pareto efficient.

Because of this nice regularity, we have a way of taking voter's preferences over possible taxes and turning that into a decision that meets many of the requirements we would any in such a process. But what tax would win this process? As we saw above, 300 will beat both of the other taxes, 75 and 2700 in pairwise votes. In fact, if we include other possible taxes, $t = 300$ will beat those in pairwise votes as well. $t = 300$ will beat *every other* tax in a pairwise vote. That is because $t = 300$ is the **median voter's ideal tax**. That is, if we line up the favorite taxes in order (75, 300, 2700), it is right in the middle. As long as there is an odd number of people in a model, the favorite tax of the median voter will always be well-defined.

Definition 14.1: Median Voter's Ideal Tax. Arrange the individually ideal taxes in ascending order. If there are an odd number of people in the model, the individually ideal tax that is in the middle of this order is the **median voter's favorite tax**.

Why will the median always win? It is because half of voters prefer a tax no more than 300 and half prefer a tax no less than 300. If we hold a vote between $t = 300$ and any number higher than 300, the median voter and any voters who prefer a tax smaller than 300 will vote for 300. That has to be at least 50% of people! If we hold a vote between $t = 300$ and any number less than 300, the median voter and any voters who prefer a tax larger than 300 will vote for 300. That has to be at least 50% of people!

Thus, the median voter's ideal tax, here $t = 300$ will beat any other tax in a majority vote. In social choice, we call it a **Condorcet winner**.

Definition 14.2: Condorcet Winner. A Condorcet winner is an option in an election or decision-making process that beats every other option in a pairwise comparison based on majority rule.

Recall that Condorcet's paradox describes a situation in social choice where a cycle occurs in the social preferences resulting from majority rule preference aggregation. In those cases, there is no clear winner based on majority voting. This can happen in environments where preferences are not single-peaked. **However, in single-peaked environments, a Condorcet winner always exists.** Specifically, the median voter's most preferred outcome becomes the Condorcet winner because it can defeat any other option in a pairwise majority vote. In the given context, the median voter's ideal tax, $t = 300$, will beat any other tax in a majority vote.

14.4 Key Topics

For questions from this chapter on an exam, I will only use the log utility function.

- Be able to the best response function of a player in a public goods game with heterogeneous preferences as in *Exercise 14.1*
- Be able to find the individually ideal total contributions for the players in a public goods game with heterogeneous preferences as in *Exercise 14.1*
- Be able to find the equilibrium total contributions in a public goods game with heterogeneous preferences and understand that only the people who care most about the public good will contribute. *Exercise 14.1*
- Be able to find the utilitarian ideal total contributions for a public goods model with heterogeneous preferences as in *Exercises 14.2*
- Be able to find the utilitarian ideal tax as in *Exercise 14.2*
- Be able to find the individually ideal taxes as in *Exercise 14.3*
- Be able to find the median voter's favorite tax as in *Exercise 14.4*
- Understand that the median voter's favorite tax is a Condorcet winner and what this means.

15 Mechanism Design for Public Goods

Suppose Alice and Bob are considering a new coffee machine for the office. It costs \$500. Each has a “valuation” v_i for having the machine in the office.

If Alice’s valuation is 400 and Bob’s is 300, then they should buy the machine. Their total value is more than the cost of the machine.

One way for them to come up with the funds for the machine is to split the cost, each paying 250. Even if they do not know each other’s valuations, they can just ask each other “what is it worth to you?”. If both say it is worth more than 250 to both, they split the cost and buy the machine.

This is an economic **mechanism**. It is a way of having people reveal their **private information** (here, the valuations) and then using that information to pick an *outcome* (here, whether the coffee machine is purchased) and to assign *transfers* (here, how they will pay for the machine).

This mechanism is *incentive compatible* (or more formally *strategy-proof*) there is no reason for Alice or Bob to lie about their valuations regardless of what they think the other person is going to claim their valuation is. That’s because if their valuation is below 250, they would rather not pay 250 for the machine, so they will get their favorite outcome by telling the truth. If one has a valuation at or above 250 then if they lie and say it is some other number above 250, nothing will change about the outcome. If they lie and say it is something below 250 there are some instances where they will then get the outcome they like less. In summary, there is never an incentive to lie.

While the mechanism above is incentive compatible, it is not efficient. That is, it does not always lead to the efficient outcome. If Alice values the coffee machine at 400 and Bob at 200, it is still efficient to buy the machine. However, since Bob’s valuation is not 250 (half the cost of the machine), they will not buy it under this mechanism.

To make the efficient decision here, we need Alice to pay more than Bob for the coffee machine. But if we create a mechanism that assigns payments in a way that Alice will pay more in this scenario, we might break the incentive compatibility. Suppose that we come up with such a mechanism that does pick the efficient outcome in this scenario, but has Alice pay 250 and Bob pay 150.

Suppose now they are in the original scenario where Alice values it 400 and Bob 300. If Bob thinks Alice will say her valuation is 300 he might now have incentive to understate his valuation to 200. The outcome will remain that they buy the coffee machine, but now Alice will pay more of the cost.

This demonstrates the tension in mechanism design. To get people to tell the truth about their valuations, you have to be very careful about the incentives that the mechanism provides.

15.1 Formal Model

In this chapter we will use the following formal model. There are n people. There is one public good that costs c to provide. If the public good is provided, person i gets utility v_i (their *valuation*). t_i is the amount person i has to pay in *transfers*. Thus, if the public good is provided and person i has to pay t_i , then person i gets utility $v_i - t_i$. If it is not provided and they do not have to pay anything, their utility is 0.

15.1.1 Efficiency

What is the utilitarian optimal outcome? If the public good is provided, total utility is $\sum_{i=1}^n (v_i) - c$ since each person gets their valuation v_i , but society needs to pay c for the public good. If it is not provided, the total utility is 0. Thus, the utilitarian welfare of providing the public good is $\frac{\sum_{i=1}^n (v_i) - c}{n}$ and the utilitarian welfare of not providing the public good is zero. It is more efficient to provide it if:

$$\frac{\sum_{i=1}^n (v_i) - c}{n} \geq 0$$

This simplifies to:

$$\sum_{i=1}^n (v_i) \geq c$$

Definition 15.1: Efficiency. We say it is **efficient** to provide the public good if the sum of the valuations exceeds the cost

$$\sum_{i=1}^n (v_i) \geq c$$

15.2 Mechanism

A mechanism is a procedure that asks everyone their valuation v_i . From these, it picks an **outcome** and **transfers**.

Definition 15.2: Incentive Compatible. A mechanism is *incentive compatible* if there is never a scenario where someone could lie about their valuation and do strictly better. That is, if it is best to tell the truth.

Definition 15.3: Efficient. A mechanism is *efficient* if it chooses to provide the public good whenever it is efficient to do so and does not provide the public good when it is not efficient to do so.

Given these definitions, it seems like our initial goal should be to try and find a mechanism that is incentive-compatible and efficient. That is, it truthfully reveals each person's valuation and then uses that information to implement the efficient outcome.

15.2.1 An aside: Incentive-Compatibility and the Revelation Principle

Incentive compatibility is important because if a mechanism is not incentive compatible, then it is operating on information that is not valid.

Furthermore, there is a result in mechanism known as the **revelation principle** that says for any mechanism that is not incentive compatible, we could find one that ends up doing the same thing but is incentive compatible. That is, in non incentive-compatible mechanism, a player has some optimal “claim” about what their valuation is that might not be true. Let’s call this the *optimal lie*. We could always add another mechanism that first asks the valuation and then automatically implements their optimal lie in the original mechanism. The resulting combination would be incentive compatible and achieve the same outcome as the original non-incentive compatible mechanism. Thus, the revelation principle tells us that if we want to explore what is possible in **any** mechanism, we can just focus on the incentive-compatible mechanisms.

15.3 Some Mechanisms

In this section we will study a few mechanisms for deciding whether to provide a public good. We will start with our mechanism from the introduction.

15.3.1 Unanimous Agreement with Fixed-Split

This mechanism asks each person their valuation v_i . Then, if everyone has valuation $v_i \geq \frac{c}{n}$, the public good is provided and each person pays $\frac{c}{n}$.

Definition 15.4: Unanimous Agreement with Fixed-Split Mechanism. Outcome: The public good is provided when $v_i \geq \frac{c}{n}$ for everyone.
Transfers: If the public good is provided, everyone pays $\frac{c}{n}$.

Notice that in this mechanism, if someone has to pay, the amount is fixed and not dependent on the valuations. This is known as a **fixed split** mechanism.

Definition 15.5: Fixed-split Mechanism. A mechanism is a **fixed-split** mechanism if the transfers always sum to c when the public good is provided and those transfers only depend on whether the good is provided but not the valuations.

A fixed-split mechanism is always **budget balanced**. This means that, if the public good is produced in this mechanism, the amount of money collected in transfers is exactly the cost of the public good.

Definition 15.6: Budget Balance. A mechanism is **budget balanced** if the total amount collected in transfers when the public good is provided always equals the total cost of c .

While the unanimous agreement mechanism is incentive compatible and budget balanced, it is not efficient. Here is an example.

Example 15.1: Inefficient Example.

Suppose there are three people with valuations

$v_1 = 1000$, $v_2 = 2500$, and $v_3 = 7500$.

The cost of the public good is $c = 9000$.

To determine efficiency, we check if it is efficient to provide the public good. According to the condition for efficiency, we need:

$$\sum_{i=1}^3 v_i \geq c$$

Calculating the sum: $1000 + 2500 + 7500 = 11000$

Since $11000 \geq 9000$, it is efficient to provide the public good.

However, using the unanimous agreement with the fixed-split mechanism, each person needs to meet the condition $v_i \geq \frac{c}{n}$.

Here, $\frac{c}{3} = 3000$.

Checking each v_i :

- Person 1: $v_1 = 1000$ which is not ≥ 3000 .
- Person 2: $v_2 = 2500$ which is not ≥ 3000 .
- Person 3: $v_3 = 7500$ which is ≥ 3000 .

Since not all $v_i \geq 3000$, the unanimous agreement mechanism will not provide the public good even though it is actually efficient to do so.

However, the outcome sometimes selects the efficient outcome.

Example 15.2: Efficient Example. Suppose there are three people with valuations

$v_1 = 3500$, $v_2 = 3500$, and $v_3 = 7500$. The cost of the public good is $c = 9000$.

To determine efficiency, we check if it is efficient to provide the public good. According to the condition for efficiency, we need:

$$\sum_{i=1}^3 v_i \geq c$$

Calculating the sum: $3500 + 3500 + 7500 = 14500$. Since $14500 \geq 9000$, it is efficient to provide the public good.

Using the unanimous agreement with the fixed-split mechanism, each person needs to meet the condition $v_i \geq \frac{c}{n}$. Here, $\frac{c}{3} = 3000$.

Checking each v_i : - Person 1: $v_1 = 3500$ which is ≥ 3000 .

- Person 2: $v_2 = 3500$ which is ≥ 3000 .

- Person 3: $v_3 = 7500$ which is ≥ 3000 .

Since all $v_i \geq 3000$, the unanimous agreement mechanism will provide the public good.

In fact, whenever this mechanism chooses to produce the good it must be efficient to do so.

Info 15.1: Unanimous Agreement with Fixed Split and Efficiency. While the Unanimous Agreement with Fixed Split mechanism may sometimes choose *not* to provide the public good when it is effective to do so, any time it chooses to provide the public good, it must be efficient to do so.

This fact is due to the fact that if the public good is provided then $v_i \geq \frac{c}{n}$ for all n people. Thus, $\sum_{i=1}^n v_i \geq \sum_{i=1}^n \frac{c}{n} = c$. That is, if everyone is willing to pay $\frac{c}{n}$ then the total value must be at least c .

15.3.2 Median Mechanism with Fixed Split

In this subsection, we explore a mechanism where the provision of a public good is determined by the median valuation. This mechanism is consistent with what would happen if we had people vote on whether they are willing to pay $\frac{c}{n}$ to provide the public good. If the median of the valuations is above $\frac{c}{n}$ then at least half of the people would be willing to pay that amount and the vote would pass. Otherwise, if median of the valuations is below $\frac{c}{n}$ then at least half of the people would *not* be willing to pay that amount and the vote would fail.

To conduct this mechanism, ask each person their valuation v_i . Arrange these in ascending order. Let v_{median} be the middle value in this order- the median.

Definition 15.7: Median Mechanism with Fixed Split. **Outcome:** The public good is provided if the median valuation v_{median} is greater than or equal to $\frac{c}{n}$. **Transfers:** If the public good is provided, each person pays $\frac{c}{n}$.

In this mechanism, it is not required that all individuals meet the threshold valuation, only that the median value does.

Example 15.3: Efficiency Example. Suppose there are three people with valuations $v_1 = 2000$, $v_2 = 3500$, and $v_3 = 7000$. The cost of the public good is $c = 9000$.

To determine if the public good is provided, we find the median valuation, $v_{\text{median}} = 3500$.

Here, $\frac{c}{3} = 3000$. Since the median valuation $3500 \geq 3000$, the good is provided. Each person pays $\frac{9000}{3} = 3000$.

The total valuation is 12500, which is greater than the cost $c = 9000$. Thus, providing the good is efficient in this case.

Example 15.4: Inefficiency Example 1. Suppose there are three people with valuations $v_1 = 1000$, $v_2 = 3500$, and $v_3 = 4000$. The cost of the public good is $c = 9000$.

To determine if the public good is provided, we find the median valuation, $v_{\text{median}} = 3500$.

Here, $\frac{c}{3} = 3000$. Since the median valuation $3500 \geq 3000$, the good is provided. Each person pays $\frac{9000}{3} = 3000$.

The total valuation is 8500, which is less than the cost $c = 9000$. Thus, providing the good is **inefficient** in this case even though it is provided.

Example 15.5: Inefficiency Example 2. Suppose there are three people with valuations $v_1 = 1000$, $v_2 = 2500$, and $v_3 = 7500$. The cost of the public good is $c = 9000$.

To determine if the public good is provided, we find the median valuation, $v_{\text{median}} = 2500$.

Here, $\frac{c}{3} = 3000$. Since the median valuation $2500 < 3000$, the good is **not** provided.

The total valuation is 11000, which is more than the cost $c = 9000$. Thus, providing the good is **efficient** in this case even though it is not provided.

15.3.3 VCG “Pivot” Mechanism

The Vickrey-Clarke-Groves (VCG) mechanism for the provision of a public good always chooses the efficient outcome. This is not a fixed-split mechanism. Transfers are based on how each individual impacts the total utility of society. If the inclusion of person i 's valuation does not change the outcome, then they have no impact and pay nothing. On the other hand, if the inclusion of person i 's valuation does change the outcome, they have an impact, and we call them **pivotal**.

Below, let v_{-i} be the total valuation of everyone except i . That is: $v_{-i} = \sum_{j=1}^n (v_j) - v_i$.

Definition 15.8: Pivotal. A person i is **pivotal** if it is efficient to provide the public good, but it would be inefficient to provide if i did not exist. That is $\sum_{i=1}^n v_i \geq c$ and $v_{-i} < c$.

Definition 15.9: VCG Mechanism with Fixed Split. **Outcome:** Provide the public good if it is efficient $\sum_{i=1}^n v_i \geq c$
Transfers: If the public good is provided, and i is **pivotal**, they pay $c - v_{-i}$

The transfers in this mechanism are carefully designed so that each participant maximizes their own utility by truthfully reporting their valuation, regardless of what other participants report. This is because each person's payment is determined based on the impact of their presence on the decision to provide the public good.

Info 15.2: VCG Incentive Compatibility. The VCG mechanism ensures that each participant's best strategy is to report their true valuation. It is **incentive compatible**.

Example 15.6: VCG Example 1. Suppose there are three people with valuations $v_1 = 1000$, $v_2 = 2500$, and $v_3 = 7500$. The cost of the public good is $c = 9000$. First, check if it's efficient to provide the public good:

$$\sum_{i=1}^3 v_i = 1000 + 2500 + 7500 = 11000.$$

Since $11000 \geq 9000$, it is efficient to provide the public good.

Now, calculate v_{-i} for each person:

$$v_{-1} = 2500 + 7500 = 10000$$

$$v_{-2} = 1000 + 7500 = 8500$$

$$v_{-3} = 1000 + 2500 = 3500.$$

Determine if any person is pivotal:

Person 1: $v_{-1} = 10000 \geq 9000$ (not pivotal)

Person 2: $v_{-2} = 8500 < 9000$ (pivotal)

Person 3: $v_{-3} = 3500 < 9000$ (pivotal)

Transfer calculations for those who are pivotal: - Person 2 pays $c - v_{-2} = 9000 - 8500 = 500$.

- Person 3 pays $c - v_{-3} = 9000 - 3500 = 5500$.

To build some intuition about the incentive compatibility of the VCG mechanism, suppose Person 2 reports $v'_2 = 2000$ instead of 2500. The new apparent total valuation is: $1000 + 2000 + 7500 = 10500 \geq 9000$, so the public good is still provided and Person 2 remains pivotal. Their transfer remains unchanged at 500. The same analysis applies to any v'_2 for which 2 would remain pivotal.

On the other hand if 2 says their valuation is $v'_2 = 0$, the new apparent total valuation is: $1000 + 0 + 7500 = 8500 < 9000$, so the public good is not provided and person 2's new utility is 0. However, when they told the truth, their utility was $2500 - 500 = 2000$ so they were better off telling the truth. This analysis applies to any v'_2 for which 2 would become non-pivotal.

This is just a demonstration that players in the VCG mechanism have no incentive to misstate their valuations. This applies in **any** scenario.

Example 15.7: VCG Example 2. Suppose there are three people with valuations $v_1 = 0$, $v_2 = 0$, and $v_3 = 10000$. The cost of the public good is $c = 9000$. First, check if it's efficient to provide the public good:

$$\sum_{i=1}^3 v_i = 0 + 0 + 10000 = 10000.$$

Since $10000 \geq 9000$, it is efficient to provide the public good.
Now, calculate v_{-i} for each person:

$$\begin{aligned} v_{-1} &= 0 + 10000 = 10000 \\ v_{-2} &= 0 + 10000 = 10000 \\ v_{-3} &= 0 + 0 = 0 \end{aligned}$$

Determine if any person is pivotal:

Person 1: $v_{-1} = 10000 \geq 9000$ (not pivotal)
Person 2: $v_{-2} = 10000 \geq 9000$ (not pivotal)
Person 3: $v_{-3} = 0 < 9000$ (pivotal)

Transfer calculations for those who are pivotal: - Person 3 pays $c - v_{-3} = 9000 - 0 = 9000$.

Notice that while the total transfers add to c in the second example above (person 3 pays 9000), in the first example there is a deficit. The total transfers add to 6000 but it costs 9000 to produce the public good.

Info 15.3: VCG and Budget Balance. The VCG Mechanism is **efficient** and **incentive compatible**, but **not budget balanced**.

15.4 A Naive Approach to Efficiency and Budget Balance

So far we have seen no mechanism that is incentive compatible, efficient, and budget balanced. Let's try to make one.

Definition 15.10: Efficient Mechanism with a Fixed Split. **Outcome:** Provide the public good if it is efficient $\sum_{i=1}^n v_i \geq c$
Transfers: If the public good is provided, everyone pays $\frac{v}{c}$

Since this mechanism uses a fixed split it is budget balanced. It is also efficient since the decision rule always chooses the efficient outcome (given the submitted valuations). However, it is not incentive compatible.

Example 15.8: Efficient Fixed Split Example. Suppose there are three people with valuations $v_1 = 1000$, $v_2 = 2500$, and $v_3 = 7500$. The cost of the public good is $c = 9000$.

Let's check what happens if everyone tells the truth.

First, check if it's efficient to provide the public good:

$$\sum_{i=1}^3 v_i = 1000 + 2500 + 7500 = 11000.$$

Since $11000 \geq 9000$, it is efficient to provide the public good.

Transfers are 3000 for each person.

Notice that person 2 have utility -500 in this example. The public good is provided and they get utility 2500 from that, but they pay 3000. However that if v_2 says their valuation is $v'_2 = 0$ then the public good is not provided and they get utility 0. Thus, person 2 has an incentive to misstate their preferences. In general, Efficient Fixed Split Example is **not incentive compatible**.

15.5 An Impossibility

Here is a summary of the mechanisms we have seen so far:

Rule	Efficient	Incentive Compatible	Budget Balanced
Unanimous + Fixed Split	×	✓	✓
Median + Fixed Split	×	✓	✓
VCG	✓	✓	×
Efficient + Fixed Split	✓	×	✓

Table 6: Comparison of Mechanisms for Public Goods

Info 15.4: Green-Laffont Impossibility. There is no mechanism for public goods provision that is simultaneously incentive compatible, efficient, and budget-balanced.

The Green-Laffont Impossibility Theorem underscores a critical limitation in mechanism design for public goods. Designers must navigate trade-offs between efficiency, and budget balance. If you insist that any time the mechanism chooses to provide the public good, enough money is raised to actually do it, you cannot provide the public good every time it is efficient. If you want to provide the public good every time it is efficient, you may have to deal with budget deficit.

15.6 Key Topics

- Understand how the Unanimous Agreement Mechanism with Fixed-Split Mechanism works and apply it to problems as in *Exercises 15.1, 15.2, 15.4*

- Understand how the Median Mechanism with Fixed-Split works and apply it to problems as in *Exercises 15.1, 15.2, 15.4*
- Understand how the Unanimous Agreement with Fixed-Split Mechanism works and apply it to problems as in *Exercises 15.2, 15.4*
- Understand how the VCG Mechanism works and apply it to problems as in *Exercises 15.3, 15.5*

Part V

Exercises

16 Exercises

16.1 Exercises for Chapter 1

For the relations R below, when a pair is not listed, assume that the relation is not true of that pair.

Exercise 1.1: Is the relation “is a sibling of” on the set of all people complete? Is it transitive? Is it symmetric?

Exercise 1.2: Is the relation “is at least as tall as” on the set of all people complete? Is it transitive? Is it symmetric?

Exercise 1.3: Is the relation “has same birthday as” on the set of all people complete? Is it transitive? Is it symmetric?

Exercise 1.4: Explain why a relation that is complete and symmetric is trivial in the sense that it relates all pairs to each other.

Exercise 1.5: For the set $X = \{x, y, z\}$, identify if the following relations are transitive:

1. $R : xRy, yRz, xRz$
2. $R : xRx, yRy, zRz$
3. $R : xRy, yRz, zRx$

Exercise 1.6: For the set $X = \{p, q, r\}$, identify if the following relations are complete and transitive. When a relation is not both of these, say which assumption fails and why.

1. $R : pRp, qRq, rRr, pRq, qRr$
2. $R : pRp, qRq, rRr, pRq, qRr, pRr$
3. $R : pRp, qRq, rRr, pRq, qRp, qRr, rRq, pRr, rRp$
4. $R : pRp, qRq, rRr, pRq, qRp, pRr$

Exercise 1.7: Consider the preference relation that describes someone’s preferences over left l and right r shoes, where they only care about the number of usable pairs of shoes they consume. Sketch the indifference curves $\sim (1, 1)$ and $\sim (2, 2)$ on graph that has l on the x-axis and r on the y-axis. Label the set $\succ (2, 2)$.

Exercise 1.8: Consider the preference relation that describes someone’s preferences for red apples r and green apples g , where they only care about the total number of apples they have but not the color. Sketch the indifference curves $\sim (1, 1)$ and $\sim (2, 2)$ on graph that has r on the x-axis and g on the y-axis. Label the set $\succ (2, 2)$.

Exercise 1.9: Plot the following rational preference relation using a graph similar to those used in this chapter.

$$a \succsim b, a \succsim c, a \succsim d, b \succsim c, b \succsim d, c \succsim b, c \succsim d, a \succsim a, b \succsim b, c \succsim c, d \succsim d$$

Exercise 1.10: Write the strict preference relation \succ induced by each of the following weak preference relations:

1. $p \succsim p, q \succsim q, r \succsim r, p \succsim q, q \succsim r, p \succsim r$
2. $p \succsim p, q \succsim q, r \succsim r, p \succsim q, q \succsim p, q \succsim r, r \succsim q, p \succsim r, r \succsim p$

Exercise 1.11: Write the indifference relation \sim induced by each of the following weak preference relations:

1. $p \succsim p, q \succsim q, r \succsim r, p \succsim q, q \succsim r, p \succsim r$
2. $p \succsim p, q \succsim q, r \succsim r, p \succsim q, q \succsim p, q \succsim r, r \succsim q, p \succsim r, r \succsim p$

16.2 Exercises for Chapter 2

Exercise 2.1: Consider bundles a , b , and c with the given utilities $U(a) = 8$, $U(b) = 15$, and $U(c) = 10$. What complete and transitive relation \succsim does this represent?

Exercise 2.2: Provide an alternative utility function that represents the same preferences as those in the previous exercise.

Exercise 2.3: Suppose that a consumer's preferences can be represented by the utility function $u(t, m) = \sqrt{t} + m$. Which is true of this consumer's preferences? $(16, 3) \succ (4, 5)$, $(4, 5) \succ (16, 3)$, or $(4, 5) \sim (16, 3)$

Exercise 2.4: Suppose that a consumer's preferences can be represented by the utility function $u(t, m) = \sqrt{t} + m$. What is the utility of $(9, 4)$. What amount of money m solves the following $(9, 4) \sim (0, m)$?

Exercise 2.5: Suppose that a consumer's preferences can be represented by the utility function $u(t, m) = \sqrt{t} + m$. Sketch the indifference curve $\sim (9, 4)$ on a graph with t on the x-axis and m on the y-axis. Feel free to use a computer to help you with this.

Exercise 2.6: Discuss the following statement: *Economists do not have to believe that utility functions exist in the minds of consumers for the concept to be useful.*

Exercise 2.7: For the set $X = \{p, q, r\}$, write down a utility function that represents each of these preference relations.

1. $p \succsim p, q \succsim q, r \succsim r, p \succsim q, q \succsim r, p \succsim r$
2. $p \succsim p, q \succsim q, r \succsim r, p \succsim q, q \succsim p, q \succsim r, r \succsim q, p \succsim r, r \succsim p$

Exercise 2.8: Suppose that a consumer's preferences can be represented by the utility function $u(x, y) = x + y$. Sketch the indifference curve $\sim (4, 4)$ on a graph with x on the x-axis and y on the y-axis.

Exercise 2.9: Suppose that a consumer's preferences can be represented by the utility function $u(x, y) = 2x + y$. Sketch the indifference curve $\sim (4, 4)$ on a graph with x on the x-axis and y on the y-axis.

Exercise 2.10: Suppose that a consumer's preferences can be represented by the utility function $u(x, y) = \min\{x, y\}$. Sketch the indifference curve $\sim (4, 4)$ on a graph with x on the x-axis and y on the y-axis.

16.3 Exercises for Chapter 3

Exercise 3.1: Add a third person (Camden) to [Example 3.3](#) who likes the smell of warm fish.

Exercise 3.2: Add a third person (Camden) to [Example 3.4](#) who is so inept that if he attempts to clean, even with the help of others, he makes the kitchen worse than if no one had tried to clean at all. For preferences, *there are many right answers*, justify your answers with an explanation of why Alice, Bob, and Camden might have those particular preferences in the context of this "story".

Exercise 3.3: Add a third person (Camden) to [Example 3.5](#) who likes the smell of warm fish. Ensure that the utilities you choose are consistent with your solution to Exercise 3.1.

Exercise 3.4: Add a third person (Camden) to [Example 3.6](#) who is so inept that if he attempts to clean, even with the help of others, he makes the kitchen worse than if no one had tried to clean at all. Ensure that the utilities you choose are consistent with your solution to Exercise 3.2.

16.4 Exercises for Chapter 4

Exercise 4.1: Alice and Bob have the following preferences over outcomes $\{a, b, c\}$. Graph the Pareto dominance relation for this example as demonstrated in this chapter.

- Alice: $a \succ b \succ c$
- Bob: $a \sim b \succ c$
- Camden: $a \sim b \succ c$

Exercise 4.2: Alice and Bob have the following preferences over outcomes $O = \{a, b, c\}$. What are the Pareto efficient outcomes?

- Alice: $a \succ b \succ c$
- Bob: $b \succ a \succ c$

Exercise 4.3: Alice, Bob, and Camden have the following preferences over outcomes $O = \{a, b, c, d\}$. What are the Pareto efficient outcomes?

- Alice: $b \succ d \succ c \succ a$
- Bob: $d \succ c \succ a \succ b$
- Camden: $a \succ c \succ d \succ b$

Exercise 4.4: Alice, Bob, Camden, and Dave have the following preferences over outcomes $O = \{a, b, c, d, e, f\}$. What are the Pareto efficient outcomes?

- Alice: $a \succ b \succ c \succ d \succ e \succ f$
- Bob: $a \succ d \succ c \succ b \succ e \succ f$
- Camden: $a \succ c \succ b \succ e \succ f \succ d$
- Dave: $b \succ c \succ a \succ f \succ d \succ e$

Exercise 4.5: Alice and Bob have the following preferences over outcomes $\{a, b, c, d\}$. Graph the Pareto dominance relation for this example as demonstrated in this chapter.

- Alice: $a \sim b \sim c \succ d$
- Bob: $a \succ b \sim c \succ d$

Exercise 4.6: For the example above, which outcomes Pareto dominate others? Which outcomes strictly Pareto dominate others?

Exercise 4.7: For the example above, what are the Pareto efficient outcomes.

Exercise 4.8: Consider the following model based on [Example 3.5](#), but where Bob can choose

to leave. Plot the possibly utility combinations and mark the Pareto efficient outcomes.

$$\begin{aligned}
 P &= \{\text{Alice}, \text{Bob}\} \\
 O &= \{\text{yes/stays}, \text{yes/leaves}, \text{no/stays}, \text{no/leaves}\} \\
 U_a(o) &= \begin{cases} 10 & o = \text{yes/stays} \\ 10 & o = \text{yes/leaves} \\ 5 & o = \text{no/stays} \\ 5 & o = \text{no/leaves} \end{cases} \\
 U_b(o) &= \begin{cases} 1 & o = \text{yes/stays} \\ 5 & o = \text{yes/leaves} \\ 10 & o = \text{no/stays} \\ 5 & o = \text{no/leaves} \end{cases}
 \end{aligned}$$

Exercise 4.9: Alice and Bob have the following preferences over outcomes $\{a, b, c, d\}$. Which outcomes Pareto dominate others? Which outcomes *strictly* Pareto dominate others?

- Alice: $a \sim b \sim c \succ d$
- Bob: $a \succ b \sim c \succ d$

Exercise 4.10: Alice and Bob have the following preferences over outcomes $\{b, c, d\}$. Which outcomes Pareto dominate others? Which outcomes *strictly* Pareto dominate others?

- Alice: $b \sim c \succ d$
- Bob: $b \sim c \succ d$

Exercise 4.11: Alice and Bob have the following preferences over outcomes $\{b, c, d\}$. Which outcomes are Pareto Efficient?

- Alice: $b \sim c \succ d$
- Bob: $b \sim c \succ d$

16.5 Exercises for Chapter 5

The following exercises make use of this example.

Example 16.1: Example 3.. There are five people and three outcomes. Their preferences are:

1: $a \succ b \succ c$

2: $a \succ b \succ c$

3: $b \succ c \succ a$

4: $b \succ c \succ a$

5: $c \succ a \succ b$

Exercise 5.1: Informally, what social preferences would you assign for this example and why?

Exercise 5.2: What social preferences result from applying *Majority Rule* to this example? Is this an example of the Condorcet Paradox? How do you know?

Exercise 5.3: What social preferences result from applying *Copeland's Method* to this example?

Exercise 5.4: What social preferences result from applying *Borda Count* to this example where a top-ranked outcome gets a score of 1, a second-ranked outcome gets a score of 0 and a last-ranked outcome gets a score of -1 .

16.6 Exercises for Chapter 6

Exercise 6.1: Ask ChatGPT or another *AI* of your choice to provide an example of a preference aggregation rule that meets completeness, transitivity, Pareto efficiency, and IIA. Can you convince it to lie to you? If so, provide the prompt you used.

Exercise 6.2: As shown in the table above **Plurality Vote** is not Pareto efficient and it does not obey IIA. Come up with counter-examples using three people and three outcomes a, b, c that show it does not meet these properties.

Exercise 6.3: As shown in the table above **Veto** is not Pareto efficient and it does not obey IIA. Come up with counter-examples using three people and three outcomes a, b, c that show it does not meet these properties.

Exercise 6.4: Come up with a preference aggregation rule that is complete, transitive, IIA, but **not Pareto efficient**.

16.7 Exercises for Chapter 7

The following exercises make use of this example.

Example 16.2: Example 3.. There are five people and three outcomes. Their preferences are:

1: $a \succ b \succ c$

2: $a \succ b \succ c$

3: $b \succ c \succ a$

4: $b \succ c \succ a$

5: $c \succ a \succ b$

Exercise 7.1: What choice/choices result from applying *Plurality Vote- Social Choice* to this example?

Exercise 7.2: What choice/choices result from applying *Borda Count- Social Choice* to this example?

The Veto social choice function chooses the outcome/s that the fewest number of people rank last.

Exercise 7.3: What choice/choices result from applying *Veto- Social Choice* to this example?

16.8 Exercises for Chapter 8

Exercise 8.1: Describe Arrow's impossibility theorem as it applies to *social choice functions* in three paragraphs to someone who has never taken economics or mathematics. Be sure to describe the properties that cannot simultaneously be achieved in as simple terms as possible.

Exercise 8.2: Write down a counter-example using three people and three outcomes a, b, c showing that Majority rule the social choice function is **not nonempty**.

Exercise 8.3: Write down a counter-example using three people and three outcomes a, b, c showing that Unanimity rule the social choice function is **not nonempty**.

Exercise 8.4: Write down a counter-example using three people and three outcomes a, b, c showing that Plurality vote the social choice function is **not IIA**.

Exercise 8.5: Write down a counter-example using three people and three outcomes a, b, c showing that Borda count the social choice function is **not IIA**.

The **veto** rule chooses the outcome or outcomes that are the least favorite of the fewest number of people.

Exercise 8.6: Write down a counter-example using three people and three outcomes a, b, c showing that Veto the social choice function is **not Pareto efficient**.

16.9 Exercises for Chapter 9

Exercise 9.1: Describe the Gibbard-Satterthwaite theorem in simple terms as-if explaining it to someone who has never studied economics or mathematics before.

Exercise 9.2: Provide an example with 5 people and 3 outcomes where someone can manipulate their preferences and make the chosen outcome better for themselves in **Plurality Vote** where ties are broken by their alphabetical order.

Exercise 9.3: Provide an example with 3 people and 3 outcomes where someone can manipulate their preferences and make the chosen outcome better for themselves in **Borda Count** where ties are broken by their alphabetical order.

16.10 Exercises for Chapter 10

Example 16.3: The Coffee Machine, Model.

$$\begin{aligned}P &= \{a, b\} \\O &= \{ab, a, b, n\} \\U_a(o) &= \begin{cases} 15 & \text{if } o = ab \\ 5 & \text{if } o = a \\ 20 & \text{if } o = b \\ 10 & \text{if } o = n \end{cases} \\U_b(o) &= \begin{cases} 15 & \text{if } o = ab \\ 20 & \text{if } o = a \\ 5 & \text{if } o = b \\ 10 & \text{if } o = n \end{cases}\end{aligned}$$

Exercise 10.1: Plot these utilities with u_a on the x-axis and u_b on the y-axis.

Exercise 10.2: Calculate the utilitarian welfare (average version) for each outcome. What outcome maximizes utilitarian welfare?

Exercise 10.3: Calculate the Rawlsian welfare for each outcome. What outcome maximizes Rawlsian welfare?

Exercise 10.4: Again plot the utility combinations of each outcome with u_a on the x-axis and u_b on the y-axis. Add the Rawlsian welfare contours that give combinations of utilities that provide the same welfare as each of the outcomes.

Exercise 10.5: Calculate the Nash welfare for each outcome. What outcome maximizes Nash

welfare?

Now consider the following model where Alice either microwaves fish (yes) or not (no) and Bob either stays in the office (stay) or leaves (leave).

Example 16.4: Microwaving Fish, Model.

$$\begin{aligned}
 P &= \{a, b\} \\
 O &= \{yes/stay, yes/leave, no/stay, no/leave\} \\
 U_a(o) &= \begin{cases} 25 & \text{if } o = (yes/stay) \\ 25 & \text{if } o = (yes/leave) \\ 10 & \text{if } o = (no/stay) \\ 10 & \text{if } o = (no/leave) \end{cases} \\
 U_b(o) &= \begin{cases} 5 & \text{if } o = (yes/stay) \\ 10 & \text{if } o = (yes/leave) \\ 20 & \text{if } o = (no/stay) \\ 10 & \text{if } o = (no/leave) \end{cases}
 \end{aligned}$$

Exercise 10.6: Plot these the utility combinations of each outcome with u_a on the x-axis and u_b on the y-axis.

Exercise 10.7: Calculate the utilitarian welfare (average version) for each outcome. What outcome maximizes utilitarian welfare?

Exercise 10.8: Again plot the utility combinations of each outcome with u_a on the x-axis and u_b on the y-axis. Add the utilitarian welfare contours that give combinations of utilities that provide the same welfare as each of the outcomes.

Exercise 10.9: Calculate the Rawlsian welfare for each outcome. What outcome maximizes Rawlsian welfare?

Exercise 10.10: Calculate the Nash welfare for each outcome. What outcome maximizes Nash welfare?

16.11 Exercises for Chapter 11

Example 16.5: Microwaving Fish, Model.

$$\begin{aligned}
 P &= \{a, b\} \\
 O &= \{yes/stay, yes/leave, no/stay, no/leave\} \\
 U_a(o) &= \begin{cases} 25 & \text{if } o = (yes/stay) \\ 25 & \text{if } o = (yes/leave) \\ 10 & \text{if } o = (no/stay) \\ 10 & \text{if } o = (no/leave) \end{cases} \\
 U_b(o) &= \begin{cases} 5 & \text{if } o = (yes/stay) \\ 10 & \text{if } o = (yes/leave) \\ 20 & \text{if } o = (no/stay) \\ 10 & \text{if } o = (no/leave) \end{cases}
 \end{aligned}$$

Exercise 11.1: Plot the set of utilities achievable with randomization. Mark the Pareto frontier.

Exercise 11.2: What outcome (no randomization) maximizes Nash welfare? What about Rawlsian welfare? What about utilitarian welfare?

Exercise 11.3: What point on the Pareto frontier, including randomization, maximizes Nash welfare?

Exercise 11.4: What point on the Pareto frontier, including randomization, maximizes Rawlsian welfare?

Exercise 11.5: What point on the Pareto frontier, including randomization, maximizes Utilitarian welfare?

Exercise 11.6: Plot the set of utilities achievable with side-payments.

Exercise 11.7: What pair of utilities simultaneously maximizes Nash, Utilitarian, and Rawlsian welfare when side-payments are permitted? To achieve this, what outcome is chosen and what payments are made?

Example 16.6: Simple Model 2.

$$\begin{aligned}P &= \{1, 2\} \\O &= \{a, b\} \\U_1(o) &= \begin{cases} 30 & \text{if } o = a \\ 0 & \text{if } o = b \end{cases} \\U_2(o) &= \begin{cases} 0 & \text{if } o = a \\ 20 & \text{if } o = b \end{cases}\end{aligned}$$

Exercise 11.8: What randomized or non-randomized outcome maximizes Utilitarian welfare? What are the expected utilities?

Exercise 11.9: What randomized or non-randomized outcome maximizes Nash welfare? What are the expected utilities?

Exercise 11.10: What randomized or non-randomized outcome maximizes Rawlsian welfare? What are the expected utilities?

Exercise 11.11: What pair of utilities simultaneously maximizes Nash, Utilitarian, and Rawlsian welfare when side-payments are permitted? To achieve this, what outcome is chosen and what payments are made?

Example 16.7: Simple Model 3.

$$\begin{aligned}P &= \{1, 2\} \\O &= \{a, b\} \\U_1(o) &= \begin{cases} 5 & \text{if } o = a \\ 0 & \text{if } o = b \end{cases} \\U_2(o) &= \begin{cases} 0 & \text{if } o = a \\ 8 & \text{if } o = b \end{cases}\end{aligned}$$

Exercise 11.12: What randomized or non-randomized outcome maximizes Utilitarian welfare? What are the expected utilities?

Exercise 11.13: What randomized or non-randomized outcome maximizes Nash welfare? What are the expected utilities?

Exercise 11.14: What randomized or non-randomized outcome maximizes Rawlsian welfare? What are the expected utilities?

Exercise 11.15: What pair of utilities simultaneously maximizes Nash, Utilitarian, and Rawlsian welfare when side-payments are permitted? To achieve this, what outcome is chosen and what payments are made?

16.12 Exercises for Chapter 12

Exercise 12.1: Consider a scenario in which two friends decide whether to go to a movie or stay at home. Each prefers to go to the movie if the other goes. They prefer to stay home if the other stays home. Formalize this situation as a game using a 2×2 table. What is/are the Nash equilibrium(a) of your game?

For the next three questions, consider a game where two companies decide whether to launch l a new product or not n . Below, the row player is player 1 and the column player is player 2.

s_1, s_2	l	n
l	(20, 20)	(30, 5)
n	(5, 30)	(10, 10)

Exercise 12.2: What are the Pareto efficient outcomes of this game?

Exercise 12.3: Write down the best response function for player 1 and player 2.

Exercise 12.4: What is/are the Nash equilibrium/a of this game?

For the next three questions, consider the following game. Alice and Bob work on different projects at work. Occasionally they have more work to do than they can handle. They cooperate c by helping the other person out when appropriate or not cooperate n and say they are too busy. Here's the game.

s_a, s_b	c	n
c	(15, 15)	(0, 20)
n	(20, 0)	(5, 5)

Exercise 12.5: What are the Pareto efficient outcomes of this game?

Exercise 12.6: Write down the best response function for Alice and Bob.

Exercise 12.7: What is/are the Nash equilibrium/a of this game?

For the next three questions, consider the following game. Two neighboring farmers must decide whether to invest i or not n in a new irrigation system.

s_1, s_2	i	n
i	(25, 25)	(35, 5)
n	(5, 35)	(10, 10)

Exercise 12.8: What are the Pareto efficient outcomes of this game?

Exercise 12.9: Write down the best response function for 1 and 2?

Exercise 12.10: What is/are the Nash equilibrium/a of this game?

16.13 Exercises for Chapter 13

3 people are considering whether to build a public good. The contribution of person i is g_i and the total contribution of everyone who is not i is g_{-i} . Each person's utility is:

$$100 \ln(g_i + g_{-i}) - g_i$$

.

Exercise 13.1: What is the best response of person i to g_{-i} ?

Exercise 13.2: What are the individually ideal total contributions of person i ? What are the total contributions in equilibrium?

Exercise 13.3: What are the utilitarian ideal total contributions when $n = 10$

Exercise 13.4: What are the utilitarian ideal total contributions for any n ?

16.14 Exercises for Chapter 14

3 people are considering whether to build a public good. The contribution of person i is g_i and the total contribution of everyone who is not i is g_{-i} . Each person's utility is $a_i \ln(g_i + g_{-i}) - g_i$. Assume $a_1 = 100$, $a_2 = 200$, $a_3 = 500$.

Exercise 14.1: If everyone chooses individually how much to contribute, what are the individually ideal total contributions of each person? What are the total contributions in equilibrium?

Exercise 14.2: What are the utilitarian ideal total contributions? What is the utilitarian ideal tax?

Exercise 14.3: What are the individually ideal taxes of each person?

Exercise 14.4: What is the median voter's favorite tax? What are the total contributions under that tax?

16.15 Exercises for Chapter 15

Exercise 15.1: Suppose there are three individuals—Alice, Bob, and Carol—considering the purchase of a new printer for the office. The printer costs \$600. Their valuations are $v_A = 300$, $v_B = 250$, and $v_C = 200$.

1. Using the **Unanimous Agreement with Fixed-Split** mechanism, determine whether the printer will be purchased and how much each person will pay.
2. Is the outcome efficient under this mechanism?

Exercise 15.2: Suppose there are three individuals—Alice, Bob, and Carol—considering the purchase of a new printer for the office. The printer costs \$600. Their valuations are $v_A = 300$, $v_B = 250$, and $v_C = 150$.

1. Apply the **Median Mechanism with Fixed Split** mechanism to determine whether the printer will be purchased and how much each person will pay.
2. Does the **Median Mechanism with Fixed Split** choose the efficient outcome in this case?
3. Apply the **Unanimous Mechanism with Fixed Split** mechanism to determine whether the printer will be purchased and how much each person will pay.
4. Does the **Unanimous Mechanism with Fixed Split** choose the efficient outcome in this case?

Exercise 15.3: Consider a public good that costs $c = \$1000$ and three individuals—Alice, Bob, and Carol—with valuations $v_a = 400$, $v_b = 350$, and $v_c = 300$ respectively.

1. What outcome does the **VCG Mechanism** select?
2. In the **VCG Mechanism** who is pivotal?
3. What are the transfers for each person?
4. Is there a budget deficit in this case?

Exercise 15.4: Suppose there are four individuals—Alice, Bob, Carol, and Dave—deciding whether to invest in a community garden costing \$2000. Their valuations are $v_1 = 600$, $v_2 = 550$, $v_3 = 500$, $v_4 = 100$ and $v_5 = 50$ respectively.

1. Apply the **Median Mechanism with Fixed Split** mechanism to determine whether the printer will be purchased and how much each person will pay.
2. Does the **Median Mechanism with Fixed Split** choose the efficient outcome in this case?

3. Apply the **Unanimous Mechanism with Fixed Split** mechanism to determine whether the printer will be purchased and how much each person will pay.
4. Does the **Unanimous Mechanism with Fixed Split** choose the efficient outcome in this case?

Exercise 15.5: Suppose there are three individuals—Alice, Bob, and Carol—deciding whether to purchase a shared online subscription costing \$150 per month. Their valuations are $v_A = 80$, $v_B = 75$, and $v_C = 60$ respectively.

1. What outcome does the **VCG Mechanism** select?
2. In the **VCG Mechanism** who is pivotal?
3. What are the transfers for each person?
4. Is there a budget deficit in this case?

Exercise 15.6: Describe the Green-Laffont Impossibility theorem in simple terms as-if explaining it to someone who has never studied economics or mathematics before.

17 Solutions

17.1 Solutions for Chapter 1

Solution for Exercise 1.1.

No, No, Yes

It is not complete since there are people who are not siblings of each other at all. It is not transitive if we think of sibling as being something broader than “having the same biological parents”. For instance, person a ’s step-sibling b might have a step-sibling c who is not a step-sibling of a . Sibling is symmetric though since if a is a sibling of b then b is a sibling of a .

Solution for Exercise 1.2.

Yes, Yes, No

Solution for Exercise 1.3.

No, Yes, Yes

Solution for Exercise 1.4.

If a relation is symmetric then if xRy is true, so is yRx . If yRx is true then so is xRy . Thus, if a relation is symmetric then either both of xRy and yRx are true or neither are true. But by completeness, for all x and y , at least one must be true. Thus, combining these, we have: for all x and y both xRy and yRx are true. Thus, every possible pair of things are related in both directions. This is the **universal** relation.

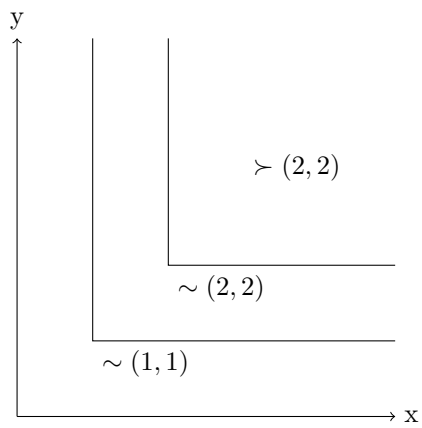
Solution for Exercise 1.5.

1. *Transitive.*
2. *Transitive.*
3. *Not Transitive. Missing xRz)*

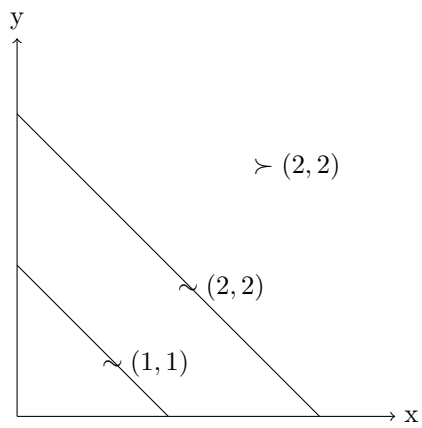
Solution for Exercise 1.6.

1. *Neither.*
2. *Both.*
3. *Both.*
4. *Neither.*

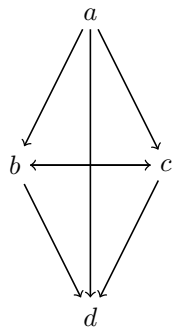
Solution for Exercise 1.7.



Solution for Exercise 1.8.



Solution for Exercise 1.9.



Solution for Exercise 1.10.

1. $p \succ q, p \succ r, q \succ r$
2. There are no strict preferences.

Solution for Exercise 1.11.

1. $p \sim p, q \sim q, r \sim r$
2. $p \sim p, q \sim q, r \sim r, p \sim q, q \sim r, p \sim r$

17.2 Solutions for Chapter 2

Solution for Exercise 2.1.

$$a \succsim a, b \succsim b, c \succsim c, b \succsim a, b \succsim c, c \succsim a$$

Solution for Exercise 2.2.

For example, $U(A) = 9, U(B) = 16, U(C) = 12$. But there are **many** possible solutions.

Solution for Exercise 2.3.

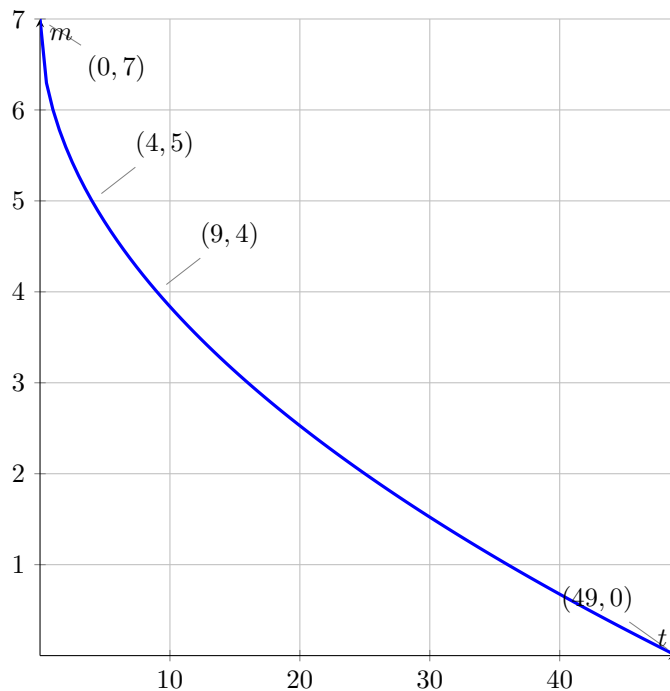
$$(4, 5) \sim (16, 3)$$

Solution for Exercise 2.4.

$$U(9, 4) = 7 \text{ and } (9, 4) \sim (0, 7).$$

Solution for Exercise 2.5.

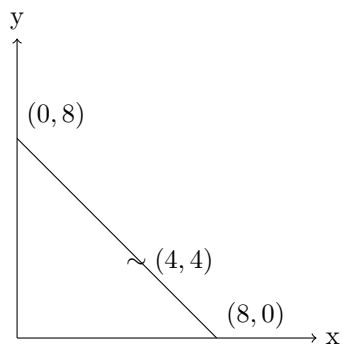
Solution for Exercise 2.6.



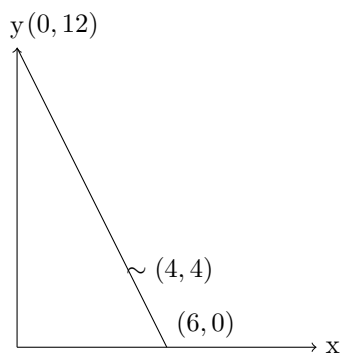
Solution for Exercise 2.7.

Utility is merely a mathematical tool to represent preferences. Preferences are really what we care about. As long as those preferences meet some basic conditions, we can use utility to represent them. This allows us to use the numerical tools of mathematics to work with preferences.

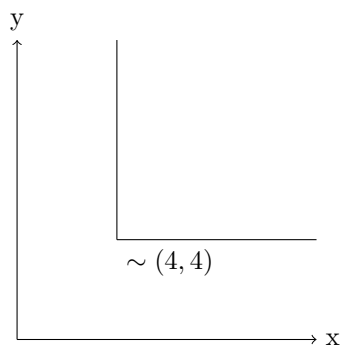
Solution for Exercise 2.8.



Solution for Exercise 2.9.



Solution for Exercise 2.10.



17.3 Solutions for Chapter 3

Solution for Exercise 3.1.

$$\begin{aligned}
 P &= \{a, b, c\} \\
 O &= \{\text{yes}, \text{no}\} \\
 \text{yes} &\succ_a \text{no} \\
 \text{no} &\succ_b \text{yes} \\
 \text{yes} &\succ_c \text{no}
 \end{aligned}$$

Solution for Exercise 3.2.

Here, I will assume that if Camden cleans, then even if someone else helps, they cannot make things any better. In this case, as long as Camden is involved, Alice and Bob just want to avoid even trying to put in work. There are many ways to answer this question though. The most important thin is that any outcome involving Camden should be lower than n for everyone.

$$\begin{aligned}
 P &= \{a, b, c\} \\
 O &= \{abc, ab, ac, bc, a, b, c, n\} \\
 b &\succ_a ab \succ_a a \succ_a n \succ_a c \sim_a bc \succ_a abc \sim_a ac \\
 a &\succ_b ab \succ_b b \succ_b n \succ_b c \sim_b ac \succ_b abc \sim_b bc \\
 a &\sim_c b \sim_c ab \succ_c n \succ_c c \sim_c ac \sim_c bc \sim_c abc
 \end{aligned}$$

Solution for Exercise 3.3.

Here, I will just assume that Camden's utility is the same as Alice's. Though, you can really choose any utility for Camden as long as the utility of yes is higher than no.

$$\begin{aligned}
 P &= \{a, b, c\} \\
 O &= \{\text{yes}, \text{no}\} \\
 u_a(o) &= \begin{cases} 10 & o = \text{yes} \\ 9 & o = \text{no} \end{cases} \\
 u_b(o) &= \begin{cases} 1 & o = \text{yes} \\ 10 & o = \text{no} \end{cases} \\
 u_c(o) &= \begin{cases} 10 & o = \text{yes} \\ 9 & o = \text{no} \end{cases}
 \end{aligned}$$

Solution for Exercise 3.4.

My goal here is just to make the numbers represent my solution to 3.2.

$$P = \{a, b, c\}$$

$$O = \{abc, ab, ac, bc, a, b, c, n\}$$

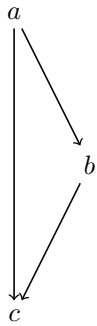
$$u_a(o) = \begin{cases} 12 & \text{if } o = ab \\ 10 & \text{if } o = a \\ 25 & \text{if } o = b \\ 5 & \text{if } o = n \\ 3 & \text{if } o = c \\ 1 & \text{if } o = ac \\ 3 & \text{if } o = bc \\ 1 & \text{if } o = abc \end{cases}$$

$$u_b(o) = \begin{cases} 12 & \text{if } o = ab \\ 25 & \text{if } o = a \\ 10 & \text{if } o = b \\ 5 & \text{if } o = n \\ 3 & \text{if } o = c \\ 3 & \text{if } o = ac \\ 1 & \text{if } o = bc \\ 1 & \text{if } o = abc \end{cases}$$

$$u_c(o) = \begin{cases} 25 & \text{if } o = ab \\ 25 & \text{if } o = a \\ 25 & \text{if } o = b \\ 5 & \text{if } o = n \\ 1 & \text{if } o = c \\ 1 & \text{if } o = ac \\ 1 & \text{if } o = bc \\ 1 & \text{if } o = abc \end{cases}$$

17.4 Solutions for Chapter 4

Solution for Exercise 4.1.



Solution for Exercise 4.2.

a and b

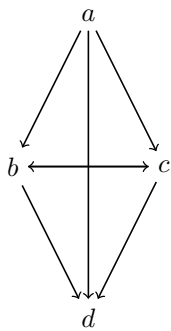
Solution for Exercise 4.3.

All outcomes: a, b, c, d

Solution for Exercise 4.4.

All outcomes: a, b, c

Solution for Exercise 4.5.



Solution for Exercise 4.6.

a Pareto dominates all others.

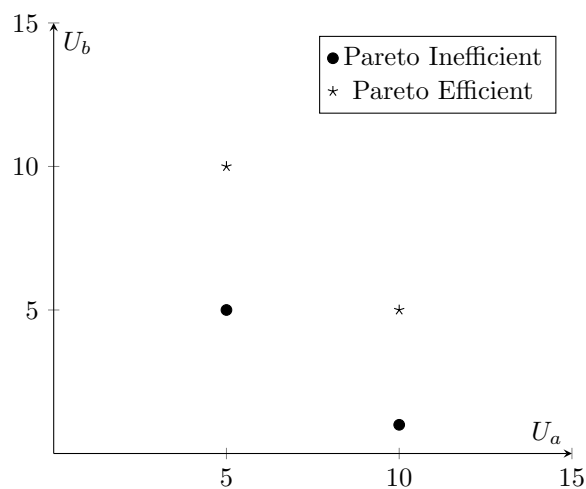
b Pareto dominates c and d

c Pareto dominates b and d

Solution for Exercise 4.7.

Only a

Solution for Exercise 4.8.



Solution for Exercise 4.9.

- a* Pareto dominates all others.
- b* Pareto dominates *c, d*
- c* Pareto dominates *b, d*
- a* Strictly Pareto dominates all others.
- b* Strictly Pareto dominates *d*
- c* Strictly Pareto dominates *d*

Solution for Exercise 4.10.

- b* Pareto dominates *d*
- c* Pareto dominates *d*
- b* Strictly Pareto dominates *d*
- c* Strictly Pareto dominates *d*

Solution for Exercise 4.11.

b, c

17.5 Solutions for Chapter 5

Solution for Exercise 5.1.

You can really say whatever you want for this. I personally like $b \succ^* a \succ^* c$ since *b* is the favorite of 2 people and second favorite of 2 people. *a* is the favorite of 2 but the second favorite of only 1.

Solution for Exercise 5.2.

$a \succ^* b, b \succ^* c, c \succ^* a$. This is condorcet Paradox example. The preferences are intransitive.

Solution for Exercise 5.3.

$a \sim^* b \sim^* c$

Solution for Exercise 5.4.

$b \succ^* a \succ^* c$

17.6 Solutions for Chapter 6

Solution for Exercise 6.1.

Greg: Give me an example of a preference aggregation rule that is complete, transitive, Pareto efficient, and obeys IIA.

ChatGPT: One classic example of a preference aggregation rule that satisfies completeness, transitivity, Pareto efficiency, and independence of irrelevant alternatives (IIA) is the Borda count.

Characteristics: Completeness: The Borda count allows for the ranking of all alternatives based on individual preferences, ensuring that every pair of options can be compared.

Transitivity: If option A is preferred to B and B is preferred to C, then A will be preferred to C when aggregating preferences, assuming individual preferences are transitive.

Pareto Efficiency: If every voter prefers option A over option B, then B cannot be chosen over A without violating Pareto efficiency.

Independence of Irrelevant Alternatives (IIA): The ranking of A over B remains unchanged if a third option, C, is introduced or removed, as the preference is solely based on the pairwise comparison of A and B.

How it Works:

In the Borda count, each voter ranks all the alternatives. Points are assigned based on the rank given (e.g., if there are 4 options, a first-place vote gets 3 points, second place gets 2 points, and so on). The alternative with the highest total score wins. This method aggregates preferences while adhering to the mentioned criteria.

Summary:

The Borda count effectively aggregates preferences while satisfying completeness, transitivity, Pareto efficiency, and IIA, making it a robust choice for preference aggregation.

Solution for Exercise 6.2.

Pareto Efficiency Counter-Example:

$$a \succ b \succ c$$

$$a \succ b \succ c$$

$$a \succ b \succ c$$

$$a \succ^* b \sim^* c$$

IIA Counter-Example:

$$a \succ b \succ c$$

$$b \succ a \succ c$$

$$c \succ a \succ b$$

$$a \sim^* b \sim^* c$$

Let's change the preferences but keep the relationship between a and b for each person.

$$a \succ b \succ c$$

$$b \succ a \succ c$$

$$a \succ b \succ c$$

$$a \succ^* b \succ^* c$$

Solution for Exercise 6.3.

Pareto Efficiency Counter-Example:

$$a \succ b \succ c$$

$$a \succ b \succ c$$

$$a \succ b \succ c$$

$$a \sim^* b \succ^* c$$

IIA Counter-Example:

$$a \succ b \succ c$$

$$b \succ a \succ c$$

$$c \succ a \succ b$$

$$a \succ^* b \succ^* c$$

$$a \succ b \succ c$$

$$b \succ c \succ a$$

$$c \succ a \succ b$$

$$a \sim^* b \sim^* c$$

Solution for Exercise 6.4.

The alphabetic rule orders the outcomes in terms of alphabetical order regardless of individual preferences. It is complete, transitive, and respects IIA.

17.7 Solutions for Chapter 7

Solution for Exercise 7.1.

a or b

Solution for Exercise 7.2.

b

Solution for Exercise 7.3.

b

17.8 Solutions for Chapter 8

Solution for Exercise 8.1.

Imagine that we had to make a collective decision over some options- like where to go for lunch. Every individual has their own opinion and preferences about where they want to go. It would be nice if we could come up with some formal way to take individual preferences and apply a rule that will result in what we should choose.

What would we want that rule to do? Well, it should always pick *something*. It should never pick something if everyone unanimously agree there is some other option that is better. And it should never make weird choices- like it picks some option a out of a, b, c but picks option b out of a, b .

Arrow's impossibility theorem says that there is no rule that does this except to just have one of the people always decide the outcome. That's called a dictatorship. This result tells us any time we want to use a formal procedure for making choices in a society, we *have* to be ok with it having some flaw. There is no perfect system.

Solution for Exercise 8.2.

Any Condorcet paradox example like:

$$a \succ b \succ c$$

$b \succ c \succ a$
 $c \succ a \succ b$

Solution for Exercise 8.3.

The same example will work:

$a \succ b \succ c$
 $b \succ c \succ a$
 $c \succ a \succ b$

Solution for Exercise 8.4.

IIA Counter-Example:

$a \succ b \succ c$

$b \succ a \succ c$

$a \succ b \succ c$

a is the social choice.

$a \succ b \succ c$

$b \succ a \succ c$

$c \succ a \succ b$

a, b or c are the social choices.

Note that nothing about preferences over a and b have changed from the first to second set of preferences. In the first a was chosen and b was not. Yet, in the second, b is also chosen. This violates IIA.

Solution for Exercise 8.5.

IIA Counter-Example:

$a \succ b \succ c$

$b \succ a \succ c$

$a \succ b \succ c$

a is the social choice.

$a \succ b \succ c$

$b \succ c \succ a$

$a \succ b \succ c$

a or b are the social choices.

Note that nothing about preferences over a and b have changed from the first to second set of preferences. In the first a was chosen and b was not. Yet, in the second, b is also chosen. This violates IIA.

Solution for Exercise 8.6.

$a \succ b \succ c$

$a \succ b \succ c$

$a \succ b \succ c$

The social choice is a or b and yet a is strictly preferred by everyone to b . b cannot be chosen by a Pareto efficient social choice function.

17.9 Solutions for Chapter 9

Solution for Exercise 9.1.

In the US voting system, there are many third party candidates. However, given how the system works, there is very little incentive for people who support those candidates to vote for them. They are very unlikely to win an election. Because of that, people who support those candidates will usually vote for one of the two major party candidates instead. They act as-if one of those candidates are their favorite. This is an instance of “strategic” voting. The result is that there is less apparent support for third-parties than there might be in a system where such strategic voting was not common.

Could we come up with a system that eliminates strategic voting and implement a system where voters would have incentive to reveal their true preferences over candidates or other options? The Gibbard-Satterthwaite theorem says **no** unless you are willing to accept a system that is either a dictatorship (one person decided the outcome) or in which there is a possibility of an outcome being chosen that is unanimously worse than some other option.

Solution for Exercise 9.2.

Suppose true preferences are:

$a \succ b \succ c$

$a \succ b \succ c$

$c \succ b \succ a$
 $c \succ b \succ a$
 $b \succ c \succ a$

Under plurality vote with the alphabetic tie-breaking rule, a is the social choice. However, if person 5 instead says their preferences are $c \succ b \succ a$ then c wins and they like this better.

Solution for Exercise 9.3.

Warning. This one is a little tougher to come up with a counter-example for than others we have looked at.

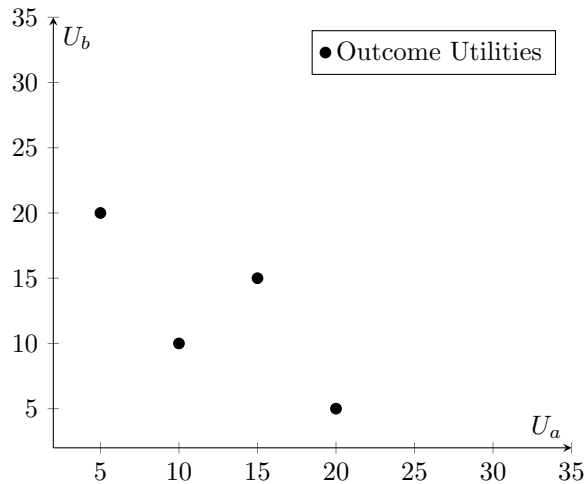
Here is one possibility. Suppose true preferences are:

$c \succ a \succ b$
 $a \succ b \succ c$
 $b \succ c \succ a$

Since all three outcomes get a score of 6 the tie is broken and a is the social choice. However, if person 3 says their preferences are $c \succ b \succ a$ then c is the social choice since the scores are now $a : 6, b : 5, c : 7$. Since person 3 prefers c to a they have incentive to manipulate their preferences this way.

17.10 Solutions for Chapter 10

Solution for Exercise 10.1.



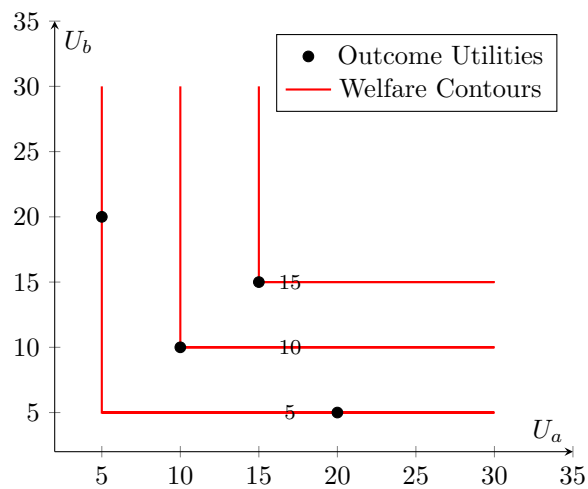
Solution for Exercise 10.2.

$W(ab) = 15, W(a) = \frac{25}{2}, W(b) = \frac{25}{2}, W(n) = 10$. ab maximizes Utilitarian welfare.

Solution for Exercise 10.3.

$W(ab) = 15, W(a) = 5, W(b) = 5, W(n) = 10$. ab maximizes Rawlsian welfare.

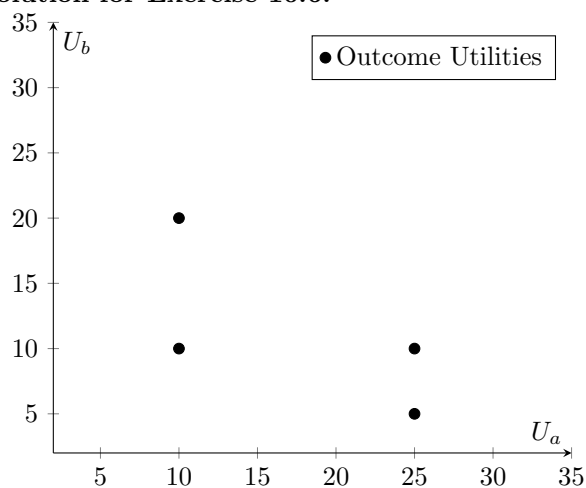
Solution for Exercise 10.4.



Solution for Exercise 10.5.

$W(ab) = 15, W(a) = 10, W(b) = 10, W(n) = 10$. ab maximizes Nash welfare.

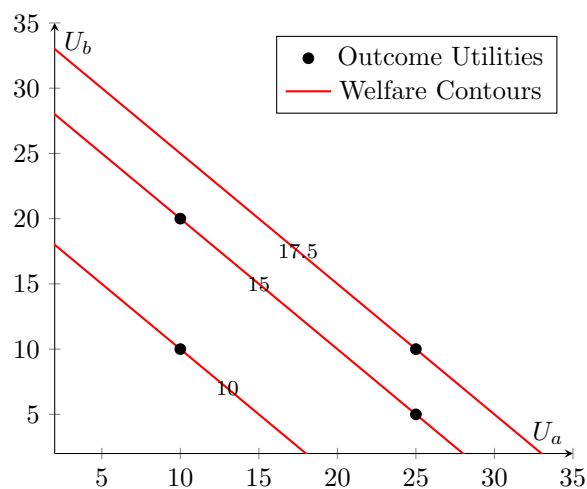
Solution for Exercise 10.6.



Solution for Exercise 10.7.

$W(yes/stay) = 15, W(yes/leave) = \frac{35}{2}, W(no/stay) = 15, W(no/leave) = 10$. Yes/leave maximizes utilitarian welfare.

Solution for Exercise 10.8.



Solution for Exercise 10.9.

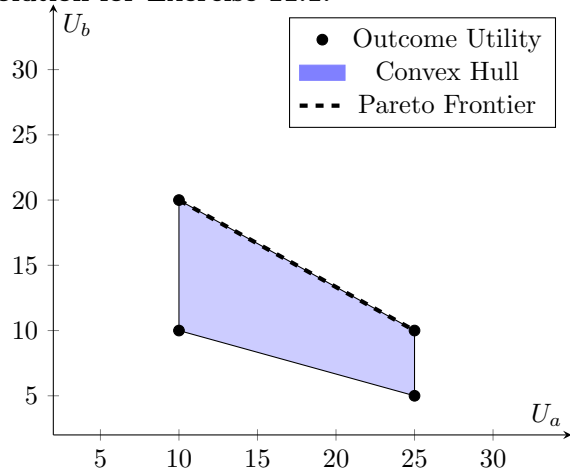
$W(\text{yes/stay}) = 5, W(\text{yes/leave}) = 10, W(\text{no/stay}) = 10, W(\text{no/leave}) = 10$. Yes/leave, No/Stay, No/Leave all maximizes Rawlsian welfare.

Solution for Exercise 10.10.

$W(\text{yes/stay}) \approx 11.1803, W(\text{yes/leave}) \approx 15.8114, W(\text{no/stay}) \approx 14.1421, W(\text{no/leave}) = 10$. Yes/leave maximizes Nash welfare.

17.11 Solutions for Chapter 11

Solution for Exercise 11.1.



Solution for Exercise 11.2.

yes/leave maximizes utilitarian, yes/leave, no/stay, no/leave all maximize rawlsian welfare, yes/leave maximizes Nash welfare.

Solution for Exercise 11.3.

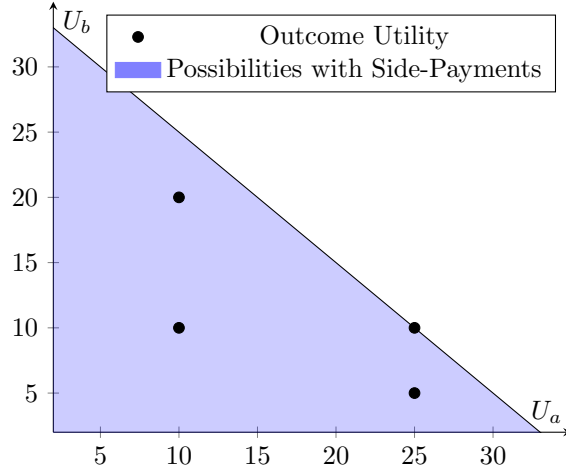
$\frac{1}{3}(10, 20) + \frac{2}{3}(25, 10)$ which is the point $(20, \frac{40}{3})$. This is achieved by picking no/stay with probability $\frac{1}{3}$ and yes/leave with probability $\frac{2}{3}$.

Solution for Exercise 11.4.

(16, 16). This is achieved by picking no/stay with probability $\frac{3}{5}$ and yes/leave with probability $\frac{2}{5}$.

Solution for Exercise 11.5.

(25, 10). This is achieved by the non-randomized outcome yes/leave.

Solution for Exercise 11.6.**Solution for Exercise 11.7.**

(17.5, 17.5) can be achieved by selecting yes/leave and having Alice pay Bob 7.5.

Solution for Exercise 11.8.

a maximizes utilitarian welfare with utilities (30, 0)

Solution for Exercise 11.9.

Choosing each outcome with probability $\frac{1}{2}$ maximizes Nash welfare and yields expected utilities (15, 10)

Solution for Exercise 11.10.

Choosing a with probability $\frac{2}{5}$ maximizes Rawlsian welfare and yields expected utilities (12, 12)

Solution for Exercise 11.11.

Choosing outcome a and having person 1 pay person 2 15 yields (15, 15) which maximizes all the welfare functions among all outcomes achievable with side-payments.

Solution for Exercise 11.12.

b maximizes utilitarian welfare with utilities (0, 8)

Solution for Exercise 11.13.

Choosing each outcome with probability $\frac{1}{2}$ maximizes Nash welfare and yields expected utilities (2.5, 4)

Solution for Exercise 11.14.

Choosing a with probability $\frac{8}{13}$ maximizes Rawlsian welfare and yields expected utilities $(\frac{40}{13}, \frac{40}{13})$

Solution for Exercise 11.15.

Choosing outcome b and having person 2 pay person 1 4 yields (4, 4) which maximizes all the welfare functions among all outcomes achievable with side-payments.

17.12 Solutions for Chapter 12

Solution for Exercise 12.1.

For example:

s_1, s_2	<i>movie</i>	<i>home</i>
<i>movie</i>	(20, 20)	(0, 0)
<i>home</i>	(0, 0)	(10, 10)

The Nash equilibria are $(movie, movie), (home, home)$

Solution for Exercise 12.2.

$(l, l), (n, l), (l, n)$

Solution for Exercise 12.3.

$B_1(l) = l, B_1(n) = l$
 $B_2(l) = l, B_2(n) = l$

Solution for Exercise 12.4.

(l, l)

Solution for Exercise 12.5.

$(n, n), (c, n), (n, c)$

Solution for Exercise 12.6.

$B_a(c) = n, B_a(n) = n, B_b(c) = n, B_b(n) = n$

Solution for Exercise 12.7.

(n, n)

Solution for Exercise 12.8.

$(i, i), (i, n), (n, i)$

Solution for Exercise 12.9.

$B_1(n) = i, B_1(i) = i$
 $B_2(n) = i, B_2(i) = i$

17.13 Solutions for Chapter 13

Solution for Exercise 13.1.

$$\frac{\partial (100 \ln(g_i + g_{-i}) - g_i)}{\partial g_i} = 0$$

$$\frac{100}{g_{-i} + g_i} - 1 = 0$$

$$g_i = 100 - g_{-i}$$

Solution for Exercise 13.2.

100 and 100

Solution for Exercise 13.3.

$$\frac{\sum_{i=1}^n (100 \ln(g_i + g_{-i}) - g_i)}{n} = \frac{n100 \ln(g) - g}{n} = 100 \ln(g) - \frac{g}{n}$$

Maximize this with respect to g :

$$\frac{\partial (100 \ln(g) - \frac{g}{n})}{\partial g} = 0$$

$$g \rightarrow 100n$$

For $n = 10$:

$$g = 1000$$

Solution for Exercise 13.4.

Using the work above, $g = 100n$

17.14 Solutions for Chapter 14**Solution for Exercise 14.1.**

Finding the best response for each person:

$$\frac{\partial (a_i \ln(g_i + g_{-i}) - g_i)}{\partial g_i} = 0$$

$$g_i = a_i - g_{-i}$$

Thus, the individual ideal total contributions are: 100, 200, 500 respectively.

Solution for Exercise 14.2.

Utilitarian welfare is:

$$\frac{(500 + 200 + 100) \ln(g) - g}{3}$$

Maximizing this with respect to g :

$$\frac{\partial (\frac{800}{3} \ln(g) - \frac{g}{3})}{\partial g} = 0$$

$$g = 800$$

This requires $t = \frac{800}{3}$

Solution for Exercise 14.3.

Person i 's utility under tax t is:

$$a_i \log(3t) - t$$

Maximize this with respect to t :

$$\frac{\partial (a_i \log(3t) - t)}{\partial t} = 0$$

$$t = a_i$$

Thus, the individually ideal taxes are 500, 200, 100 respectively.

Solution for Exercise 14.4.

200

17.15 Solutions for Chapter 15**Solution for Exercise 15.1.**

Since $v_i \geq \frac{600}{3} = 200$ for each person, they all agree purchase the printer and each person pays 200. This is the efficient outcome.

Solution for Exercise 15.2.

They purchase the printer under the median mechanism since the median person's valuation $250 \geq 200$. Each pays 200. This is the efficient outcome.

They do not purchase the printer under the unanimous mechanism since $v_c < 200$. This is not the efficient outcome.

Solution for Exercise 15.3.

Since the sum of the valuations is greater than 1000, the VCG mechanism chooses to provide the public good.

Everyone is pivotal.

$$t_a = 1000 - 650 = 350$$

$$t_b = 1000 - 700 = 300$$

$$t_c = 1000 - 750 = 250$$

There is a budget deficit.

Solution for Exercise 15.4.

The median mechanism chooses for them to build the garden since the median valuation $500 \geq \frac{2000}{5}$. Each person pays 400. This is not the efficient outcome since the sum of the valuations is less than 2000.

The unanimous mechanism chooses for them to *not* build the garden since at least one person has a valuation less than $\frac{2000}{5}$. This is the efficient outcome.

Solution for Exercise 15.5.

It chooses for them to purchase the online subscription since the sum of valuations is greater than the cost. a and b are pivotal and pay $t_a = 150 - (75 + 60) = 15$ and $t_b = 150 - (80 + 60) = 10$. There is a budget deficit.

Solution for Exercise 15.6.

Let's suppose some roommates are trying to decide whether to buy a new TV. The TV costs 1000 but each person has a different maximum they are willing to pay (their valuation) for the TV. No one knows the other's valuations. If the sum of their valuations is more than the cost of the TV they should buy it, otherwise they should not (this is called the "efficient" outcome. A "mechanism" is something that allows each individual to say what the TV is worth to them, then the mechanism decides whether it should be purchased and how much each person pays. The Green-Laffont impossibility says that there is no mechanism that makes it so everyone always has incentive to tell the truth, the mechanism always picks the efficient outcome, and always ensures the amount each person pays sum to the cost of the TV if it is supposed to be purchased.

Part VI

Appendix

Part VII

Appendix

A Optimization

A.1 Unconstrained One-Dimensional Optimization

Optimization involves finding the minimum or maximum of a function $f(x)$. Here, we focus on instances where f is one-dimensional. The goal is to determine the value of x that maximizes (or minimizes) $f(x)$. *Unconstrained* means that we will not place any restrictions on what x can be.

Imagine that you are hiking on a mountain trail. If the slope of the trail is positive, then moving forward will bring you to a higher point. If the slope of the trail is negative, then moving *backward* will bring you to a higher point. Thus, **the slope must be zero at the peak**. This is demonstrated in Figure [Figure A.1](#).

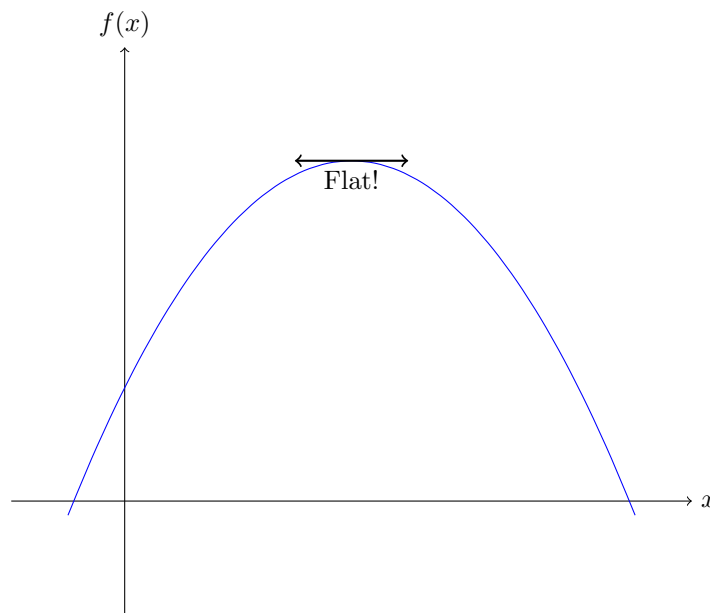


Figure A.1: Slope is Zero at the Peak

One issue with using this fact to find a maximum is that the slope can also be zero at a minimum and also at places that are “local” maxima.

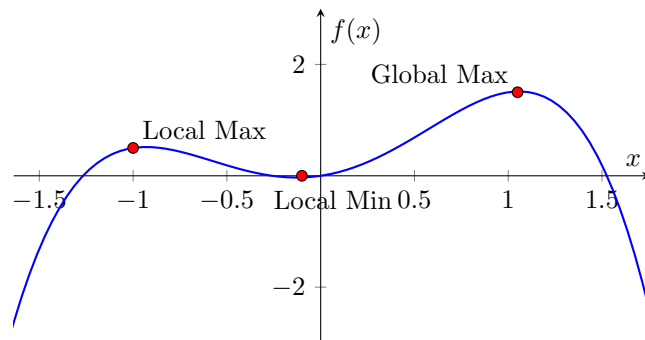


Figure A.2: Not every point of zero slope is a global maximum!

To account for this, we should remember that when we find a point of zero slope, it is only a *candidate* for a global maximum.

Info A.1: Unconstrained Optimization. How to find the unconstrained maximum of a one-dimensional function: For a function $f(x)$:

1. Find the first derivative $f'(x)$.
2. Set the first derivative to zero: $f'(x) = 0$.
3. Solve for x . These are your candidates.
4. Which, if any is a global max?

A.2 Unconstrained Multi-Dimensional Optimization

The intuition of the slope being zero at the maximum holds even when there are multiple directions in which you can move. Imagine trying to find the peak of a mountain when you are not on a trail. You can move east/west or north/south. In fact, you can also move in combinations of these directions, like the northwest. But at the peak, you better not be able to move east/west and get to a higher altitude. **The slope has to be zero in the east/west direction.** Similarly, **the slope has to be zero in the north/south direction.** One of the nice things about *smooth* functions is that if the slope is zero in these two cardinal directions, it will be zero even if you try to move northwest, or southeast, or any other direction. **Figure A.3** demonstrates this. Notice that at the peak, the slope is zero in both the x direction and the y direction, and also in all other directions.

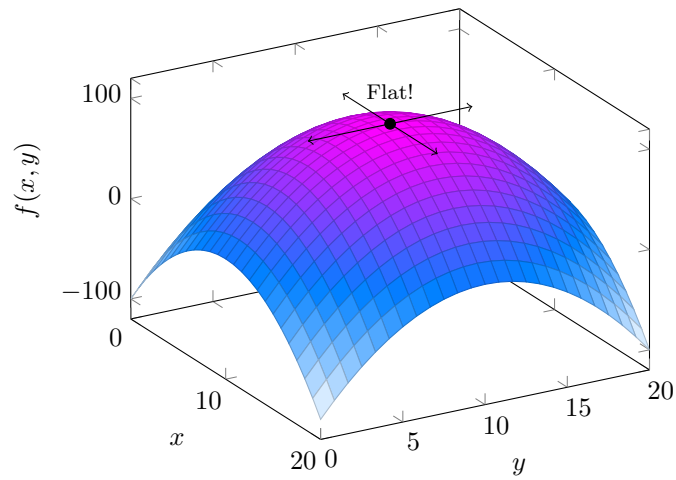
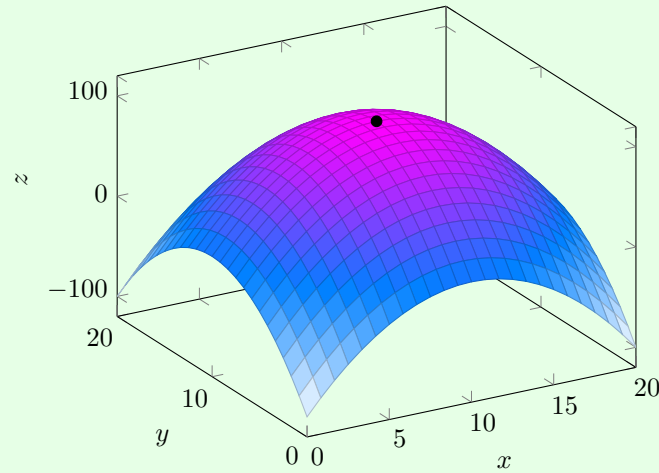


Figure A.3: Slope is Zero in All Directions!

Info A.2: Unconstrained Multi-Dimensional Optimization. To maximize a function $f(\mathbf{x})$ where $\mathbf{x} = (x_1, x_2, \dots, x_n)$:

1. Find all partial derivatives $\left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right)$.
2. Set all partial derivatives to zero: $\frac{\partial f}{\partial x_i} = 0$.
3. Solve the resulting system of equations for (x_1, \dots, x_n) . These solutions are your candidates.
4. Determine which, if any, of these solutions is a global maximum.

Example A.1: Unconstrained Maximum. Maximize $100 - (x - 10)^2 - (y - 10)^2$. Let's look at this function first. The global maximum (black dot) occurs where $x = 10$ and $y = 10$.



Now, we confirm that this is the maximum formally using the procedure in A.2. The partial derivatives are $\frac{f(x,y)}{x} = -2(x - 10)$ and $\frac{f(x,y)}{y} = -2(y - 10)$. Setting these to zero, we get the equations:

$$\frac{f(x,y)}{x} = -2(x - 10) = 0$$

$$\frac{f(x,y)}{y} = -2(y - 10) = 0$$

Solving these gives us the (x, y) where the function has zero slope. The only solution is $x = 10, y = 10$.

We can see by inspecting the function that this must be the global maximum.

A.3 Constrained Multi-Dimensional Optimization

Suppose that we want to maximize a function $f(x, y)$ (**the objective**) but where the set of x and y we can choose from is constrained in some way (**the constraint**).

Let's have a look at how adding a constrained complicates [Example A.1](#).

Example A.2: Constrained Maximum. Maximize $100 - (x - 10)^2 - (y - 10)^2$ subject to $x + y \leq 10$.

Let's look at this function first. As we found in [Example A.1](#), the global maximum (black dot) occurs where $x = 10$ and $y = 10$, but that violates the constraint since $10 + 10 > 10$. We are not allowed to go past the red line. The maximum within that area occurs at $x = 5$ and $y = 5$ (green dot).

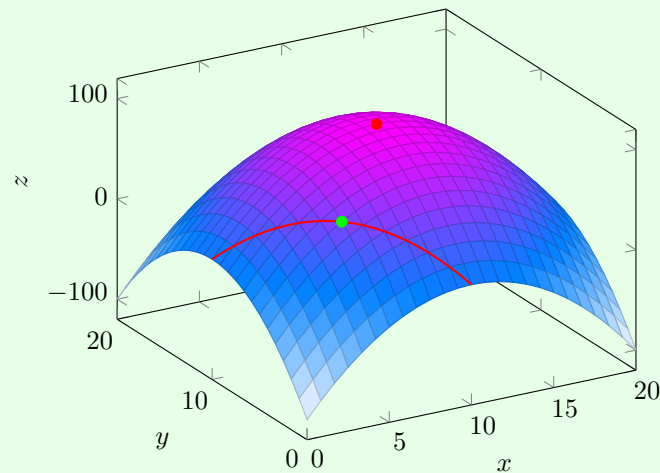
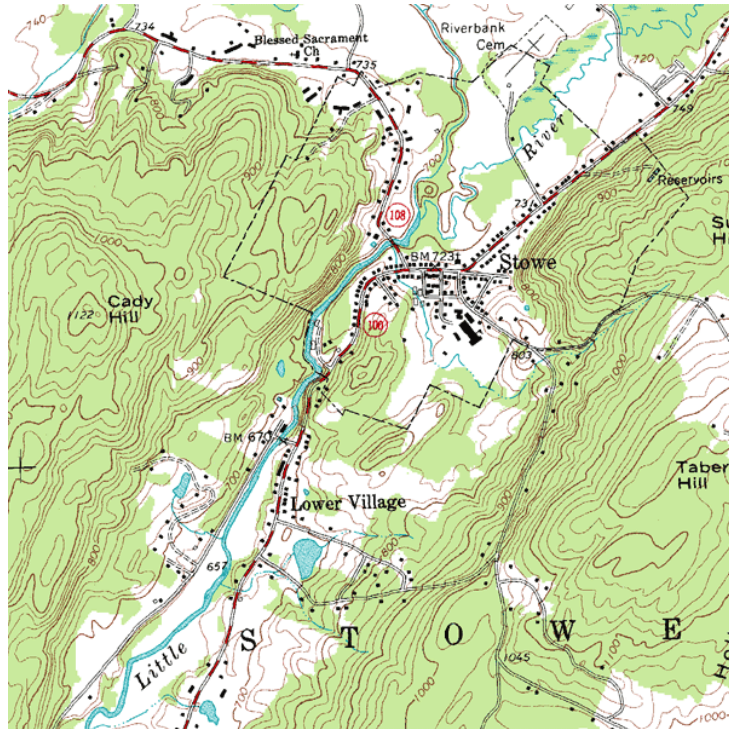


Figure A.4: 3d Plot of $100 - (x - 10)^2 - (y - 10)^2$

How should we formalize the process of finding the constrained optimum? Let's work through a few concepts and return to this example later in the chapter.

A.4 Contours

It can be very useful to think of three-dimensional plots in terms of their contours. [Figure A.5](#) shows a real-world example of how contours are used on a topographic map, which is a 2d map that includes information about elevation through contour lines. Look at the line labeled “1000” near Cady Hill. This is a line connecting places that all have an elevation of 1000 feet.



Taken from the public domain USGS Digital Raster Graphic file o44072d6.tif for the Stowe, VT quadrangle.

Figure A.5: A topographic map of Stowe Mountain.

Let's add some contours to our function at an “elevation” of 25, 50, 75, and 99 (right near the peak). In the context of mathematics, this is known as a “contour” plot.

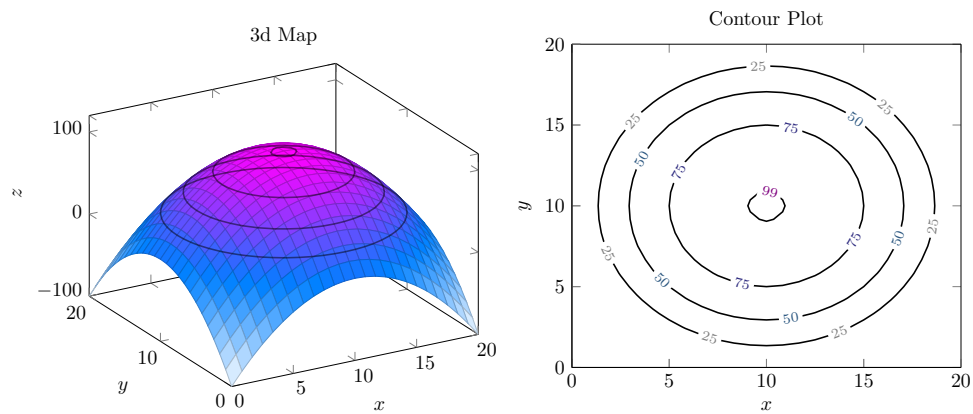


Figure A.6: A function and its contours.

A.5 Monotonicity

Imagine standing at the point $(0,0)$ on the function plotted in [Figure A.6](#). If you walk in the northwest direction (increasing x and y) the function increases. That is, you are increasing in elevation. In fact, this is true whenever $0 \leq x \leq 10$ and $0 \leq y \leq 10$. [Figure A.7](#) shows a plot of the function in [Example A.2](#) limited to this region. Notice how the function always slopes up when moving in the northwest direction regardless of where you are.

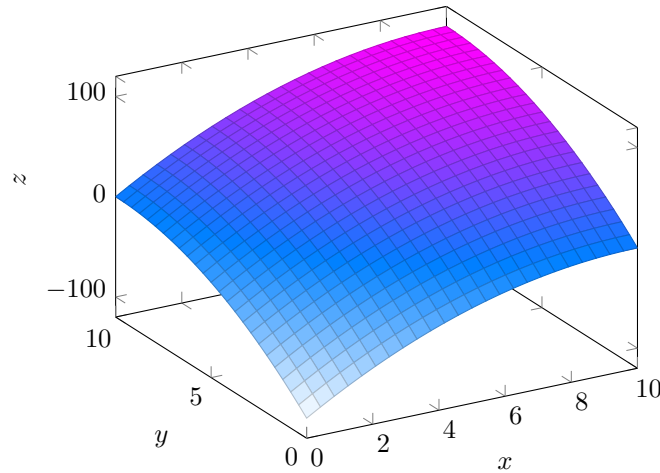


Figure A.7: 3d Plot of $100 - (x - 10)^2 - (y - 10)^2$ where $x \leq 10$ and $y \leq 10$.

When a function increases when you increase all of its variables, we say that it is monotonic. As we will see, this property comes in handy.

Definition A.1: Monotone. $f(x, y)$ is said to be **monotone** when:

1. $x' \geq x$ and $y' \geq y$ implies $f(x', y') \geq f(x, y)$.
2. $x' > x$ and $y' > y$ implies $f(x', y') > f(x, y)$.

A.6 Three Possibilities for an Optimal Point

Let's continue looking at [Example A.2](#). Let's look at the contour plot where $x \leq 10$ and $y \leq 10$ and add the line $x + y = 10$. This is shown in [Figure A.8](#).

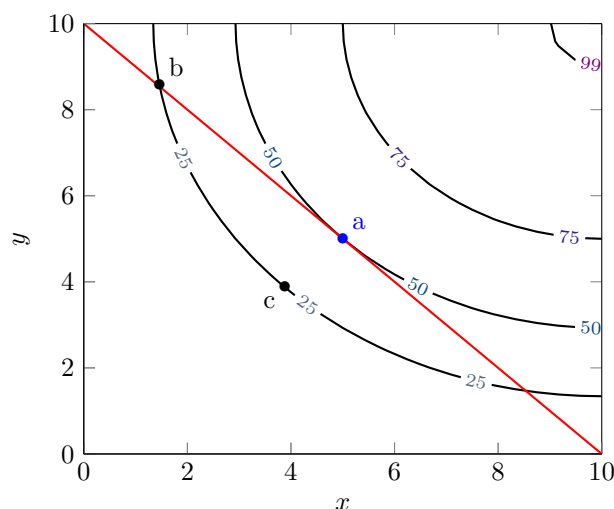


Figure A.8: Contour plot with constraint.

Here, the constraint is the area southwest of the red line. The red line is the "boundary" of the constraint, and the area to the south-west of that line is called the interior of the constraint. For example, the points b and c are on the boundary of the constraint and c is on the interior.

First, notice that point c could never be optimal. Why? If we are on the interior, we could always move up and to the right a little (increasing both x and y) and still meet the constraint. *Since the function is monotonic*, the result **must be better!** Here, for example, we could move from c to a . Thus, we can see that if a function is monotonic, the optimal point cannot be on the interior of the constraint. But it has to meet the constraint. Thus, **when a function is monotonic, the optimal point must be on the boundary of the constraint.**

The point b is on the boundary. Can it be optimal? No, it is on the same contour as c . Because c cannot be optimal, neither can b . They have the same value. This shows us that a point like b , which is on a contour that passes through the interior of the constraint, can never be optimal.

What we have seen so far is that whatever point is optimal must be on the boundary of the constraint and not on a contour that passes into the interior of the constraint. The only way for this to happen is if the contour at the optimal point *just touches* the constraint. See point c for instance. When the contours are smooth, the only way for this to happen is if the contour and the constraint have the same slope.

There are only three possibilities for an optimal point. These are enumerated below.

Info A.3: Three Possibilities for a Constrained Optimum. When the objective is monotonic, the optimum must meet one of the following three conditions.

1. **(Tangent)** It is at a point where the contour of the objective at that point had the same slope constraint.
2. **(Touching but not Smooth)** The point is a “non-smooth” point on the contour of the objective, but the that point just touches the constraint.
3. **(Boundary)** The point is at one of the boundaries of the constraint.

A.7 Slopes of Contours

Many of the optimization problems we will encounter in this course will be “smooth”. In that case, the first possibility “tangent” from box A.3 is relevant. To find such a tangency point, we need to know how to find the slopes of functions.

In most cases, we will be dealing with functions of just two variables. For instance, lines like $ax_1 + bx_2 = 10$ or non-linear functions like $x_1^2x_2^2 = 10$. How do we find the slope of functions like these at particular points?

For the linear case $ax_1 + bx_2 = 10$, we can put it in the conical form of a line $x_2 = -\frac{a}{b}x_1 + \frac{10}{b}$ then read the slope right off. Here it is $\frac{a}{b}$ and since this is a line, that is the slope at every point on the function. But what about a nonlinear function like $x_1^2x_2^2 = 10$? That slope depends on the point you are considering. Here is a plot.

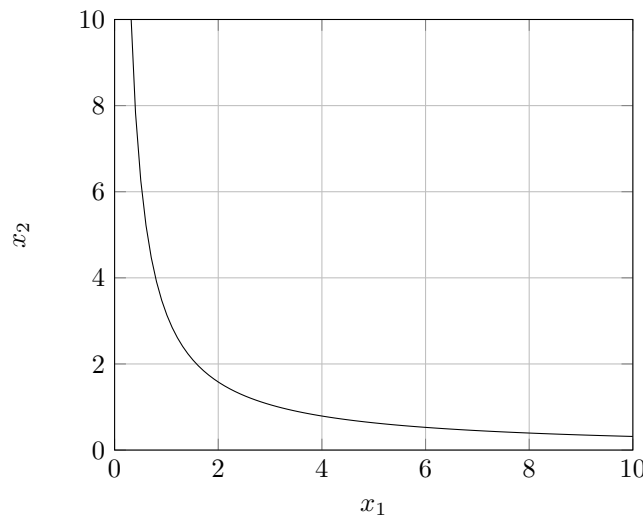


Figure A.9: Plot of $x_1^2x_2^2 = 10$

We can use something called the implicit function theorem to find the slope, but let’s work through it intuitively before I give you the general result.

The slope of a function is really measuring something like this: When you move horizontally a

little, how much do you have to move vertically to get back to the function? For a line with a slope of -1 if you move to the right by one unit, you have to move down by one unit to get back on the function. Look at the plot above. When x_1 is small and x_2 is large, if you move to the right a little, you have to move down a lot to get back to the function. If x_1 is large but x_2 is small, if you move to the right a little, you don't have to move down much at all to get back to the function.

A function like $x_1^2 x_2^2 = 10$ defines a set of points (x_1, x_2) that all meet some condition. Here, the value of the function $f(x_1, x_2) = x_1^2 x_2^2$ in the set of points is equal to 10.

The partial derivative of f with respect to x_1 tells us how much f changes when we increase x_1 a little. This is denoted by $\frac{\partial f}{\partial x_1}$. The partial derivative with respect to x_2 , denoted $\frac{\partial f}{\partial x_2}$ tells us how much f changes when we increase x_2 by a little.

Suppose $\frac{\partial f}{\partial x_1} = 1$ and $\frac{\partial f}{\partial x_2} = 1$. Roughly, if we increase x_1 by a little, $f()$ increases by 1 unit. What do we have to do to get back to the function? We cannot increase x_2 . That will only make $f()$ even bigger. We have to **decrease** x_2 by one unit. The slope is -1 .

Suppose $\frac{\partial f}{\partial x_1} = 1$ and $\frac{\partial f}{\partial x_2} = 2$. Roughly, if we increase x_1 by a little, $f()$ increases by 1 unit. What do we have to do to get back to the function? If we decrease x_2 by one unit, $f()$ will decrease by 2. That's too much! Instead, we decrease it by $\frac{1}{2}$. Then $f()$ will decrease by 1. The slope is $-\frac{1}{2}$.

Finally, suppose $\frac{\partial f}{\partial x_1} = 1$ and $\frac{\partial f}{\partial x_2} = \frac{1}{2}$. Roughly, if we increase x_1 by a little, $f()$ increases by 1 unit. What do we have to do to get back to the function? If we decrease x_2 by one unit, $f()$ will be decreased by $\frac{1}{2}$. That is not enough! Instead, we decrease it by 2. Then $f()$ will decrease by 1. The slope is -2 .

Note that in each case, we find that the slope is the **negative** of the **ratio** of the **partial derivatives**. That is a general result.

Info A.4: Slope of an Implicit Function. The slope of a function $f(x_1, x_2) = y$ at the point (x_1, x_2) is $-\frac{\frac{\partial f(x_1, x_2)}{\partial x_1}}{\frac{\partial f(x_1, x_2)}{\partial x_2}}$.

A.8 Solution - Two Dimensional Constrained Optimization

Now that we know how to find the slope of implicit functions, we have an easy way to find places where the slope of the contour is the same as the slope of the constraint.

Definition A.2: First-Order Condition. For maximizing $f()$ subject to $g() \leq c$, the **first-order condition** is:

$$-\frac{\frac{\partial f(x_1, x_2)}{\partial x_1}}{\frac{\partial f(x_1, x_2)}{\partial x_2}} = -\frac{\frac{\partial g(x_1, x_2)}{\partial x_1}}{\frac{\partial g(x_1, x_2)}{\partial x_2}}$$

Notice that the first-order condition gives us just one equation (the number of variables). But the optimal point has two unknowns. This is not enough for the optimal point. Fortunately, we already know something else about the optimal point. It must occur *on the constraint*.

Info A.5: Solving a Constrained Maximum. To solve a constrained maximum problem:

1. Find $-\frac{\frac{\partial f(x_1, x_2)}{\partial x_1}}{\frac{\partial f(x_1, x_2)}{\partial x_2}}$ and $-\frac{\frac{\partial g(x_1, x_2)}{\partial x_1}}{\frac{\partial g(x_1, x_2)}{\partial x_2}}$
2. Simplify $-\frac{\frac{\partial f(x_1, x_2)}{\partial x_1}}{\frac{\partial f(x_1, x_2)}{\partial x_2}} = -\frac{\frac{\partial g(x_1, x_2)}{\partial x_1}}{\frac{\partial g(x_1, x_2)}{\partial x_2}}$.
3. Plug the result into the constraint to get the solution.

A.9 Gradients

All of the above assumed we are maximizing a two-dimensional function $f(x, y)$. There, we can find a place where the slope of a contour is equal to the slope of the constraint using the formula in **Info Box A.4**. But for functions with more variables like $f(x, y, z)$ it is more convenient to instead calculate the **gradient** of the objective and constraint. What is that?

Definition A.3: Gradient. The gradient of a function is a vector that points in the direction of the *fastest rate of increase of the function*. It is denoted by ∇f and is the vector of partial derivatives of the function:

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right)$$

where f is a function of n variables x_1, x_2, \dots, x_n .

Info A.6: Gradients are Perpendicular to Contour. The gradient of the function is always perpendicular to the contour lines. This is because the gradient points in the direction of the steepest ascent, while the contour lines represent points of zero ascent.

In summary, the gradient ∇f is perpendicular to the contour lines of the function f , and its magnitude indicates the rate of change of the function in the direction of the gradient.

A.10 First Order Condition and Solution

Now that we know that the gradient of a function at some point is always perpendicular to the contour of a function at some point, we have an easy way to find places where the slope of the contour is the same as the slope of the constraint.

Definition A.4: First-Order Condition. For maximizing $f()$ subject to $g() \leq c$, the **first order condition** is:

$$\nabla f(x) = \lambda \nabla g(x) \quad (6)$$

where $\nabla f(x)$ is the gradient of the objective function, $\nabla g(x)$ is the gradient of the constraint function, and λ is just some number.

This is equivalent to the following n equations:

$$\begin{aligned} \frac{\partial f}{\partial x_1} &= \lambda \frac{\partial g}{\partial x_1} \\ \frac{\partial f}{\partial x_2} &= \lambda \frac{\partial g}{\partial x_2} \\ &\dots \\ \frac{\partial f}{\partial x_n} &= \lambda \frac{\partial g}{\partial x_n} \end{aligned} \quad (7)$$

The λ comes from the fact that the two gradients need not have the same *magnitude*, only the same direction! The λ allows their magnitude to differ.

Notice that the first-order condition gives us n equations (the number of variables). But the optimal point has n unknowns. This is not quite enough for the optimal point. Fortunately, we already know something else about the optimal point. It must occur *on the constraint*.

Info A.7: Solving a Constrained Maximum. To solve a constrained maximum problem with n variables.

1. For each variable x_1, x_2, \dots, x_n . Find $\frac{\partial f}{\partial x_i}$ and $\frac{\partial g}{\partial x_i}$.
2. Solve the resulting first-order condition, together with the constraint.

$$\begin{aligned} \frac{\partial f}{\partial x_1} &= \lambda \frac{\partial g}{\partial x_1} \\ \frac{\partial f}{\partial x_2} &= \lambda \frac{\partial g}{\partial x_2} \\ &\dots \\ \frac{\partial f}{\partial x_n} &= \lambda \frac{\partial g}{\partial x_n} \\ g(x_1, x_2, \dots, x_n) &= c \end{aligned} \quad (8)$$

A.11 Lagrange

Notice that first-order condition for *constrained optimization* involves the n equations of the form:

$$\frac{\partial f}{\partial x_i} = \lambda \frac{\partial g}{\partial x_i} \quad (9)$$

... together with the constraint $g(x_1, x_2, \dots, x_n) = c$.

As it turns out, there is always an *unconstrained optimization* problem that has the exact same first-order conditions. That is, the optimization of the following Lagrange function for variables $x_1, x_2, \dots, x_n, \lambda$.

$L = f(x_1, x_2, \dots, x_n) - \lambda(g(x_1, x_2, \dots, x_n) - c)$. Recall above that the first-order condition for an unconstrained problem is just that all the first derivatives have to be zero.

Taking those first , we get something familiar:

$$\begin{aligned}\frac{\partial f}{\partial x_1} - \lambda \frac{\partial g}{\partial x_1} &= 0 \\ \frac{\partial f}{\partial x_2} - \lambda \frac{\partial g}{\partial x_2} &= 0 \\ &\dots \\ \frac{\partial f}{\partial x_n} - \lambda \frac{\partial g}{\partial x_n} &= 0 \\ g(x_1, x_2, \dots, x_n) - c &= 0\end{aligned}\tag{10}$$

These can be rearranged exactly to the first-order conditions of the constrained optimization problem.

A.12 Examples

Let's work through [Example A.2](#) using the Lagrange method.

Example A.3: Solution to [Example A.2](#).

We have our objective: $100 - (x - 10)^2 - (y - 10)^2$ and our constraint $x + y \leq 10$.

We begin by turning this constrained optimization into the unconstrained optimization problem of the Lagrange function:

$$L(x, y) = 100 - (x - 10)^2 - (y - 10)^2 - \lambda(x + y - 10)$$

Now we find the first-order conditions of this unconstrained problem with respect to the three variables x, y, λ . These are:

$$-2(x - 10) - \lambda = 0$$

$$-2(y - 10) - \lambda = 0$$

$$x + y - c = 0$$

We can solve these three equations by first eliminating λ from the first two equations:

$$-2(y - 10) = -2(x - 10)$$

$$x = y$$

Now we plug this into the third (constraint) equation $x + y = 10$ to get:

$$x = 5$$

$$y = 5$$

A.13 Exercises

Assume for each of the following problems that $x \geq 0$ and $y \geq 0$.

Exercise 13.1: Maximize the function $f(x) = -x^2 + 4x + 4$.

Exercise 13.2: Maximize the function $f(x) = \ln(x) - \frac{1}{4}x + 4$.

Exercise 13.3: Maximize the function $f(x, y) = -x^2 - y^2 + 2x + 2y$.

Exercise 13.4: Maximize the function $f(x, y) = x + y$ subject to the constraint $x + 2y \leq 60$.

Exercise 13.5: Maximize $f(x, y) = xy$ subject to the constraint $x + 2y \leq 60$.

Exercise 13.6: Maximize the function $f(x, y) = x^{\frac{1}{2}} + y^{\frac{1}{2}}$ subject to the constraint $x + 2y \leq 60$.

Exercise 13.7: Maximize the function $f(x, y) = \min\{x, y\}$ subject to the constraint $x + 2y \leq 60$.

Exercise 13.8: Solve the constrained maximization problem in [Example A.2](#) but change the constraint to $x + 2y \leq 10$

A More Preference Aggregation Rules

A.1 Plurality Vote

Plurality vote focuses on the goal of giving as many people as possible their top-ranked outcome. Because of this, it throws away most of the information about preferences and just focuses on the top of each individuals' ranking.

Definition A.1: Plurality Vote. The **score** of an outcome is the number of people who rank that outcome highest. Social preferences are determined by score.

Example A.1: Example 1.

- 1: $a \succ b \succ c$
- 2: $a \succ c \succ b$
- 3: $c \succ a \succ b$

Scores. $a : 2, b : 0, c : 1$

$$a \succ^* c \succ^* b$$

Example A.2: Example 2.

- 1: $a \succ c \succ b$
- 2: $a \succ c \succ b$
- 3: $b \succ c \succ a$
- 4: $b \succ a \succ c$
- 5: $c \succ a \succ b$

Scores. $a : 2, b : 2, c : 1$

$$a \sim^* b \succ^* c$$

A.2 Veto

While plurality vote attempts to maximize the number of people who get their favorite outcome, this method attempts to do the opposite: minimizing the number of people who get their least favorite outcome.

Definition A.2: Veto. The score of an outcome is the negative of the number of people who rank it last. Social preferences are determined by the score as in the other scoring methods above with a higher score being ranked higher.

Example A.3: Example 1.

- 1: $a \succ b \succ c$
- 2: $a \succ c \succ b$
- 3: $c \succ a \succ b$

Scores. $a : 0, b : -2, c : -1$

$$a \succ^* c \succ^* b$$

Example A.4: Example 2.

- 1: $a \succ c \succ b$
- 2: $a \succ c \succ b$
- 3: $b \succ c \succ a$
- 4: $b \succ a \succ c$
- 5: $c \succ a \succ b$

Scores. $a : -1, b : -3, c : -1$

$$a \sim^* c \succ^* b$$

References

- [1] Jeremy Bentham. *An Introduction to the Principles of Morals and Legislation*. 1789.
- [2] Vincenzo Denicolò. Fixed agenda social choice theory: Correspondence and impossibility theorems for social choice correspondences and social decision functions. *Journal of Economic Theory*, 59(2):324–332, 1993.
- [3] John Stuart Mill. *Utilitarianism*. 1863.
- [4] John Rawls. *A Theory of Justice*. Harvard University Press, 1971.
- [5] Kota Saito. Social preferences under risk: Equality of opportunity versus equality of outcome. *American Economic Review*, 103(7):3084–3101, 2013.