Economics 1011a Section Notes

Kevin He

January 1, 2015

This file collects together all the section notes I wrote as a Teaching Fellow for an undergraduate micro-economic theory course (Economics 1011a) in Fall 2014. These notes are **not** self-contained, as they are meant to accompany weekly lectures and the textbook. For instance, the notes provide very few numerical examples, which are already in ample supply in these other sources. Instead, I try to emphasize the "theory" in "intermediate microeconomic theory".

Some of the material in these notes borrow from previous Teaching Fellows and many images are taken from the Internet. Specific sources and acknowledgements appear on the first page of each week's notes. I am very grateful for the work of these individuals.

Contents

(1) Partial differentiation; **(2)** Total differentiation; **(3)** Implicit functions; **(4)** Optimization

TF: Kevin He (he02@fas.harvard.edu)

1 Partial differentiation

1.1 *What is a partial derivative?* Let's think back to single-variable calculus for a moment. For a function $g(t)$, the derivative $g'(t)$ describes the rate at which the output changes when the input changes. In multivariate calculus, when a function $f(x,y)$ depends on **several** inputs, we are often interested in the rate at which **each** of these inputs affects output. Partial differentiation allows us to disentangle the effects of these different inputs. Formally, starting at some point (a,b), define the partial derivative of f with respect to x at (a,b) to be:

$$\boxed{\frac{\partial f}{\partial x}(a,b) := \lim_{\Delta x \to 0} \frac{f(a+\Delta x, b) - f(a,b)}{\Delta x}} \tag{1}$$

In other words, although f **actually** depends on the **pair** of inputs (x,y), in computing $\frac{\partial f}{\partial x}$ we treat y as a fixed constant. This way, we can interpret $\frac{\partial f}{\partial x}$ as the marginal contribution of **just the first input** to the output. Indeed, for small enough Δx, we have the approximation:

$$f(a+\Delta x, b) \approx f(a,b) + \frac{\partial f}{\partial x}(a,b) \cdot \Delta x$$

Of course, we could have done the above exercise for y holding x fixed, with analogous interpretations.

1.2 *The geometric interpretation of partial derivative.* Back in single-variable calculus, we could read off the value of $g'(t_0)$ as the **slope** of the **tangent line** to $g(t)$ at t_0. A similar procedure can be performed for partial differentiation. Given function $f(x,y)$ and a point (a,b), consider the curve that f would trace out in $\mathbb{R}^3$ if its second argument were fixed at b while its first argument varies (curve C_1 in Figure 1). Construct the tangent line to this curve at point a (line T_1 in Figure 1). The slope of this tangent line is $\frac{\partial f}{\partial x}(a,b)$.

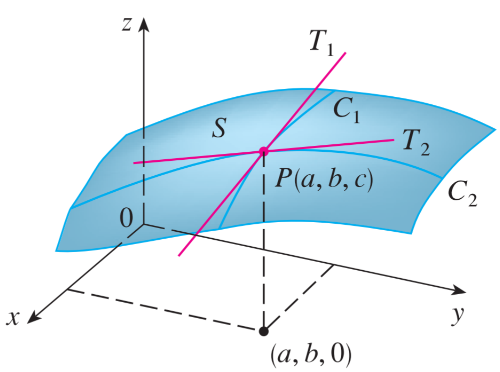


Figure 1: The geometric interpretation of partial derivatives. The curve C_1 traces out f with y fixed at b, while curve C_2 traces out f with x fixed at a. The slopes of the tangent lines to these two curves at (a,b), T_1 and T_2, equal to $\frac{\partial f}{\partial x}(a,b)$ and $\frac{\partial f}{\partial y}(a,b)$, respectively.

[1]This week's section material borrows from the work of previous TFs, in particular the notes of Zhenyu Lai. Image credits: Calculus: Early Transcendentals, Wikimedia Commons.

1.3 *Partial differentiation as a functional operator.* It is important to remember that the partial derivative $\frac{\partial f}{\partial x}(x, y)$ is a **function** defined on the same domain as $f(x, y)$. It is **not a number**, unless evaluated at some specific point $(a, b) \in \mathbb{R}^2$. Let me make this point clearer by viewing the act of partial differentiation as a functional operator.

Most familiar functions take numbers as inputs and return a number as output. But a function could also take a **function as input**, returning another **function as output**. Such functions are called **functional operators**, in that they "operate" on the input function and give back a transformed function[2].

Partial differentiation is a functional operator. Let's write $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$ for partial differentiation against first and second arguments. Take $f(x, y) = x^3y^2$. Then $\frac{\partial}{\partial x}(f)$ is another function, satisfying

$$\left(\frac{\partial}{\partial x}(f)\right)(x, y) = 3x^2y^2$$

Further, since $\frac{\partial}{\partial x}(f)$ is just another function, we can operate on it with $\frac{\partial}{\partial x}$ again, obtaining $\frac{\partial}{\partial x}\left(\frac{\partial}{\partial x}(f)\right)$, which is yet another function – namely,

$$\left(\frac{\partial}{\partial x}\left(\frac{\partial}{\partial x}(f)\right)\right)(x, y) = 6xy^2$$

It can get quite tiresome to write expressions like $\left(\frac{\partial}{\partial x}\left(\frac{\partial}{\partial x}(f)\right)\right)$, so people have developed some shorthands. For example, $\frac{\partial f}{\partial x}$ is just the shorthand for $\frac{\partial}{\partial x}(f)$. Table 1 shows some other common shorthands for partial derivatives. No matter how they are written though, always remember that partial derivatives are functions.

Equivalent notations			
$\frac{\partial}{\partial x}(f)$	$\frac{\partial f}{\partial x}$	f_x	f_1
$\frac{\partial}{\partial x}\left(\frac{\partial}{\partial x}(f)\right)$	$\frac{\partial^2 f}{\partial x^2}$	f_{xx}	f_{11}
$\frac{\partial}{\partial y}\left(\frac{\partial}{\partial x}(f)\right)$	$\frac{\partial^2 f}{\partial y \partial x}$	f_{xy}	f_{12}

Table 1: All notations in each row are equivalent.

One last thing. Using the specific example $f(x, y) = x^3y^2$, you may have noticed that

$$\frac{\partial}{\partial x}\left(\frac{\partial}{\partial y}(f)\right) = \frac{\partial}{\partial y}\left(\frac{\partial}{\partial x}(f)\right)$$

Turns out this is a general result. **Young's theorem** states that, provided f has continuous second derivatives,

$$\boxed{f_{xy} = f_{yx}} \tag{2}$$

For the purposes of this class, assume Young's theorem always holds.

[2]Also called "higher order functions" by computer scientists.

2 Total differentiation

2.1 *What is total differentiation?* Partial differentiation is useful for isolating the effect of one input variable on the output when the other input variables are held fixed. Sometimes, however, there is another variable that determines the values of the input variables. For instance, suppose in the function $f(x, y)$, x and y are actually functions of s. In that case, the output is really determined as $f(x(s), y(s))$.

When all inputs to a multivariate function are themselves functions of a common variable, we may ask how this common variable affects the final output. That is, we may want to compute $\frac{df}{ds}$. This calculation is called **total differentiation**, since we are computing the **total dependency** of f on s without assuming that any of the input variables is held fixed. (Usually, the only time it makes sense to talk about the "total derivative" of a multivariate function is when all of its arguments depend on a single common variable.)

Chain rule allows us to write a total derivative in terms of partial derivatives. Suppose $f : \mathbb{R}^n \to \mathbb{R}$ while each of the arguments to f, namely $x_1, x_2, ..., x_n$, is a function of s. Then **chain rule** implies:

$$\boxed{\frac{df}{ds} = \frac{\partial f}{\partial x_1} \cdot \frac{dx_1}{ds} + \frac{\partial f}{\partial x_2} \cdot \frac{dx_2}{ds} + ... + \frac{\partial f}{\partial x_n} \cdot \frac{dx_n}{ds}} \tag{3}$$

2.2 An example of total differentiation. Suppose utility is a function of income (Y) and leisure (L). Further, each of Y and L is a function of hours worked, H. Specifically, $Y(H) = wH$ (for some positive number w representing hourly wage) while $L(H) = 24 - H$. In that case, the utility function, $U(Y, L)$, is totally dependent on H and might be re-written as $U(Y(H), L(H))$. By above formula, its total derivative with respect to H is:

$$\frac{dU}{dH} = U_Y \cdot \frac{dY}{dH} + U_L \cdot \frac{dL}{dH} = wU_Y - U_L$$

What is the sign of the total derivative? Ha, **trick question**! Remember that the partial derivatives U_Y and U_L are **functions** (of Y, L), so $\frac{dU}{dH}$ is also a function. Its value, hence its sign, depends on values of Y and L, which in turn depend on H. In economic terms, whether working one more hour improves your total utility depends on how many hours you are already working.

3 Implicit functions

3.1 From relation to function, an example. So far, we have dealt with multivariable **functions**, which are mappings from $\mathbb{R}^n$ to $\mathbb{R}$. We next turn to multivariable **relations**, which are equations of the form $R(x_1, ..., x_n) = 0$. Rough speaking, we would like to find some function $h(x_2, ..., x_n)$ so that

$$R(h(x_2, ..., x_n), x_2, ..., x_n) = 0$$

holds for a "rich" set of $(x_2, ..., x_n)$. That is, h is a **function implicit within the relation** $R(x_1, ..., x_n) = 0$, which says what the value of x_1 has to be for every $n - 1$ tuple $(x_2, ..., x_n)$ in order to solve the equation. Said another way, h defines x_1 as a function of the other variables, under the constraint that the relation $R(x_1, ..., x_n) = 0$ must be maintained.

An example might clarify the desired construction. Suppose $n = 2$ and $R(x, y) = x^{\frac{1}{2}} y^{\frac{1}{2}} - 6$. To add economic motivation, view $V(x, y) = x^{\frac{1}{2}} y^{\frac{1}{2}}$ as the utility function of some consumer, so that the (x, y) pairs satisfying the relation $R(x, y) = 0$ are the set of consumption pairs carrying a utility level of exactly 6. We want to find a function $h(y)$ so that $h(y)^{\frac{1}{2}} y^{\frac{1}{2}} - 6 = 0$ for "many" values of y. Simple enough: we rearrange to get $h(y) = \frac{36}{y}$. This is the function "hidden inside" the relation $R(x, y) = 0$, in the sense that $R(h(y), y) = 0$, at least for $y > 0$. We can even compute $\frac{dh}{dy}$, which is the marginal rate of substitution between x and y (how much less x is needed to reach a utility level of 6 if I get an extra unit of y?).

3.2 The "implicit function assumption". Here is the bad news: It is not true that given a relation

$$R(x_1, ..., x_n) = 0$$

and some point $(a_1, ..., a_n) \in \mathbb{R}^n$ where the relation holds, we can necessarily find h so that

$$R(h(x_2, ..., x_n), x_2, ..., x_n) = 0$$

holds in even a small disk around $(a_2, ..., a_n)$. Worse, even when such h exists, it may not be differentiable. There is a result in differential topology that gives sufficient conditions for existence and differentiability of h. But for the purposes of this class, we will instead make the "**implicit function assumption**":

Suppose the relation $R(x_1, ..., x_n) = 0$ is satisfied at some point $(a_1, ..., a_n) \in \mathbb{R}^n$. Then assume there is a differentiable h_1 so that $R(h_1(x_2, ..., x_n), x_2, ..., x_n) = 0$ for all $(x_2, ..., x_n)$ sufficiently close to $(a_2, ..., a_n)$. Analogous functions $h_2, h_3, ..., h_n$ are also assumed to exist.

3.3 Implicit function as a tool for comparative statics. Many of the models you will encounter in this class involve solving **optimization problems**. Once you have found the optimizer (eg. the input value that maximizes the output), you might want to know how the optimizer changes when some parameter of the problem changes. For example, we can view the income and leisure problem from before as an optimization problem over H, the number of hours spent working. One might ask how the utility-maximizing number of working hours changes when hourly wage changes. These kinds of exercises are called **comparative statics**.

Often, we cannot write down an explicit formula for the optimizer. We will instead have some relation characterizing the optimizer, perhaps from first order conditions of optimization. To derive comparative statics, use the implicit function assumption.

Recall the optimizer in the income and leisure problem is characterized by $\frac{dU}{dH} = 0$, i.e.

$$wU_Y - U_L = 0$$

Recall U_Y, U_L are functions of Y, L, which are in turn functions of H. So, above is more formally written as:

$$wU_Y(wH, 24 - H) - U_L(wH, 24 - H) = 0$$

Put $R(w, H) = wU_Y(wH, 24 - H) - U_L(wH, 24 - H)$ and consider the relation $R(w, H) = 0$, which is satisfied at (w_0, H_0) whenever H_0 solves the utility maximization problem under parameter w_0. Implicit function assumption allows us to say there is some differentiable $H^*(w)$ such that $R(w, H^*(w)) = 0$, i.e. the optimizer H^* exists and is differentiable with respect to parameter w. Under this assumption we have:

$$wU_Y(wH^*(w), 24 - H^*(w)) - U_L(wH^*(w), 24 - H^*(w)) = 0$$

Differentiating both sides by $\frac{d}{dw}$ via chain rule (eventually) obtains the desired comparative statics.

4 Optimization

4.1 One variable, no constraints. Suppose we want to find local maximizers for a smooth function $f(x)$. As you might remember, we do this by considering the first order condition (**FOC**). If x^* is a local maximizer of f, then:

$$\boxed{f'(x^*) = 0} \tag{4}$$

If $f'(x^*) \neq 0$, then x^* cannot be a local maximizer, for the intuitive reason that we can find a $\hat{x}$ "sufficiently close to x^*" with $f(\hat{x}) > f(x^*)$.

The FOC is **necessary, but not sufficient**. While the FOC can help cut down possible candidates for local maximizers, $f'(x^*) = 0$ is not a guarantee that x^* is a local maximizer. Sufficient conditions come from the second derivative. Suppose x^* meets the FOC. Then...

$$\boxed{f''(x^*) \begin{cases} < 0 & \text{means local maximizer} \\ = 0 & \text{means } \textbf{no information!} \\ > 0 & \text{means local minimizer} \end{cases}} \tag{5}$$

4.2 Multiple variables, no constraints. Suppose we want to find local maximizers for smooth $f(x_1, ..., x_n)$. The **FOC** is analogous to the single variable case, except it applies to every partial derivative. If $(x_1^*, ..., x_n^*)$ is a local maximizer, then

$$\boxed{f_1(x_1^*, ..., x_n^*) = f_2(x_1^*, ..., x_n^*) = ... = f_n(x_1^*, ..., x_n^*) = 0} \tag{6}$$

As before, the FOC is not a guarantee that $(x_1^*, ..., x_n^*)$ is a local maximizer. Unfortunately, in the multivariate case the sufficient second order condition is complicated, except in two special cases:

$$\boxed{\text{If } f \text{ is concave, then FOC is also sufficient.}} \tag{7}$$

If $n = 2$, a set of sufficient second order conditions is:

$$\boxed{f_{11} < 0 \text{ and } f_{22} < 0 \text{ and } f_{11}f_{22} - (f_{12})^2 > 0, \text{ where all partials are evaluated at } (x_1^*, x_2^*)} \tag{8}$$

4.3 Constrained optimization via Lagrange multiplier. So far we have dealt with global optimization problems, where we were free to pick any point as an input to our function. Another important class of maximization problems puts constraints on allowable inputs. Consider:

$$\max_{x,y} f(x, y) \qquad \text{subject to } g(x, y) = c$$

That is, out of the pairs (x, y) that solve the equation $g(x, y) = c$, we are looking for the one that maximizes $f(x, y)$. Such restriction on the domain of of optimization might arise due to, for example, budget constraints. One way to solve constrained optimization is to explicitly write down the implicit function within the relation $g(x, y) = c$, expressing x as a function of y. This transforms the problem into an unconstrained optimization over one variable. However, if the function g is complicated, then it may be difficult to find the exact expression for the relationship between x and y.

Lagrange multiplier method facilitates solving constrained optimization problems by transforming them into unconstrained optimization problems **in a mechanical way**. For a constrained optimization problem as above, the associated Lagrangian is a three variable function defined as:

$$\boxed{\mathcal{L}(x, y, \lambda) := f(x, y) + \lambda(c - g(x, y))} \tag{9}$$

Now, optimize $\mathcal{L}(x, y, \lambda)$ as an **unconstrained three variable maximization problem**. Under regularity conditions (which we always assume to hold for this class), if (x^*, y^*, λ^*) satisfies FOC of optimization, then (x^*, y^*) is a constrained local maximum in the original problem.

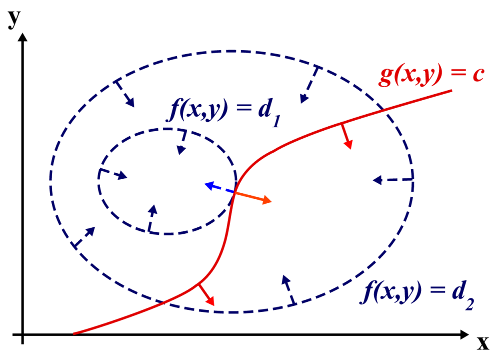


Figure 2: An illustration of constrained maximization problem. The two ellipses show two level curves of the function $f(x, y)$. The constraint is $g(x, y) = c$. The constrained maximum of f must occur at a point of tangency between the level curve $g(x, y) = c$ and some level curve of f.

To understand why this might work, consider the geometry of the constrained optimization problem, illustrated in Figure 2. The ellipses are two level curves for f, while the attached arrows point in the directions of fastest increase for f. Imagine the ellipses as snapshots of an inflating balloon. The balloon starts as a single point at the local maximum of f and expands outwards, reaching ever **decreasing** values of f as it expands. Now suppose we place a **physical barrier** in the shape of $g(x, y) = c$ and inflate the same balloon again. Initially, the values of f are high, but the balloon has not reached the feasible set $g(x, y) = c$. Eventually, the balloon touches the barrier at some point, which is the constrained maximum of f. Further, physical intuition suggests the point of first contact between the balloon and the barrier should be a **point of tangency**. That is, at the constrained maximum, level curves of f and g must share the same "slope" – which is to say $\begin{pmatrix} f_x(x^*, y^*) \\ f_y(x^*, y^*) \end{pmatrix}$ must be a scalar multiple of $\begin{pmatrix} g_x(x^*, y^*) \\ g_y(x^*, y^*) \end{pmatrix}$.

However, the FOC of $\mathcal{L}$ with respect to x and y are:

$$\begin{cases} f_x(x^*, y^*) = \lambda^* g_x(x^*, y^*) \\ f_y(x^*, y^*) = \lambda^* g_y(x^*, y^*) \end{cases}$$

which enforces exactly the said relationship, with λ^* as the scalar. At the same time, the FOC of $\mathcal{L}$ with respect to λ rearranges to

$$g(x^*, y^*) = c$$

which ensures the constraint is satisfied.

Economics 1011a ▷ Section 1 : Modeling and "Verbal Problems"[3] 9/9/2014

(1) Modeling tips; **(2)** 1-variable,1-parameter maximization; **(3)** Firm's problem

TF: Kevin He (he02@fas.harvard.edu)

1 Modeling tips

1.1 *Optimization is king...* The models you write for this class should always feature **optimization**. One central tenet of economics is that economic agents are rational actors who **respond to incentives**. As such, we would like to think that (to a first approximation) consumers, firms, etc., always optimally respond to their economic environment.

1.2 *...and simplicity is queen.* Think about the key economic idea that the question is getting at. It might be useful to say to yourself, "gee, why would Giacomo ask this question?" Once you identify the key idea, write a "**minimum viable model**" to capture it. No bells and whistles! Only include what you absolutely need!

Think back to the first model we saw in the class involving fuel-efficient cars. That model was very parsimonious, focusing on a particular aspect of the impact of fuel-efficient vehicles on the environment. The model **did not** include, for example: (i) how the production processes of fuel-efficient car versus standard car differentially impact the environment; (ii) how fuel-efficient cars implement some green technologies to make them less damaging to the environment for reasons other than high mpg; (iii) whether the intrinsic pride of owning a hybrid vehicle would affect driving habits for reasons beyond mpg. Instead, we zoomed in on fuel economy and focused on the key idea that improving fuel-efficiency carries the side-effect of reducing the "price" for each mile traveled, resulting in higher demand for miles.

1.3 *Practical advice.* Focusing on the theme of optimization and simplicity, here are some steps you should follow for modeling problems.

$$\boxed{\text{Step 1: Write down the relevant variables.}}$$

Writing down important variables for the economic environment is a good first step towards a complete model. Be sure to distinguish between **endogenous variables** and **parameters**.

For the sake of simplicity, include only the endogenous variables you absolutely need, and even then see if you can **reduce the dimension** of the problem by expressing some of them as dependent on a common choice variable. For example, two relevant variables for the optimal labor supply problem were income Y and leisure L. We expressed each of them as a function of the choice variable H.

When writing down parameters, include **only those parameters you need for comparative statics**. For example, we include wage w since we want to compute $\frac{\partial H^*}{\partial w}$. Should we also include the price of the consumption good, p? After all, income Y only indirectly yields us utility. It is total consumption Y/p that directly makes us happier. But it would not be a good idea to include p, unless you also plan on computing $\frac{\partial H^*}{\partial p}$. Else, just assume price level stays constant and leave p out of the model.

$$\boxed{\text{Step 2: Think about the trade-off.}}$$

Decision makers face trade-offs in just about every interesting economic situation. Ask yourself, what is it that prevents economic actors from getting "infinite happiness" in this problem? For instance,

[3] This week's section material borrows from the work of previous TFs, in particular the notes of Zhenyu Lai.

in the labor supply problem, an individual cannot send $U(Y, L)$ to infinity since higher Y trades-off for lower L.

The trade-off will suggest either a **dependency** of the variables on a common choice variable, or a **constraint**. Indeed, we have formalized the trade-off between Y and L by writing both as functions of H, with $\frac{dY}{dH} > 0$ but $\frac{dL}{dH} < 0$. But we could have also formalized the trade-off by writing down the constraint $Y = w(24 - L)$ and maximize the objective $U(Y, L)$ subject to this constraint.

$$\boxed{\text{Step 3: Set up an optimization problem.}}$$

This could be a constrained optimization problem if you wrote the trade-off as a constraint, or an unconstrained one if you managed to express some endogenous variables as functions of a common choice variable. If there are multiple ways to set up the optimization problem, choose the representation where the **choice variables are the ones you want to take comparative statics on**. For example, even though the labor supply problem could have been rephrased in terms of constrained maximization over (Y, L) pairs, we were really interested in an optimization problem with H as a choice variable since we wanted to compute $\frac{\partial H^*}{\partial w}$.

$$\boxed{\text{Step 4: Do calculus.}}$$

Now that you have the optimization problem, the next steps are mechanical. Find the **FOC** and express the optimal value of choice variable as a function of the parameter of interest. Then, **totally differentiate** both sides with respect to this parameter. Rearrange to isolate the desired **comparative statics**. Discuss the **sign** of each summand in the expression that emerges.

2 One-variable, one-parameter maximization

2.1 *A function involving both choice variable and parameter.* In lecture we saw a general maximization problem of the form:

$$\max_{x \in \mathbb{R}} V(x; z)$$

We can think of V as **a family of** single-variable functions, indexed by z. That is, different z's correspond to different economic environments. For example...

- z is wage, x is number of hours worked, $V(x; z) = U(z \cdot x, 24 - x)$ gives individual's utility
- z is wage, x is amount of labor employed, $V(x; z) = p_0 \cdot f(x) - z \cdot x$ gives firm's profit where f is the production function (assuming p_0 price of output never changes)

The decision maker does not get to choose which economic environment she faces. Therefore, z is a parameter for the optimization problem, not a choice variable. However, we might still be interested in how the agent's optimal choice responds to changes in the environment. This involves calculating $\frac{\partial x^*}{\partial z}$.

2.2 *Computing* $\frac{\partial x^*}{\partial z}$. This computation frequently comes up in different contexts. So let's derive it carefully for the general case. Starting with the objective function, take partial derivative with respect to x:

$$V_x(x, z)$$

FOC says if x^* is a local maximum under parameter value $z = z_0$, it must be the case that $V_x(x^*, z_0) = 0$. Using the **implicit function assumption**, there is a differentiable function $x^*(\cdot)$ so that

$$V_x(x^*(z), z) = 0$$

holds for all z close enough to z_0. Differentiate both sides with respect to z obtains:

$$V_{xx}(x^*(z), z) \cdot \frac{dx^*}{dz}(z) + V_{xz}(x^*(z), z) = 0$$

Which we rearrange to get:

$$\boxed{\frac{dx^*}{dz} = -\frac{V_{xz}(x^*(z), z)}{V_{xx}(x^*(z,), z)}} \tag{10}$$

Furthermore, since $x^*(z_0)$ is a local maximum for $V(x, z_0)$, then **SOC** must say $V_{xx}(x^*(z_0), z_0) \leq 0$. Therefore,

$$\boxed{\frac{dx^*}{dz}(z_0) \propto V_{xz}(x^*(z_0), z_0)} \tag{11}$$

In particular, if $V_{xz} > 0$ everywhere, then $\frac{dx^*}{dz} > 0$ everywhere.

3 Firm's problem

3.1 *Firm's problem with one input.* Suppose firm optimizes:

$$\max_{L \geq 0} [\pi(L; w)] = \max_{L \geq 0} [pf(L) - wL]$$

where $f : \mathbb{R} \to \mathbb{R}$ is a production function, p is price of output, and w is wage that must be paid for labor. Two remarks:

- Wait, isn't π also dependent on p? Yes, but suppose we only want comparative statics with respect to w.
- Sometimes firm's optimal choice is the **corner solution** $L = 0$. This might happen if, say, $f(L) = L$ and $p < w$. Applying just the FOC will fail to pick up this corner solution!

Now assume the optimum is an interior solution. Then results from general one-variable, one-parameter optimization apply, with π as V, L as x, w as z. So,

$$\frac{dL^*}{dw} \propto \pi_{Lw} = -1$$

That is, firm always demands less labor when wage increases.

3.2 *Firm's problem with two inputs via Cramer's rule.* Suppose firm maximizes:

$$\max_{K,L \geq 0} [\pi(K, L; p, r, w)] = \max_{K,L \geq 0} [pf(K, L) - rK - wL]$$

FOC requires $\pi_K = 0$ and $\pi_L = 0$, that is:

$$\begin{cases} pf_K(K^*(w,r,p), L^*(w,r,p)) - r &= 0 \\ pf_L(K^*(w,r,p), L^*(w,r,p)) - w &= 0 \end{cases}$$

where we view K^*, L^* as functions of w by implicit function assumption. Differentiate both sides by w,

$$\begin{cases} p(f_{KK}\frac{\partial K^*}{\partial w} + f_{KL}\frac{\partial L^*}{\partial w}) &= 0 \\ p(f_{LK}\frac{\partial K^*}{\partial w} + f_{LL}\frac{\partial L^*}{\partial w}) &= 1 \end{cases}$$

We can view this as a system of two linear equations in two unknowns ($\frac{\partial K^*}{\partial w}$ and $\frac{\partial L^*}{\partial w}$). **Cramer's rule** gives an explicit solution for this class of problems. The system

$$\begin{cases} a_1 x + b_1 y &= c_1 \\ a_2 x + b_2 y &= c_2 \end{cases}$$

is solved by:

$$\boxed{x = \frac{\begin{vmatrix} c_1 & b_1 \\ c_2 & b_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}}, y = \frac{\begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}}} \tag{12}$$

where $|\cdot|$ denotes **matrix determinant**, $\begin{vmatrix} a & b \\ c & d \end{vmatrix} := ad - bc$. Applied to our particular case, where

$$\begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix} = \begin{bmatrix} pf_{KK} & pf_{KL} \\ pf_{LK} & pf_{LL} \end{bmatrix}$$

and

$$\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

deduce:

$$\frac{\partial K^*}{\partial w} = \frac{0 - pf_{KL} \cdot 1}{pf_{KK} \cdot pf_{LL} - pf_{KL} \cdot pf_{LK}} = -\frac{1}{p}\frac{f_{KL}}{f_{KK}f_{LL} - f_{KL}^2}$$
$$\frac{\partial L^*}{\partial w} = \frac{pf_{KK} \cdot 1 - 0}{pf_{KK} \cdot pf_{LL} - pf_{KL} \cdot pf_{LK}} = \frac{1}{p}\frac{f_{KK}}{f_{KK}f_{LL} - f_{KL}^2}$$

as in lecture.

1 Envelope theorem

1.1 *The value function.* Consider again the general 1-variable, 1-parameter maximization problem:

$$\max_{x \in \mathbb{R}} V(x; z)$$

We write $V^*(z) := \max_{x \in \mathbb{R}} V(x; z)$, the **value function** associated with the optimization problem. Just like how $x^*(z)$ tells us what the argmax is for each parameter z, the function $V^*(z)$ returns the maximized value of $V(x; z)$ for each z. Indeed, $V^*(z) = V(x^*(z))$. Graphically, $V^*(z)$ forms an "**upper envelope**" on the graph of $V(x; z)$, as in Figure 3.

We might ask how the value function responds to changes in z. For instance, if we think of V as firm's profit in single input case, with x being labor and z being wage, then $\frac{dV^*}{dz}$ represents the marginal change in the firm's maximized profit for a dollar increase in worker's wage. Graphically, we are asking for the **slope of the upper envelope** as the parameter varies.

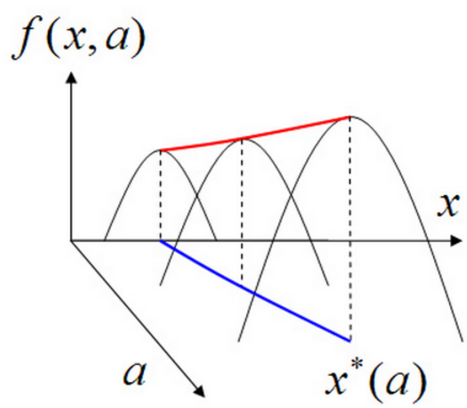


Figure 3: Suppose we are interested in $\max_{x \in \mathbb{R}} f(x; a)$ where a is a parameter. Then the value function forms an upper envelope (red curve) over f.

1.2 *The unconstrained envelope theorem.* You might think that a term involving $\frac{\partial x^*}{\partial z}$ should appear in the expression for $\frac{dV^*}{dz}$. After all, a change in the environment should change the optimal choice in the optimization, which will in turn affect the maximized function value. However, the envelope theorem says $\frac{\partial x^*}{\partial z}$ does not actually appear:

$$\boxed{\frac{dV^*}{dz}(z) = V_z(x^*(z); z)} \tag{13}$$

To derive this result, write by chain rule:

[4] Part 1 of this week's section material borrows from the work of previous TFs, in particular the notes of Zhenyu Lai. Image credit: http://exactitude.tistory.com/229

$$\frac{dV^*}{dz} = \underbrace{V_x(x^*(z);z)\cdot\frac{dx^*(z)}{dz}}_{\text{re-optimization effect}} + \underbrace{V_z(x^*(z);z)}_{\text{direct effect}}$$

However, $(x^*(z), z)$ pair **must set the FOC of optimization to 0**, since it is a maximizer. Therefore, $V_x(x^*(z);z) = 0$, meaning $V_x(x^*(z);z)\cdot\frac{dx^*(z)}{dz} = 0$.

1.3 *Some examples.* As a first example, suppose $V(x;z) = -(x-z)^2 + z^2$. Geometrically, $V(x;z)$ is a **parabola**. Increasing z does two things: (1) It changes where the parabola is maximized, i.e. it changes $x^*(z)$; (2) It shifts the parabola upwards, i.e. $V_z > 0$. Let's write $V^*(z)$ for the maximum height of the parabola for parameter z. The envelope theorem asserts that $\frac{dV^*}{dz}(z) = V_z(x^*(z);z)$. Taking the appropriate partial we find

$$V_z = 2(x-z) + 2z$$

so

$$\frac{dV^*}{dz}(z) = V_z(x^*(z);z) = 2(x^*(z)-z) + 2z = 2z$$

which we can verify geometrically. Remarkably, we **never had to compute** $\frac{dx^*}{dz}$.

In general, the envelope theorem can greatly **simplify computation** of $\frac{dV^*}{dz}$ when $\frac{dx^*}{dz}$ is complicated. Consider $V(x;z) = \sin(z)\cdot x + \mathfrak{F}(x)$ where $\mathfrak{F}(x)$ is a complicated concave function. In this case, even $V_x(x;z)$ might be hard to compute due to difficulties associated with differentiating $\mathfrak{F}$. Fortunately, $\frac{dV^*}{dz}$ is not hard to write down, as it is just $\cos(z)\cdot x^*(z)$ according to envelope theorem.

1.4 *The constrained envelope theorem.* Suppose instead we are dealing with a **constrained** optimization problem:

$$\max_{x,y\in\mathbb{R}} V(x,y;z) \quad \text{s.t. } g(x,y;z) = c$$

We form the Lagrangian:

$$\mathcal{L}(x,y,\lambda;z) = V(x,y;z) + \lambda(g(x,y;z) - c)$$

Analogous to the value function for unconstrained optimization, we can also define the value function for constrained optimization

$$V^*(z) = V(x^*(z), y^*(z); z)$$

where $x^*(z), y^*(z)$ are constrained maximizers at parameter z. The constrained envelope theorem states that:

$$\boxed{\frac{dV^*}{dz}(z) = \frac{d\mathcal{L}}{dz}(x^*(z), y^*(z), \lambda^*(z); z) = V_z(x^*(z), y^*(z); z) + \lambda^* g_z(x^*(z), y^*(z); z)} \tag{14}$$

2 Profit and cost functions

2.1 *The value function framework.* Both profit function and cost function can be viewed as **examples of value functions**. As such, **envelope theorem** applies to them.

The profit function arises out of firm's profit maximization problem, which is an instance of **unconstrained maximization problem**:

$$\max_{K,L} \{V(K, L; p, r, w)\}$$

Specifically, $V(K, L; p, r, w) = p \cdot f(K, L) - rK - wL$, so the profit maximization problem is explicitly written as:

$$\max_{K,L} \{p \cdot f(K, L) - rK - wL\}$$

The value function of this maximization problem is called the **profit function**, that is:

$$\begin{aligned} \pi(p, r, w) := V^*(p, r, w) &= V(K^*_\pi(p, r, w), L^*_\pi(p, r, w); p, r, w) \\ &= p \cdot f(K^*_\pi(p, r, w), L^*_\pi(p, r, w)) - rK^*_\pi(p, r, w) - wL^*_\pi(p, r, w) \end{aligned}$$

where K^*_π and L^*_π are profit-maximizing capital and labor demands.

On the other hand, the cost function is related to firm's cost minimization problem, which is an instance of **constrained minimization problem**:

$$\min_{K,L} \left\{\hat{V}(K, L; q, r, w)\right\} \text{ s.t. } g(K, L) = q$$

Specifically, $\hat{V}(K, L; q, r, w) = rK + wL$ and $g(K, L) = f(K, L)$ (where f is the production function), so the cost minimization problem is explicitly written as:

$$\min_{K,L} rK + wL \text{ s.t. } f(K, L) = q$$

This is the problem facing a firm that must produce q units of output and wishes to minimize the expenditure on capital and labor inputs in completing this task. The value function of this constrained minimization problem is called the **cost function**, that is:

$$\begin{aligned} C(q, r, w) := \hat{V}^*(q, r, w) &= \hat{V}(K^*_C(q, r, w), L^*_C(q, r, w); q, r, w) \\ &= r \cdot K^*_C(q, r, w) + w \cdot L^*_C(q, r, w) \end{aligned}$$

where K^*_C and L^*_C are cost-minimizing capital and labor demands.

Be careful: $K^*_\pi(p, r, w)$ and $L^*_\pi(p, r, w)$ are **not the same** as $K^*_C(q, r, w)$ and $L^*_C(q, r, w)$. These two pairs of functions refer to optimal capital and labor inputs for **two different** optimization problems: profit maximization and cost minimization. K^*_π, L^*_π are functions of output price p, since this parameter matters for profit maximization. K^*_C, L^*_C are functions of quantity of output q, since this parameter matters for cost minimization.

2.2 *Some properties of profit and cost functions.* In Table 2 we record some useful properties of π and C.

	$\pi(p,r,w)$	$C(q,r,w)$
Interpretation	Maximized profit at price p, capital rental r, wage w	Minimized cost for producing q output, under capital rental r, wage w
Partials	$\frac{\partial \pi}{\partial p} = f(K_\pi^*, L_\pi^*)$ ["Hotelling's Lemma"]	–
	$\frac{\partial \pi}{\partial r} = -K_\pi^*$	$\frac{\partial C}{\partial r} = K_C^*$ ["Shephard's Lemma"]
	$\frac{\partial \pi}{\partial w} = -L_\pi^*$	$\frac{\partial C}{\partial w} = L_C^*$ ["Shephard's Lemma"]
Homogeneity	$\pi(\lambda p, \lambda r, \lambda w) = \lambda \pi(p,r,w)$	$C(q, \lambda r, \lambda w) = \lambda C(q,r,w)$
Curvature	convex in (p,r,w)	concave in (r,w)

Table 2: Some useful properties of profit function and cost function.

3 Production functions

3.1 *Some functional forms for production functions.*

(A) Separable. Generally, $f(x_1, x_2, ..., x_n) = g_1(x_1) + g_2(x_2) + ... + g_n(x_n)$. As an example, $f(x_1, x_2, x_3) = \alpha x_1 + \alpha x_2^\beta + \ln(1 + x_3)$. The mixed second partial derivatives of a separable function are **all 0**. Further, separable production functions yield very **tractable first order conditions** for profit maximization. Indeed, the problem

$$\max_{x_1, x_2, \ldots, x_n} p \cdot (g_1(x_1) + g_2(x_2) + ... + g_n(x_n)) - w_1 x_1 - w_2 x_2 - ... - w_n x_n$$

has the FOCs $p \cdot g_j'(x_j^*) = w_j$ for all $1 \le j \le n$.

(B) Leontief. $f(x,y) = \min\left\{\frac{x}{a_x}, \frac{y}{a_y}\right\}$ for $a_x, a_y > 0$. In this production function, x and y are **perfect complements**. A cost-minimizing firm will always demand inputs in "multiples" of (a_x, a_y). Each such multiple produces one unit of output.

(C) Linear. $f(x,y) = a_x x + a_y y$ with $a_x, a_y > 0$. Technically this is a case of separable production function, but it exhibits some very special features that deserve highlighting. A firm aiming to produce a certain quantity Q of output will demand only x or only y, depending on which is more cost-efficient (i.e. pick the larger $\frac{a_j}{p_j}$). In the knife-edge case where the ratio of productivity exactly equals the ratio of prices, i.e. $\frac{a_x}{a_y} = \frac{p_x}{p_y}$, then any combination of inputs that produces the output goal of Q is cost-minimizing.

3.2 *Concavity and returns to scale.* In lecture, we saw several concepts for describing production functions: concavity, convexity, and returns to scale. It turns out several of these concepts are connected.

Concavity and convexity describe the curvature in a function. For $f : \mathbb{R}^n \to \mathbb{R}$, say f is **concave** if[5]

$$f(\alpha \mathbf{x} + (1-\alpha)\mathbf{y}) \ge \alpha f(\mathbf{x}) + (1-\alpha) f(\mathbf{y})$$

for all $\alpha \in (0,1)$, $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Conversely, call f **convex** if

$$f(\alpha \mathbf{x} + (1-\alpha)\mathbf{y}) \le \alpha f(\mathbf{x}) + (1-\alpha) f(\mathbf{y})$$

[5] Here and henceforth, variable in boldface denotes vector.

for all $\alpha \in (0,1)$, $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. For instance, take $f(L) = L^\alpha$. Then f is concave when$\alpha \leq 1$, convex when $\alpha \geq 1$. Concavity of f is important in deciding whether FOC will find the maximizers for the profit function. Taking FOC of the firm's profit maximization problem when $\alpha = 2$, that is $\frac{d}{dL}(pL^2 - wL)$, yields the **local minimum** $L^* = \frac{w}{2p}$.

Return to scale captures how well the production function scales up. Graphically, one starts with some initial level $\mathbf{x} \in \mathbb{R}^n$ of inputs, which can be drawn as a ray segment from the origin. Next, extend this ray segment to λ times its previous length, so that point $\lambda\mathbf{x}$ is reached. Return to scale involves comparing the production level at the tip of this extended ray segment to the original production level. We say a production function exhibits **increasing returns to scale** if $f(\lambda\mathbf{x}) > \lambda f(\mathbf{x})$ for $\lambda > 1$, **decreasing returns to scale** if $f(\lambda\mathbf{x}) < \lambda f(\mathbf{x})$ for $\lambda > 1$, and **constant returns to scale** if $f(\lambda\mathbf{x}) = \lambda f(\mathbf{x})$ for $\lambda > 0$.

Suppose the production function is such that $f(\vec{0}) = 0$. Then a strictly concave f exhibits decreasing returns to scale, while a strictly convex f exhibits increasing returns to scale. To see this, suppose f is strictly concave and fix some $\mathbf{x}_0 \in \mathbb{R}^n$ and $\lambda > 1$. Let $\mathbf{x} = \vec{0}$, $\mathbf{y} = \lambda\mathbf{x}_0$ in the definition of concavity and obtain:

$$f\left(\left(1 - \frac{1}{\lambda}\right)\mathbf{x} + \frac{1}{\lambda}\mathbf{y}\right) > \left(1 - \frac{1}{\lambda}\right) f(\mathbf{x}) + \left(\frac{1}{\lambda}\right) f(\mathbf{y}) = 0 + \frac{1}{\lambda} f(\mathbf{y})$$

But $\left(1 - \frac{1}{\lambda}\right)\mathbf{x} + \frac{1}{\lambda}\mathbf{y} = \mathbf{x}_0$ and $\mathbf{y} = \lambda\mathbf{x}_0$, so $f(\mathbf{x}_0) > \frac{1}{\lambda} f(\lambda\mathbf{x}_0)$, that is $f(\lambda\mathbf{x}_0) < \lambda f(\mathbf{x}_0)$.

3.3 *Jensen's inequality.* Suppose X is a real-valued random variable and $\mathbb{E}$ is the expectation operator. You might remember that

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$$

If we define $\ell(x) := ax + b$, then we can rewrite the above equality as $\mathbb{E}[\ell(X)] = \ell(\mathbb{E}[X])$. That is, the expectation operator **commutes with linear transformations**. However, if φ is a convex function, then in general $\mathbb{E}[\varphi(X)] \neq \varphi(\mathbb{E}[X])$. **Jensen's inequality** tells us the precise consequence of exchanging expectation and a convex function:

$$\boxed{\mathbb{E}[\varphi(X)] \geq \varphi(\mathbb{E}[X]) \text{ for convex } \varphi} \tag{15}$$

An immediate application: put $\varphi(x) := x^2$. Then Jensen's inequality implies $\mathbb{E}[X^2] \geq (\mathbb{E}[X])^2$, which is true since $\mathbb{E}[X^2] = (\mathbb{E}[X])^2 + \mathrm{Var}(X)$ and variance is always nonnegative.

Most functions economists deal with are concave instead of convex. That's okay too. The **"concave version" of Jensen's inequality** asserts that:

$$\boxed{\mathbb{E}[\psi(X)] \leq \psi(\mathbb{E}[X]) \text{ for concave } \psi} \tag{16}$$

As a quick check, suppose X takes on only two values, x and y, with probabilities α and $1 - \alpha$ respectively. Then above inequality implies:

$$\alpha\psi(x) + (1 - \alpha)\psi(y) \leq \psi(\alpha x + (1 - \alpha)y)$$

for concave ψ. But this is exactly the definition of concavity. Now suppose X takes on values x, y, z, with three probabilities $\alpha, \beta, (1 - \alpha - \beta)$ adding up to 1. Then deduce

$$\alpha\psi(x) + \beta\psi(y) + (1 - \alpha - \beta)\psi(z) \leq \psi(\alpha x + \beta y + (1 - \alpha - \beta)z)$$

for concave ψ. In this way, we can view the two versions of Jensen's inequality as **generalizing the definitions of convexity and concavity** for weighted average across more than two points.

To relate Jensen's inequality to production functions, suppose a firm's production process employs exactly one worker. The output is determined as $f(L)$, where L reflects the skill level of the worker. The firm is currently picking between two candidates. Candidate A's skill level is $L_A = 0.5$ with certainty. Candidate B's exact skill level is unknown, but the firm believes that L_B is distributed uniformly on $[0, 1]$. If the two candidates cost the same to hire and firm wants to maximize expected output, then the firm should pick candidate A, since under the usual concavity assumption on f,

$$\mathbb{E}[f(L_B)] \leq f(\mathbb{E}[L_B]) = f(L_A)$$

4 Spatial equilibrium

4.1 *Spatial equilibrium via no-arbitrage condition.* The **no-arbitrage condition** is one of the central assumptions of Economics. Roughly speaking, if people are free to choose from a family of options $\{c_1, c_2, ..., c_n\}$ and we observe a nonzero number of people choosing each option, then the costs and benefits of these n options must be **equalized** in such a way that no one can gain by making a different choice. Put another way, an **arbitrage** is a situation where a decision-maker gets a strict gain by switching from choice c_i to choice c_j. No-arbitrage condition requires the nonexistence of arbitrage. No-arbitrage appears as an assumption in many fields of Economics. For example, **asset pricing** interprets $\{c_i\}$ as financial derivatives and deduces a set of prices for the assets as to preclude arbitrage.

Spatial equilibrium from urban economics is just another form of no-arbitrage condition. However, instead of buying and selling assets, people are choosing where to live. Spatial equilibrium requires that the benefits and costs of living in different cities must be equalized in such a way that people are exactly indifferent between different choices. Else, there would be a **complete migration** from one city to another, just as a **mispricing** in the asset market would lead everyone to sell the asset that is priced too high and buy the asset that is priced too low.

Formally, suppose utility from living in a city with wage w, housing price p, and "amenities" a is $U(w, p, a)$. Suppose everyone is free to move between a set of n cities, $\mathcal{C} = \{c_1, c_2, ..., c_n\}$. Then we may construct functions $w : \mathcal{C} \to \mathbb{R}$, $p : \mathcal{C} \to \mathbb{R}$, $a : \mathcal{C} \to \mathbb{R}$ which return the wage, housing price, and amenities in different cities. This way, all inputs to U have been expressed as functions of a common variable c, so that we may write

$$U(w(c), p(c), a(c))$$

Spatial equilibrium is the requirement that this function is **constant across all** $c \in \mathcal{C}$. Graphically, if we construct the "level surface" of U at utility level $U(w(c_1), p(c_1), a(c_1))$, then the n points $(w(c_i), p(c_i), a(c_i))$ for $1 \leq i \leq n$ also belong to the same level surface.

4.2 *Spatial equilibrium depends on mobility.* It is important to remember that spatial equilibrium is only an **assumption** we impose regarding how wage, housing price, and amenities covary. Certainly, we would not expect $U(w(c), p(c), a(c))$ to be constant across c if w, p, a were arbitrary functions of c. Whether or not spatial equilibrium holds (at least approximately) in reality depends critically on **mobility between cities**.

We have assumed throughout that people are free to choose between cities $c_1, c_2, ..., c_n$. If there were **barriers** to migration between cities, then we would not expect spatial equilibrium to hold. For instance, it would not be appropriate to use spatial equilibrium to compare cities in different countries. Almost surely one would find that cities in developing countries have lower net income (i.e. wage minus housing price) and worse amenities than cities in developed countries. This does not mean people are missing an arbitrage opportunity. Rather, the phenomenon arises due to **immigration barriers** preventing the free movement of people internationally.

Economics 1011a ▷ Section 3 : Foundations of Consumer Behavior[6] 9/23/2014

(1) Constructing utility; **(2)** Monotonic transformation of utility; **(3)** Consumer's problem

TF: Kevin He (he02@fas.harvard.edu)

1 Constructing utility

1.1 *Preference on a finite set as a tournament.* Suppose X is a finite set and $\succsim$ is a preference relation on X. For concreteness, think of $X = \{p, q, r, s\}$ with p = peach, q = quiche, r = red grapefruit, s = steak. Assume, as usual, that $\succsim$ is rational – that is to say, complete and transitive:

- **Completeness**: for any $a, b \in X$, either $a \succsim b$ or $b \succsim a$ (or both).
- **Transitivity**: for any $a, b, c \in X$, if $a \succsim b$ and $b \succsim c$, then $a \succsim c$.

One example of $\succsim$ is given in Table 3.

$\succsim$	p	q	r	s
p	•	•	•	•
q		•	•	•
r			•	
s		•	•	•

Table 3: A rational preference on X. Here, a • in row i, column j means the i^{th} item is weakly preferred to the j^{th} item. Nobody likes grapefruits.[8]

Suppose we want to find a utility function $u : X \to \mathbb{R}$ that **represents** $\succsim$. That is, $a \succsim b$ if and only if $u(a) \geq u(b)$. One way to do this is to imagine running a **round-robin tournament** amongst the 4 food items in X, with $\succsim$ as the judge. Let's start with p, who will play against opponents p, q, r, s in the tournament (yes, even against itself!) Food p wins the match against food x if and only if $p \succsim x$, in which case p gains one point in the tournament. (If both $p \succsim x$ and $x \succsim p$, then both p and x "won" and each gets a point.) Add up the points for p in the entire tournament and assign this number as $u(p)$. That is, the utility of p is defined as the **number of food items** in X that p "beats", where "beats" really means "being weakly preferred to". When applied to the particular example in Table 3, we obtain $u(p) = 4$, $u(q) = 3$, $u(r) = 1$, $u(s) = 3$. You can check that u represents $\succsim$ in this example.

More generally, the u that emerges out of this construction always represents $\succsim$, thanks to **transitivity**. Suppose $p \succsim q$. That means p must beat any food that q beats, by transitivity. Hence, p's total points in the tournament can be no fewer than q's total points, implying $u(p) \geq u(q)$ as needed. (Try explaining why $u(p) \geq u(q)$ implies $p \succsim q$ yourself.)

1.2 *Lexicographic preference on* $\mathbb{N} \times \mathbb{N}$. Let's think about a more complicated preference. The domain of the preference is $\mathbb{N} \times \mathbb{N}$, i.e. pairs of natural numbers. **Lexicographic preference** is

[6] Part 3 of this week's section material borrows from the work of previous TFs, in particular the notes of Zhenyu Lai.

[7] See, for example, http://xkcd.com/388/.

defined so that the decision-maker compares two pairs using the **first coordinate** in each pair. If this results in a tie, then (and only then) the **second coordinate** is considered for tie-breaking. For instance, $(3,2) \succsim (2,1)$, $(3,2) \succsim (2,5)$, $(3,4) \succsim (3,2)$. The preference is called "lexicographic" since the coordinate-by-coordinate comparison of two pairs resembles the ordering of words in a dictionary. To add economic interpretation, suppose each consumption alternative is described by two attributes, where the first attribute is **infinitely more important** than the second attribute. Then the consumer would exhibit lexicographic preference.

The "tournament" construction no longer works, since almost all pairs in $\mathbb{N} \times \mathbb{N}$ are weakly preferred to infinitely many other pairs. We have to try another strategy. As a first step, suppose we fix utility levels for pairs of the form $(n,0)$ as $u(n,0) = n$. Then u certainly represents $\succsim$, provided we restrict attention to those pairs whose second coordinate is 0. To extend u as to represent $\succsim$ on all of $\mathbb{N} \times \mathbb{N}$, we know that since $\succsim$ is lexicographic, $(n,k) \succsim (n,0)$ but $(n+1,0) \succsim (n,k)$ for any $k \in \mathbb{N}$. Therefore we must arrange for $u(n,0) \leq u(n,k) \leq u(n+1,0)$, that is $n \leq u(n,k) \leq n+1$ for any $k \in \mathbb{N}$.

The task, therefore, is to "**compress**" all of $\mathbb{N}$ into an interval of length 1. One way to do this is to set $u(n,k) = n + \xi(k)$, where

$$\xi(k) := \left(\frac{1}{2}\right)^1 + \left(\frac{1}{2}\right)^2 + ... + \left(\frac{1}{2}\right)^k$$

and $\xi(0) := 0$. Since $0 \leq \xi(k) < 1$ for any $k \in \mathbb{N}$, we guarantee $n \leq u(n,k) \leq n+1$. Further, as the second coordinate of (n,k) increases, $u(n,k)$ also increases, preserving the representation.

This construction teaches several lessons about utility functions in general:

- Utility functions are **not unique**. Indeed, we have many choices as how to define $u(n,k)$ for $k \neq 0$ but chose to use $\xi(k)$ on a whim.
- Utility functions **do not express the intensity of preference**. In our construction, we could have instead started with $u(n,0) = n^2$. This new utility function would assign 25 utils to $(5,0)$ instead of 5 utils, yet would still represents the same preference.

Finally, here are two (optional) exercises for you to try at home:

Exercise A: Suppose $\succsim$ is the lexicographic preference on $\mathbb{N} \times \mathbb{R}$. Construct a utility function $u : \mathbb{N} \times \mathbb{R} \to \mathbb{R}$ that represents $\succsim$.

Exercise B: Suppose $\succsim$ is the lexicographic preference on $\mathbb{Q} \times \mathbb{R}$. Construct a utility function $u : \mathbb{Q} \times \mathbb{R} \to \mathbb{R}$ that represents $\succsim$.

1.3 *A theorem on the utility representation of continuous preferences.* Emboldened by our success, we might conjecture that any complete, transitive preference relation admits a utility representation. Unfortunately, **this is false**. Rationality is necessary but not sufficient for utility representation. As a counterexample, the lexicographic preference on $\mathbb{R} \times \mathbb{R}$ cannot be represented by **any**[8] function $u : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$.

[8]That is, no representation exists even if we allow for discontinuous u.

Proof: Suppose, by way of contradiction, that there exists some $u : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ that represents $\succsim$. This allows the construction of a one-to-one function $\varphi : \mathbb{R} \to \mathbb{Q}$ as follows. For any $r \in \mathbb{R}$, lexicographic preference says $(r,1) \succ (r,0)$. Therefore, $u(r,1) > u(r,0)$ for each $r \in \mathbb{R}$ since u represents $\succsim$. By density of $\mathbb{Q}$ in $\mathbb{R}$, there exists at least one rational number q satisfying $u(r,1) > q > u(r,0)$. Pick one such q and define $\varphi(r) := q$. Repeating this construction for every $r \in \mathbb{R}$ defines φ on all real numbers. Notice φ is strictly increasing. Indeed, for any two real numbers $r_1 < r_2$, lexicographic preference requires:

$$(r_1, 0) \prec (r_1, 1) \prec (r_2, 0) \prec (r_2, 1)$$

therefore, by construction,

Gérard Debreu proved in 1954 that under the additional assumption of **continuity**, representation can be guaranteed. As a bonus, we can even find a continuous utility function to represent $\succsim$. We state his result below:

Theorem. *(Debreu) Preference $\succsim$ on $X \subseteq \mathbb{R}^n$ is complete, transitive, and continuous if and only if there exists a continuous $u : X \to \mathbb{R}$ that represents $\succsim$.*

A preference is **discontinuous** if we have a sequence of choice objects $a_n \to a$ and another alternative b with $a_n \succsim b$ for every n but $a \not\succsim b$. That is, weak preference has not been preserved by limits. The lexicographic preference on $\mathbb{Q} \times \mathbb{R}$ is an example of a discontinuous preference. We have $(1 + \frac{1}{m}, 0) \succsim (1, 1)$ for each $m \in \mathbb{N}$, yet $(1 + \frac{1}{m}, 0) \to (1, 0)$ with $(1, 0) \not\succsim (1, 1)$. Therefore, Debreu's theorem tells us this preference does not admit a continuous utility representation. (However, **Exercise B** does have a solution! How do you reconcile these two facts?)

2 Monotonic transformations of utility

2.1 *Ordinal representations and monotonic transformations.* For most of this course, we will follow Samuelson and treat **utility functions as ordinal**. We consider preference relations as primitive while utility functions are but convenient, **numerical summaries** about these binary relations. Their sole role is ease of use – it is easier to work with real-valued functions than binary relations and it is more convenient to think about the ordering of real numbers than a complicated (but equivalent) ordering on an abstract set X.

Since utility functions are just carriers of **rank information,** we are allowed to transform them in any way that preserves this information. In particular, we are free to apply **monotonic transformations** to them. Consider a function $h : \mathbb{R} \to \mathbb{R}$.

For $A \subseteq \mathbb{R}$, call h monotonic on A if $x_1, x_2 \in A$ with $x_1 < x_2$ implies $h(x_1) < h(x_2)$.

Provided our utility function u only takes values in A, we can compose it with a monotonic function on A to form $h \circ u$, which will represent the **same preference**. It is important to remember that some functions are not monotonic everywhere, but only monotonic on a subset of $\mathbb{R}$. For example, it would not be okay to apply $h(v) = v^2$ to a utility function that also takes negative values, such as $u(x) = \ln(x)$. The newly formed $(h \circ u)(x) = (\ln(x))^2$ does not represent the same preference.

2.2 *Monotonic transformations can remove "nice" properties.* Remember, monotonic functions aren't always very "nice" functions. For example, consider the transformation:

$$h(v) = \begin{cases} v & \text{if } v < 0 \\ v + 1 & \text{if } v \geq 0 \end{cases}$$

Then h is monotonic on $\mathbb{R}$ but **not continuous**. Suppose we have a continuous utility function, $u(x) = \ln(x)$. Then $h \circ u$ represents the same preference but is discontinuous!

$$u(r_1, 0) < \varphi(r_1) < u(r_1, 1) < u(r_2, 0) < \varphi(r_2) < u(r_2, 1)$$

so in particular we see $r_1 < r_2 \Rightarrow \varphi(r_1) < \varphi(r_2)$. This shows φ is one-to-one. But no one-to-one function can exist from $\mathbb{R}$ to $\mathbb{Q}$, since the former is uncountably infinite while the latter is countably infinite. Contradiction. □

Remark: Actually, there is one wrinkle in this "proof". Which famous axiom from set theory did I forget to invoke?

$$(h \circ u)(x) = \begin{cases} \ln(x) & \text{if } x < 1 \\ \ln(x) + 1 & \text{if } x \geq 1 \end{cases}$$

Moral of the story: you should avoid making claims of the form "every utility representation of this preference relation satisfies [`some nice property`]", because one can often apply a nasty monotonic transformation to a utility function with [`nice property`] and make the property go away, while still representing the same preference.

In particular, one should remember that Debreu's theorem **does not** assert that **every** utility representation of a complete, transitive, and continuous preference relation is continuous, only that there **exists at least one** utility representation that is continuous.

2.3 *Wait, what about production functions?* One practical use of monotonic transformations is to **simplify** utility maximization problems. Since

$$U(x, y) = x^{\alpha} y^{1-\alpha}$$

represents the same preference as

$$\hat{U}(x, y) = \alpha \ln(x) + (1 - \alpha) \ln(y)$$

we can find the same Marshallian demand whether we solve the original

$$\max_{x,y} x^{\alpha} y^{1-\alpha} \text{ s.t. } p_x \cdot x + p_y \cdot y \leq Y$$

or the easier

$$\max_{x,y} \alpha \ln(x) + (1 - \alpha) \ln(y) \text{ s.t. } p_x \cdot x + p_y \cdot y \leq Y$$

Recalling the painful algebra you had to endure for firm optimization, you might get the idea of applying monotonic transformations to a firm's production function as to simplify computations. This is **not valid**! Whereas a utility function is a "second-class citizen" that merely represent an underlying preference, a production function is a "first-class citizen" whose numerical output **literally** refers to the output level of the firm. Production function is "**real**" in a sense that utility function is not. It is not okay to, say, apply natural log to a production function. More concretely,

$$\max_{K,L} p \cdot K^{1/3} L^{1/3} - rK - wL$$

is not equivalent to:

$$\max_{K,L} p \cdot \left(\frac{1}{3} \ln(K) + \frac{1}{3} \ln(L) \right) - rK - wL$$

This highlights a fundamental difference between utility functions and production functions: the former is **ordinal** while the latter is **cardinal**.

3 Consumer's problem

3.1 *Utility maximization.* In the last few weeks, we saw that the theory of firm optimization involves two canonical problems: **profit maximization** and **cost minimization**. The theory of consumer optimization involves two analogous problems: **utility maximization** and **expenditure minimization**.

Utility maximization is a constrained maximization problem facing a consumer with a limited budget:

$$\max_{\mathbf{x}} u(\mathbf{x}) \text{ s.t. } \mathbf{p}\cdot\mathbf{x} \leq Y$$

Here, **consumption bundle** $\mathbf{x} = (x_1, x_2, ..., x_n)$ is the choice variable while **prices** $\mathbf{p} = (p_1, p_2, ..., p_n) \in \mathbb{R}^n_+$ and **income** $Y \in \mathbb{R}$ are parameters. As always, an optimization problem is naturally associated with two functions. The first one expresses the dependence of maximizer on parameters,

$$\mathbf{x}(\mathbf{p}, Y) = (x_1(\mathbf{p}, Y), x_2(\mathbf{p}, Y), ..., x_n(\mathbf{p}, Y))$$

This is called the **Marshallian demand**. The second one is the value function, which is written as $v(\mathbf{p}, Y)$ and called **indirect utility function** in the context of utility maximization problems. By definition,

$$v(\mathbf{p}, Y) = u(\mathbf{x}(\mathbf{p}, Y))$$

To add geometric interpretation, suppose $n = 2$ and the two goods are called x and y (rather than x_1 and x_2). Fixing p_y and Y, we investigate the function $p_x \mapsto x(p_x, p_y, Y)$. That is, how does the Marshallian demand for good x change as its price changes? This function is plotted in Figure 4.

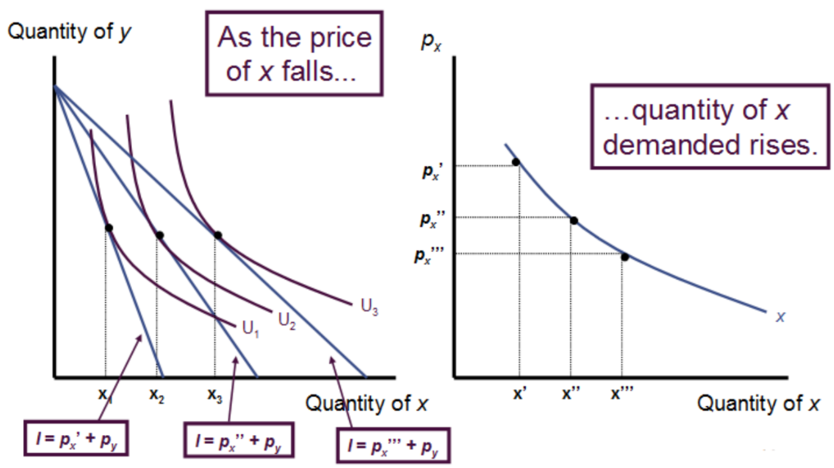


Figure 4: Geometric interpretation of Marshallian demand. The slope of the budget constraint changes as p_x changes. This affects the Marshallian demand, which is characterized by a point of **tangency** between the budget constraint and some level curve of the utility function $\left(\frac{u_x}{u_y} = \frac{p_x}{p_y}\right)$. As the slope of the budget constraint varies, the point of tangency occurs on level curves belong to **different** utility levels. **Remark**: in this example, Marshallian demand for x decreases when p_x increases, but this need not hold in general.

But $p_x \mapsto x(p_x, p_y, Y)$ has a well-known name – it's the **demand curve**! In an introductory economics class, the demand curve is usually taken as a **primitive**, without questioning where the curve comes from. But we can answer this question in our utility maximization framework: demand curve is a "slice" of the maximizer function arising from the utility maximization problem.

3.2 *Expenditure minimization.* **Expenditure minimization** refers to the constrained minimization problem facing a consumer who **must** achieve utility level $\bar{u}$ and who wishes to spend the **least amount of money** in completing this task.

$$\min_{\mathbf{x}} \mathbf{p} \cdot \mathbf{x} \text{ s.t. } u(\mathbf{x}) \geq \bar{u}$$

Here, **consumption bundle** $\mathbf{x} = (x_1, x_2, ..., x_n)$ is the choice variable while **prices** $\mathbf{p} = (p_1, p_2, ..., p_n) \in \mathbb{R}^n_+$ and **desired utility level** $\bar{u} \in \mathbb{R}$ are parameters. Again, the optimization problem naturally induces two functions. One expresses the cost-minimizing choice of consumption as a function of the parameters. This is usually written as $\mathbf{h}(\mathbf{p}, \bar{u}) = (h_1(\mathbf{p}, \bar{u}), h_2(\mathbf{p}, \bar{u}), ..., h_n(\mathbf{p}, \bar{u}))$ and called the **Hicksian demand**. The value function of this problem is written as $e(\mathbf{p}, \bar{u})$ and called the **expenditure function**. By definition,

$$e(\mathbf{p}, \bar{u}) = \mathbf{p} \cdot \mathbf{h}(\mathbf{p}, \bar{u})$$

To add geometric interpretation, suppose $n = 2$ and the two goods are called x and y (rather than x_1 and x_2). Fixing p_y and $\bar{u}$, we investigate the function $p_x \mapsto h_x(p_x, p_y, \bar{u})$. That is, how does the Hicksian demand for good x change as its price changes? This function is plotted in Figure 5.

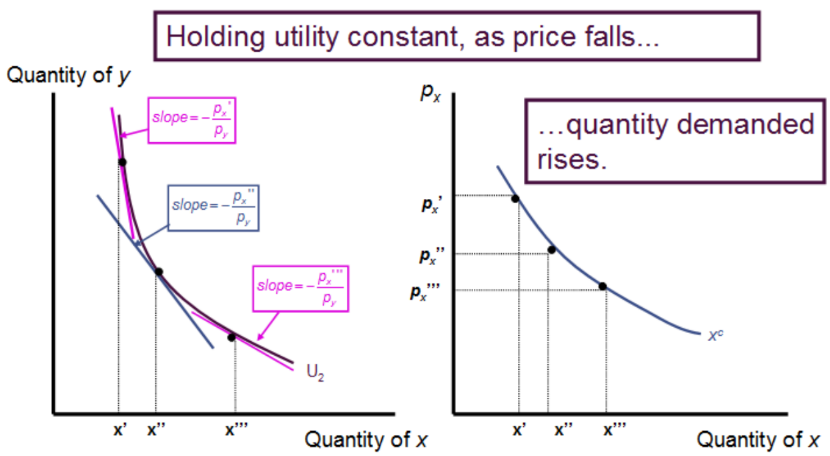


Figure 5: Geometric interpretation of Hicksian demand. The slope of the budget constraint changes as p_x changes. However, unlike in Marshallian demand where the budget constraint stays at a fixed level (namely, Y), in Hicksian demand it is the **utility level curve** that **remains fixed** (at $\bar{u}$). Instead, consumer's **budget** level is allowed to **grow** or **shrink** as to become tangent to the utility level curve $\left(\frac{u_x}{u_y} = \frac{p_x}{p_y}\right)$. This point of tangency is the Hicksian demand.

3.3 *Properties of indirect utility function and expenditure function.* Table 4 collects together some useful properties of the indirect utility function and expenditure function.

	$v(\mathbf{p}, Y)$	$e(\mathbf{p}, \bar{u})$
Interpretation	Maximized utility at prices $\mathbf{p}$, income Y	Minimized expenditure for getting $\bar{u}$ utility, under prices $\mathbf{p}$
Partials	$-\frac{v_{p_i}(\mathbf{p},Y)}{v_Y(\mathbf{p},Y)} = x_i(\mathbf{p}, Y)$ ["Roy's identity"]	$e_{p_i}(\mathbf{p}, \bar{u}) = h_i(\mathbf{p}, \bar{u})$ ["Shephard's Lemma"]
Monotonicity	Non-increasing in $\mathbf{p}$ Non-decreasing in Y	Non-decreasing in $\mathbf{p}$ Non-decreasing in $\bar{u}$
Homogeneity	$v(\lambda\mathbf{p}, \lambda Y) = v(\mathbf{p}, Y)$	$e(\lambda\mathbf{p}, \bar{u}) = \lambda e(\mathbf{p}, \bar{u})$
Continuity	$v(\mathbf{p}, Y)$ is continuous	$e(\mathbf{p}, \bar{u})$ is continuous
Curvature	quasi-convex in $(\mathbf{p}, Y)$	concave in $\mathbf{p}$

Table 4: Some properties of indirect utility function and expenditure function.

Economics 1011a ▷ Section 4 : Duality and Its Implications[9] —— 9/30/2014

(1) Duality in optimization; **(2)** The Slutsky equation; **(3)** Identities from the budget constraint

TF: Kevin He (he02@fas.harvard.edu)

1 Duality in optimization

1.1 *Duality in consumer's problems.* Recall the two canonical problems in consumer theory: utility maximization problem (**UMP**) and expenditure minimization problem (**EMP**).

$$\max_{\mathbf{x}\geq 0} u(\mathbf{x}) \text{ s.t. } \mathbf{p}\cdot\mathbf{x}\leq Y \tag{\textbf{UMP}}$$

$$\min_{\mathbf{x}\geq 0} \mathbf{p}\cdot\mathbf{x} \text{ s.t. } u(\mathbf{x})\geq \bar{u} \tag{\textbf{EMP}}$$

We can identify two parallels between these two problems just by inspection:

- The objective of **UMP** is the constraint of **EMP**, while the constraint of **UMP** is the objective of **EMP**
- The **UMP** is a constrained maximization problem with a less-than-or-equals-to constraint. The **EMP** is a constrained minimization problem with a greater-than-or-equals-to constraint.

But the most important connection between these two optimization problems is that they **obtain the same optimizer** when their parameters **align** in a particular way. Suppose we begin with the **UMP** with parameters $(\mathbf{p}, Y)$. In general, an **EMP** with parameters $(\hat{\mathbf{p}}, \bar{u})$ does not yield the same solution in $\mathbf{x}$ as the original **UMP**. However, if $\hat{\mathbf{p}} = \mathbf{p}$ (prices align) and $\bar{u} = v(\mathbf{p}, Y)$ (utility parameter equals to maximized objective of **UMP**), then **EMP** does give the same solution as **UMP**. This is stated more succinctly as:

$$\boxed{\mathbf{h}(\mathbf{p}, v(\mathbf{p}, Y)) = \mathbf{x}(\mathbf{p}, Y)} \tag{17}$$

In words, the bundle that yields the highest welfare under prices $\mathbf{p}$ and income Y is the same as the bundle which minimizes spending when facing prices $\mathbf{p}$ and aiming for the utility goal $v(\mathbf{p}, Y)$. We can heuristically verify this statement through an **"ϵ-perturbation" argument**. Suppose $\mathbf{h}(\mathbf{p}, v(\mathbf{p}, Y)) \neq \mathbf{x}(\mathbf{p}, Y)$. This means there exists some bundle, namely $\mathbf{h}(\mathbf{p}, v(\mathbf{p}, Y))$, that achieves the same utility as $\mathbf{x}(\mathbf{p}, Y)$ but is strictly cheaper[10]. Construct a new bundle by adding ϵ units of consumption to every coordinate of $\mathbf{h}(\mathbf{p}, v(\mathbf{p}, Y))$. For small enough ϵ, this new bundle remains affordable under $(\mathbf{p}, Y)$, yet yields strictly higher utility than $\mathbf{x}(\mathbf{p}, Y)$ provided u is strictly increasing. This contradicts the fact that $\mathbf{x}(\mathbf{p}, Y)$ is the solution to **UMP**.

The "converse" relation to equation (17) also holds if we start with an **EMP** parametrized by $(\mathbf{p}, \bar{u})$ and try to "line up" a **UMP** with respect to it. In order for the **UMP** to yield the same solution, the prices must again match and we must have $Y = e(\mathbf{p}, \bar{u})$ (the income parameter equals minimized objective of **EMP**). This is stated more succinctly as:

[9] This week's section material borrows from the work of previous TFs, in particular the notes of Zhenyu Lai. Image credit: Paolo Crosetto (http://paolocrosetto.files.wordpress.com/2010/10/pset4_solution_handout.pdf)

[10] You might worry that $\mathbf{h}(\mathbf{p}, v(\mathbf{p}, Y))$ could cost the same as $\mathbf{x}(\mathbf{p}, Y)$. But $\mathbf{h}(\mathbf{p}, v(\mathbf{p}, Y))$ costing exactly Y would mean it is a second solution to **UMP**. Under regularity conditions, **UMP** has exactly one solution.

$$\boxed{\mathbf{x}(\mathbf{p}, e(\mathbf{p}, \bar{u})) = \mathbf{h}(\mathbf{p}, \bar{u})} \tag{18}$$

To convince yourself of this, suppose $\mathbf{x}(\mathbf{p}, e(\mathbf{p}, \bar{u})) \neq \mathbf{h}(\mathbf{p}, \bar{u})$. This means there exists some bundle, namely $\mathbf{x}(\mathbf{p}, e(\mathbf{p}, \bar{u}))$, that costs the same as $\mathbf{h}(\mathbf{p}, \bar{u})$ but yields strictly more utility. Construct a new bundle by removing ϵ units of consumption from every coordinate of $\mathbf{x}(\mathbf{p}, e(\mathbf{p}, \bar{u}))$. For small enough ε, this new bundle continues to deliver more than $\bar{u}$ utility, yet costs strictly less than $\mathbf{h}(\mathbf{p}, \bar{u})$. This contradicts the fact that $\mathbf{h}(\mathbf{p}, \bar{u})$ is the solution to **EMP**.

1.2 *"Duality" in firm's problems.* The firm's profit maximization problem (**PMP**) and cost minimization problem (**CMP**) exhibit a kind of parallel similar to the duality in consumer's problems. Recall the two problems of the firm:

$$\max_{K,L \geq 0} pf(K, L) - rK - wL \tag{\textbf{PMP}}$$

$$\min_{K,L \geq 0} rK + wL \text{ s.t. } f(K, L) \geq q \tag{\textbf{CMP}}$$

Write $K_\pi^*(p, r, w)$ and $L_\pi^*(p, r, w)$ for the profit-maximizing capital and labor demand arising from **PMP**. Write $K_C^*(q, r, w)$ and $L_C^*(q, r, w)$ for the expenditure-minimizing capital and labor demand arising from **CMP**. Then:

$$K_\pi^*(p, r, w) = K_C^*(Q^*(p, r, w), r, w)$$

$$L_\pi^*(p, r, w) = L_C^*(Q^*(p, r, w), r, w)$$

where $Q^*(p, r, w) = f(K_\pi^*(p, r, w), L_\pi^*(p, r, w))$ refers to the profit-maximizing level of output. To interpret, suppose we start with **PMP** under prices (p, r, w). Then the **CMP** yields the same factor demand if its parameters are "aligned" with **PMP** in a suitable sense. In particular, this happens when **CMP** has same input prices as **PMP** and the target quantity is precisely the profit-maximizing quantity from **PMP**.

Unfortunately no analogous "converse" relationship exists.

2 The Slutsky equation

2.1 *Slutsky equation as an implication of duality.* Start with the **duality relation** in (18):

$$x_i(\mathbf{p}, e(\mathbf{p}, \bar{u})) = h_i(\mathbf{p}, \bar{u})$$

Differentiating both sides by p_j obtains:

$$\frac{\partial x_i}{\partial p_j} + \frac{\partial x_i}{\partial Y} \cdot \frac{\partial e}{\partial p_j}(\mathbf{p}, \bar{u}) = \frac{\partial h_i}{\partial p_j}$$

All partial derivatives are **functions**, but we especially highlight that $\frac{\partial e}{\partial p_j}$ is evaluated at $(\mathbf{p}, \bar{u})$ as this fact will prove important in the sequel.

Shephard's lemma implies $\frac{\partial e}{\partial p_j} = h_j$, so above equality is equivalent to:

$$\frac{\partial x_i}{\partial p_j} + \frac{\partial x_i}{\partial Y} \cdot h_j(\mathbf{p}, \bar{u}) = \frac{\partial h_i}{\partial p_j}$$

But using the other **duality relation** from (17) to rewrite $h_j(\mathbf{p}, \bar{u}) = x_j(\mathbf{p}, e(\mathbf{p}, \bar{u}))$, we deduce:

$$\frac{\partial x_i}{\partial p_j} + \frac{\partial x_i}{\partial Y} \cdot x_j(\mathbf{p}, e(\mathbf{p}, \bar{u})) = \frac{\partial h_i}{\partial p_j}$$

The arguments are often suppressed, leading to the celebrated **Slutsky equation**:

$$\boxed{\frac{\partial x_i}{\partial p_j} = \frac{\partial h_i}{\partial p_j} - \frac{\partial x_i}{\partial Y} \cdot x_j} \tag{19}$$

2.2 Geometric interpretation of the Slutsky decomposition. The Slutsky equation is also called the **Slutsky decomposition**. This is because the equation breaks down the response of Marshallian demand for good i to a change in price of good j into two additive pieces: the **substitution effect** $\left(\frac{\partial h_i}{\partial p_j}\right)$ and the **income effect** $\left(-\frac{\partial x_i}{\partial Y} \cdot x_j\right)$.

It is perhaps easier to visualize these two effects in a picture. Consider a world of two goods, x and y. Let us analyze how the Marshallian demand for x changes when its price drops from p_x to p'_x. Normally, we would proceed in 4 steps:

1. Draw the old budget constraint under p_x.
2. Draw the utility level curve tangent to the old budget constraint, thereby obtaining $x(p_x)$.
3. Draw the new budget constraint under p'_x.
4. Draw the utility level curve tangent to the new budget constraint, thereby obtaining $x(p'_x)$.

In this usual sequence of steps, we **jump directly** from $x(p_x)$ to $x(p'_x)$. However, Slutsky equation allows us to decompose this change in Marshallian demand into a **two parts** by filling in an **intermediate step** between steps 2 and 3.

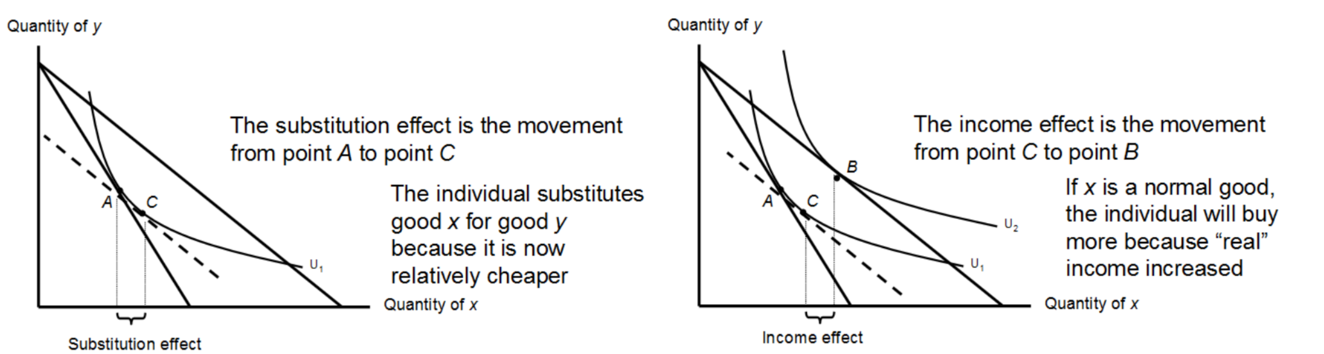


Figure 6: An illustration of the substitution effect and income effect.

Consider a "**step** $2\frac{1}{2}$" where the price of x has changed to p'_x but the consumer is trying to solve **EMP** instead of **UMP**. The desired utility level in this **EMP** is the consumer's old utility level (denote it by $U_1 \in \mathbb{R}$) before the price change. This amounts to finding a point of tangency between the old utility level curve and a budget line with the new slope given by p'_x. Note that under the

old prices, $x(p_x)$ **was** the Hicksian demand for utility level U_1. Thus, for this new **EMP**, the **departure** of its solution from $x(p_x)$ corresponds to $\frac{\partial h_x}{\partial p_x}$. We interpret this partial derivative as the "substitution effect", for it captures the part of the change in Marshallian demand that originates from a substitution towards the recently cheaper x in the **EMP** that holds old utility fixed.

Moving from step $2\frac{1}{2}$ to step 3 involves **shifting out** the budget line that solved **EMP** to a new, parallel budget line that reflects the consumer's actual income. The **EMP** budget line was at an incorrect level of income, since the change of p_x to p'_x carries also the side-effect of making the consumer richer. In particular, it is as if the consumer has **gained** $(p_x - p'_x)$ dollars **for every** unit of x he was consuming before the price change. This extra wealth causes the consumer to consume more of every normal good. In particular, how much more x he consumes as a result of this income effect depends both on his previous level of x consumption (which determines the extent of wealth increase due to price change) and how Marshallian demand of x varies with income. Put together, $-\frac{\partial x_i}{\partial Y} \cdot x_j$ is the income effect.

2.3 *Relationships between **UMP** and **EMP**.* Figure 7 collects together some relationships between functions associated with **UMP** and **EMP**.

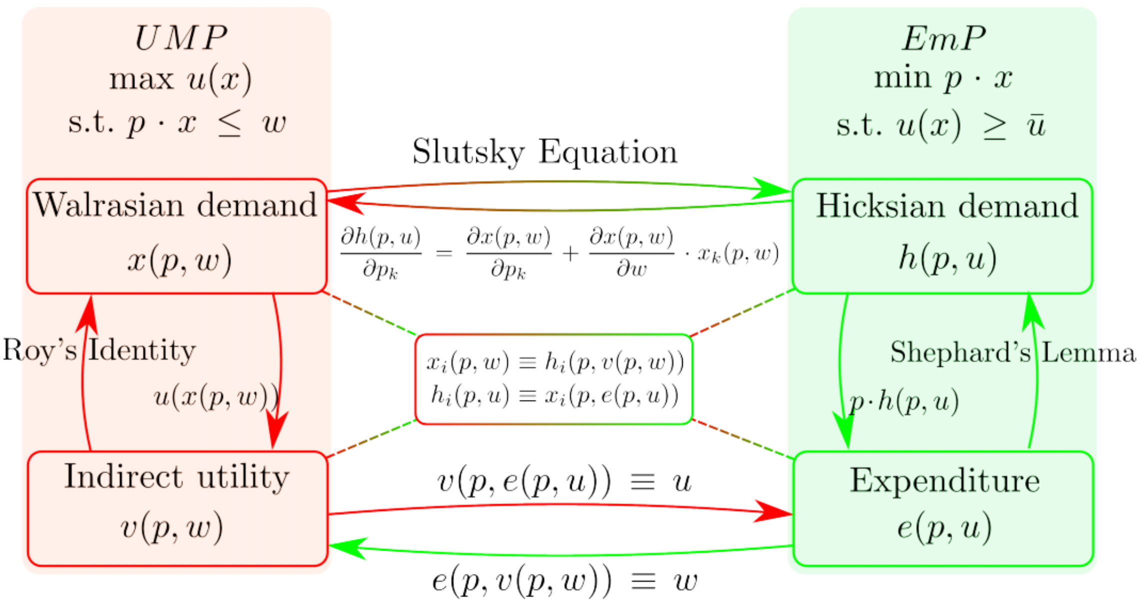


Figure 7: Some important relationships between **UMP** and **EMP**. Shamelessly taken from Paolo Crosetto's website, which uses w instead of Y to denote consumer's wealth.

3 Identities from the budget constraint

3.1 *Some notations.*

- $\eta_i := \frac{p_i x_i}{Y}$ is the **budget share** of good i
- $\varepsilon^i_{p_j} := \frac{\partial x_i}{\partial p_j} \frac{p_j}{x_i}$ is the **elasticity** of good i with respect to the price of good j
- $\varepsilon^i_Y := \frac{\partial x_i}{\partial Y} \frac{Y}{x_i}$ is the **elasticity** of good i with respect to income Y

In general, the **elasticity** of some variable X with respect to one of its arguments z is defined as $\frac{\partial X}{\partial z} \frac{z}{X}$. Elasticity is meant to formalize "how much does X change in percentage terms when z changes by 1%"? Compared to the partial derivative $\frac{\partial X}{\partial z}$, the metric $\frac{\text{percent change in } X}{\text{percent change in } z}$ is **unit-free**. If X were re-denominated in a different unit (pounds to kg, dollars to cents, etc.), then its elasticity with respect to z would not change but $\frac{\partial X}{\partial z}$ would.

To connect the motivation of elasticity with its definition, consider the elasticity of good i with respect to the price of good j. Suppressing dependence of Marshallian demand x_i on arguments other than p_j, consider a small change $\Delta \in \mathbb{R}$ to p_j. We form the ratio of percent changes:

$$\left(\frac{x_i(p_j+\Delta)-x_i(p_j)}{x_i(p_j)}\right) / \left(\frac{(p_j+\Delta)-p_j}{p_j}\right)$$

... and take the **limit** of this expression as $\Delta \to 0$:

$$\begin{aligned}
\lim_{\Delta\to 0}\left[\left(\frac{x_i(p_j+\Delta)-x_i(p_j)}{x_i(p_j)}\right) / \left(\frac{(p_j+\Delta)-p_j}{p_j}\right)\right] &= \lim_{\Delta\to 0}\left[\left(\frac{x_i(p_j+\Delta)-x_i(p_j)}{x_i(p_j)}\right)\cdot\left(\frac{p_j}{(p_j+\Delta)-p_j}\right)\right] \\
&= \lim_{\Delta\to 0}\left[\frac{x_i(p_j+\Delta)-x_i(p_j)}{\Delta}\right]\cdot\frac{p_j}{x_i(p_j)} \\
&= \frac{\partial x_i}{\partial p_j}\cdot\frac{p_j}{x_i}
\end{aligned}$$

which is the definition of $\varepsilon^i_{p_j}$.

3.2 *Useful identities.* Since the Marshallian demand is a function of $(\mathbf{p}, Y)$, substituting it into the budget constraint yields a relation that holds for all $(\mathbf{p}, Y)$.

$$\sum_{i=1}^{n} p_i x_i(\mathbf{p}, Y) = Y$$

This relation can be **differentiated** with respect to different arguments of x_i to obtain a number of identities.

Differentiating with respect to Y gives:

$$\begin{aligned}
\sum_{i=1}^{n} p_i \frac{\partial x_i}{\partial Y} &= 1 \\
\sum_{i=1}^{n} \frac{p_i x_i}{Y}\frac{\partial x_i}{\partial Y}\frac{Y}{x_i} &= 1
\end{aligned}$$

which is to say,

$$\boxed{\sum_{i=1}^{n} \eta_i \varepsilon^i_Y = 1} \tag{20}$$

In words, the average of income elasticities, **weighted by budget shares**, must add up to 1. To convince yourself this makes sense, imagine the consumer receives an **unexpected windfall** that increases her income by **1%**. This causes her to increase expenditure on item i by ε^i_Y percent relative to current spending level, which translates into an increase of $\eta_i \varepsilon^i_Y$ percent in total expenditure. (The weighing makes sense since a percentage-based increase in expenditure in a consumption category that makes up very little of her spending share does not change total expenditure by much.) The identity is saying the percent increase in total expenditure must equal the 1% increase in income.

If we instead differentiated both sides with respect to p_j, we would get:

$$
\begin{aligned}
x_j + \sum_{i=1}^{n} p_i \frac{\partial x_i}{\partial p_j} &= 0 \\
\frac{p_j x_j}{Y} &= -\sum_{i=1}^{n} p_i \frac{p_j}{Y} \frac{\partial x_i}{\partial p_j} \\
\eta_j &= -\sum_{i=1}^{n} \frac{p_i x_i}{Y} \frac{\partial x_i}{\partial p_j} \frac{p_j}{x_i}
\end{aligned}
$$

using compact notation, this says:

$$
\boxed{\eta_j = -\sum_{i=1}^{n} \eta_i \varepsilon^i_{p_j}} \tag{21}
$$

In words, the budget share of good j is the negative of average elasticity of different goods with respect to price of j, **weighted by their own budget shares**. To understand this identity intuitively, imagine a consumer facing a sudden 1% increase in p_j. Since she was spending η_j of her total income on good j, it is as if she has **suddenly lost** η_j percent of her income. To **re-balance** her budget, she must cut consumption across the board. In particular, she changes consumption of good i by $\varepsilon^i_{p_j}$ percent, which translates into a $\eta_i \varepsilon^i_{p_j}$ percent change in total expenditure. The identity is saying she successfully re-balances her budget, since the sum of percent changes in her expenditure on the right hand side amounts to exactly η_j.

Economics 1011a ▷ Section 5 : Time and Preferences[11] — 10/7/2014

(1) Exponential discounting; **(2)** Quasi-hyperbolic discounting; **(3)** Axioms of time discounting

TF: Kevin He (he02@fas.harvard.edu)

1 Exponential time discounting

1.1 *Exponential time discounting from a standard UMP point of view.* Marty has a two-period intertemporal utility function with exponential discounting

$$U(c_1, c_2) = u(c_1) + \delta u(c_2)$$

where $\delta \in (0,1)$. He has income y_1 in period 1, income y_2 in period 2, and can save or borrow at an interest rate of $r > 0$ between periods. Every unit of consumption costs \$1 in each period.

It is instructive to "**flatten**" Marty's two-period intertemporal optimization problem into a standard two goods problem that we have been studying. The net present value of Marty's income is $Y := y_1 + \frac{y_2}{(1+r)}$ while the price of c_1 and c_2 in period 1 dollars are $p_1 := 1$ and $p_2 := \frac{1}{1+r}$ respectively. So Marty's problem is equivalent to:

$$\max_{c_1,c_2} \; U(c_1,c_2) \text{ s.t. } p_1 c_1 + p_2 c_2 \le Y$$

which is a very familiar **UMP**. To interpret, Marty can simply **"forget" the time structure of the problem** and compute everything in terms of period 1 dollars as he makes his consumption plans. In this **UMP** re-formulation, equating the ratio of marginal utilities of c_1 and c_2 with the ratio of their prices yields:

$$\frac{U_{c_1}(c_1^*, c_2^*)}{U_{c_2}(c_1^*, c_2^*)} = \frac{p_1}{p_2}$$

But since $U(c_1, c_2) = u(c_1) + \delta u(c_2)$, $p_1 = 1$, $p_2 = \frac{1}{1+r}$, this expression rearranges to:

$$\frac{u'(c_1^*)}{\delta u'(c_2^*)} = \frac{1}{1/(1+r)}$$

equivalently,

$$\boxed{u'(c_1^*) = (1+r)\delta u'(c_2^*)} \tag{22}$$

which is the **Euler equation**!

We have thus transformed a two-period intertemporal optimization problem into a standard **UMP** that suppresses any mention of time. This transformation extends readily into T-periods. Indeed, a T-period intertemporal optimization problem is formally equivalent to a T-goods **UMP**, where the consumer solves:

[11] Breaking with tradition, this week's section material did not borrow from the work of previous TFs, (not even the notes of Zhenyu Lai.)
Optional exercise: This handout contains several examples with named characters. What do these characters share in common?

$$\max_{c_1, c_2, \ldots, c_T} \; U(c_1, c_2, ..., c_T) \text{ s.t. } \sum_{t=1}^{T} p_t c_t \leq Y$$

Here, U has the usual exponential discounting functional form with $\delta \in (0,1)$,

$$U(c_1, c_2, ..., c_T) = \sum_{t=1}^{T} \delta^{t-1} u(c_t)$$

Y is the net present value of T periods of income stream,

$$Y := \sum_{t=1}^{T} \frac{y_t}{(1+r)^{t-1}}$$

and each p_t is price of period t consumption phrased in terms of period 1 dollars,

$$p_t := \frac{1}{(1+r)^{t-1}}$$

Despite this formal connection, it is worth pointing out several important differences between the intertemporal optimization problem and an **UMP** with T-goods:

(1) The **parameters are different**. In the case of $T = 2$, the parameters of intertemporal optimization are r, y_1, y_2. The parameters of **UMP** are Y, p_1, p_2. In the **UMP**s we studied previously, we could ask what would happen if price p_2 changes. But in the **UMP** generated by intertemporal optimization, the **UMP** parameters are **functions of intertermporal parameters**. So, the only way to change p_2 is to change r, which will also cause a change in Y. This is why comparative statics with respect to r in an intertemporal problem tend to be very messy. Changing r is equivalent to simultaneous changing two parameters in the associated **UMP**.

(2) Intertemporal optimization makes **functional form assumptions on** U. All sorts of functional forms for U are acceptable for an unrestricted **UMP**, but the **UMP** generated by intertemporal optimization requires that $U(\mathbf{c}) = \sum_{t=1}^{T} \delta^{t-1} u(c_t)$. This requires **additive separability** across dimensions of $\mathbf{c}$, so that changing the level of consumption in one period does not affect the utility derived in another period. Further, the T additive functions evaluating the T dimensions of $\mathbf{c}$ are **time independent**, in that we have a common $u(\cdot)$ and not a time-varying $u_t(\cdot)$. This captures the assumption that the consumer's taste does not change over time.

1.2 *The discount factor* δ. What relates consumption utility between different periods is a discount factor, usually assumed to be valued in the open interval $(0,1)$. We discuss several justifications for this assumption.

The first justification is **psychological**. Many laboratory studies confirm that humans are impatient and prefer getting a positive outcome sooner rather than later.

The second justification is **technical**. In many economic models, we want to consider an **infinite time horizon**. That is, we need to let $T = \infty$. In this case, $\delta \in (0,1)$ helps ensure the summability of infinitely many utility terms:

$$U(c_1, c_2, c_3, ...) = \sum_{t=1}^{\infty} \delta^{t-1} u(c_t)$$

For instance, if the per-period utility $u : \mathbb{R} \to \mathbb{R}$ is bounded with $0 \leq u(c_t) \leq M < \infty$ for any argument c_t, then the infinite series on right hand side converges when $\delta \in (0,1)$ by comparison to a geometric series.

We can offer a third justification that re-interprets discounting in the framework of **survival analysis**. Consider a decision-maker Zhang Xiao who will potentially live for T periods. However, each period starting from $t = 2$ she suffers a $(1-\delta)$ chance of dying. These chances are independent, so it is as if she tosses a (biased) coin at the start of each period $t \geq 2$ to decide if she survives. If she dies at the beginning of period t, she does not gain any utility from her planned consumption in period t or any future periods. How should Xiao value a consumption plan $(c_1, c_2, ..., c_T)$?

It is clear that she gets utility from period t consumption if and only if she succeeds in all of her first $(t-1)$ tosses of the coin. Let X be the random variable denoting how many heads in a row Xiao gets until the first tail when she tosses the biased coin. Her expected utility from the consumption plans $(c_1, c_2, ..., c_T)$ is thus:

$$\mathbb{E}\left[1_{\{X\geq 0\}}u(c_1) + 1_{\{X\geq 1\}}u(c_2) + 1_{\{X\geq 2\}}u(c_3) + ... + 1_{\{X\geq T-1\}}u(c_T)\right]$$

where $1_{\{X\geq s\}}$ is an indicator random variable that equals 1 if $X \geq s$, equals 0 otherwise. By linearity of expectation, rewrite above as:

$$\sum_{t=1}^{T}\mathbb{E}\left[1_{\{X\geq t-1\}}\right]u(c_t)$$

but $\mathbb{E}\left[1_{\{X\geq t-1\}}\right] = \text{Prob}[X \geq t-1] = \delta^{t-1}$, so that her expected utility is:

$$\sum_{t=1}^{T}\delta^{t-1}u(c_t)$$

which is the formula for exponential discounting. Therefore, we can think of $1-\delta$, the complement to the discount factor, as a **hazard rate** for the decision-maker.

Must we always have $\delta < 1$? In this course and in most areas of economics, $\delta < 1$ is a **very standard assumption**. But some authors in behavioral economics have studied the possibility of $\delta > 1$. One situation where later consumption could be valued more highly than earlier consumption involves **anticipation**. Suppose you must decide when to go on a one-day vacation in a given month $(t = 1, 2, ..., 30)$. Assume being on vacation gives you positive utility, but so does anticipating a future vacation. In this case you might decide to delay the vacation as much as possible in the month, so that you can enjoy more anticipatory utility on the days leading up to it. That is, you behave as if $\delta > 1$. (For this story to make sense, however, anticipation must be "stronger" than impatience. Further, we are assuming away any utility the decision-maker might get from the memory of being on vacation.)

1.3 *Time discounting in the presence of risk.* We have seen two lectures on intertemporal optimization and will switch gears to risk preference on Thursday. We can put these two subjects together by studying a model of consumption in **two periods that also involves risk**.

So far, we have assumed the decision-maker foresees their own future income exactly at time $t = 1$. This is unrealistic as the future is often uncertain. Consider Okabe who lives for two periods and has intertemporal utility function

$$U(c_1, c_2) = u(c_1) + \delta u(c_2)$$

where $\delta \in (0, 1)$. He has income y_1 in period 1, which he knows for certain. However, he does **not** know his income in the future period, which he views as a random variable $\tilde{y}_2$ with $\mathbb{E}[\tilde{y}_2] = y_1$.

As usual, Okabe can save or borrow at an interest rate of $r > 0$ between periods. Every unit of consumption costs \$1 in each period.

Suppose Okabe consumes c_1 in period 1. Then his period 2 consumption is

$$\tilde{c}_2 = \tilde{y}_2 + (1+r)(y_1 - c_1)$$

Notice $\tilde{c}_2$ is also a random variable, since it depends on the random future income. The optimization problem is:

$$\max_{c_1} \ \mathbb{E}\left[u(c_1) + \delta u(\tilde{y}_2 + (1+r)(y_1 - c_1))\right]$$

Taking first order condition obtains[12]:

$$\mathbb{E}\left[u'(c_1^*) - (1+r)\delta u'(\tilde{c}_2^*)\right] = 0$$

Since c_1, $1-r$, and δ are not random, we can rearrange:

$$u'(c_1^*) = (1+r)\delta \mathbb{E}\left[u'(\tilde{c}_2^*)\right]$$

This is a **stochastic** version of the **Euler equation**!

To focus on the effect of risky income on consumption decisions and abstract away from other parts of the model, assume $(1+r)\delta = 1$, so that the Euler equation simplifies to

$$u'(c_1^*) = \mathbb{E}\left[u'(\tilde{c}_2^*)\right]$$

In a counterfactual world where Okabe's future income is not random, but fixed at $\mathbb{E}[\tilde{y}_2] = y_1$, he would choose $c_1^* = c_2^* = y_1$. But now suppose $u'(\cdot)$ is a **strictly convex function**. (This would happen if you assume $u(c) = \ln(c)$ or $u(c) = \sqrt{c}$, for example.) Could it still be the case that $c_1^* = y_1$? No, because if Okabe made that decision then

$$\mathbb{E}[\tilde{c}_2^*] = \mathbb{E}[\tilde{y}_2] = y_1 = c_1^*$$

Yet **Jensen's inequality** would imply:

$$u'(c_1^*) = u'(\mathbb{E}[\tilde{c}_2^*]) < \mathbb{E}\left[u'(\tilde{c}_2^*)\right]$$

so Euler's equation fails to hold. Indeed, Jensen's inequality says the marginal utility in period 1 is too low relative to the optimum, which means consuming all of y_1 in period 1 was consuming too much. Instead, the optimal solution would involve Okabe **saving some money in period 1**. This phenomenon where future uncertainty leads a consumer to save more today is called "**precautionary savings effect**". It is one way to model the observation that many decision-makers keep more money "saved for a rainy day" than a standard model with income certainty would predict.

[12] Assuming we can exchange differentiation and integration.

2 Quasi-hyperbolic time discounting

2.1 Present bias and time inconsistency. To motivate the **quasi-hyperbolic time discounting** model, we begin by investigating two empirical puzzles not explained by exponential time discounting.

The first puzzle is **present bias**. In exponential time discounting, the amount of discounting between two neighboring periods t and $t+1$ is constant for all $t = 1, 2, 3, ..., T-1$. Yet this is empirically false as many people exhibit significant discounting between "**today**" and "**tomorrow**", but far less discounting between "**tomorrow**" and "**the day after**". As a concrete illustration, let each period be one hour and consider a choice between the two alternatives:

1a. Eat a snack now.

1b. Eat a better snack in 1 hour.

Now consider instead a choice between these two other alternatives:

2a. Eat a snack in 14 days and 10 hours.

2b. Eat a better snack in 14 years and 11 hours.

A significant fraction of people choose **1a** over **1b,** but almost no one chooses **2a** over **2b**. But this **cannot be rational under exponential discounting**, where the discount factor between $t = 1$ and $t = 2$ is equal to the the discount factor between $t = 346$ and $t = 347$. On the other hand, this behavior does make psychological sense. A decision-maker might make a very sharp distinction between time periods close to the present. But as she contemplates time periods further and further into the future, she might not feel there is such a large gap between those distant neighboring time periods.

The second puzzle is **time inconsistency**, which refers to the decision-maker setting one set of plans about future consumption but failing to follow her own plans if given the chance to change her mind later. Time inconsistency does not exist in an exponential discounting model. Indeed, this is what allowed Marty to plan out all of his consumptions in period 1, as he knows he will not wish to re-optimize in later time periods.

It turns out most real-life consumers do not behave like Marty. In 1998, two researchers conducted a food experiment where they asked subjects to choose what snack they want to receive when they return to the lab **in one week**. 74% chose a fruit while 26% chose chocolate. When the same subjects returned the following week, the experimenters claimed that the subjects' snack choice data from a week ago was lost. The subjects were asked to choose again and they would receive their choice **immediately**. This time, 30% chose a fruit while 70% chose chocolate. It was as if the subjects were making choices about future consumption using one preference but making choices about present consumption using a different preference.

The quasi-hyperbolic discounting model is a generalization of exponential discounting model that can explain present bias and time inconsistency.

$$\boxed{U(c_1, c_2, ..., c_T) = u(c_1) + \beta\left(\delta u(c_2) + \delta^2 u(c_3) + ... + \delta^{T-1} u(c_T)\right) = u(c_1) + \beta \sum_{t=2}^{T} \delta^{t-1} u(c_t)} \tag{23}$$

Where $\delta \in [0,1)$ and $\beta \in [0,1]$. If $\beta = 1$ then we are back in the exponential discounting model. To gain intuition, suppose $\delta \approx 1$ while $\beta = \frac{1}{2}$. Then:

$$U(c_1, c_2, ..., c_T) \approx u(c_1) + \frac{1}{2}u(c_2) + \frac{1}{2}u(c_3) + ... + \frac{1}{2}u(c_T)$$

The **present bias** is clear in this extreme case. Most of the discounting occurs between $t = 1$ and $t = 2$, while little discounting occurs between future periods.

Importantly, the trade-off between $t = 2$ and $t = 3$ depends on whether the agent is currently at time $t = 1$ or time $t = 2$, so one might imagine plans about future consumption need not be carried through, leading to **time inconsistency**. The next example will flesh this out in more detail.

2.2 *Procrastination and time inconsistency: a primer.* Assume an infinite horizon, $T = \infty$. Suppose $\beta = \frac{1}{2}$ and $\delta = 0.99$.

Consider an action which carries some **immediate cost** but brings **future gains**. Examples include: cleaning your apartment, physical exercise, working on a problem set. We will stick with physical exercise, though the quasi-hyperbolic discounting model makes the same prediction in each of these situations.

Suppose Aaron needs to choose **exactly one day** in which to undertake **physical exercise**. Exercise is costly (time cost, physical exhaustion, etc.) and he gets -4 utility on the day of the exercise. However, the day after he collects +6 utility from delayed benefits of exercise. Utility on all other days are 0.

At $t = 1$, Aaron will **plan to exercise** on $t = 2$. This is because exercising today yields

$$-4 + \beta(\delta \cdot 6 + \delta^2 \cdot 0 + \delta^3 \cdot 0 + ...) \approx -1$$

Exercising on $t = 2$ yields:

$$0 + \beta(\delta \cdot (-4) + \delta^2 \cdot 6 + \delta^3 \cdot 0 + ...) = 3\delta^2 - 2\delta \approx 0.96$$

Exercising on $t = 3$ yields:

$$0 + \beta(\delta \cdot 0 + \delta^2 \cdot (-4) + \delta^3 \cdot (6) + \delta^4 \cdot 0 + ...) = \delta \cdot \left(3\delta^2 - 2\delta\right) \approx 0.95$$

And so forth. Exercising today is a bad idea, as Aaron feels the cost today is not worth the (heavily discounted) benefits tomorrow. But exercising tomorrow sounds good. This is because there is **much less discounting** between tomorrow and the day after tomorrow versus between today and tomorrow. Indeed, from the perspective of the Aaron at $t = 1$, the effective discount rate between $t = 2$ and $t = 3$ is 0.99, which makes the trade-off of -4 one period for +6 the next period worthwhile. Putting off exercise later than $t = 2$ is also a bad idea, since Aaron is impatient and would rather receive the net benefit of $-4 + 0.99 \cdot 6$ sooner than later.

However, even though he made plans to exercise on $t = 2$, when $t = 2$ rolls around Aaron **will not actually exercise**. In fact, from the perspective of $t = 2$, exercising yields:

$$-4 + \beta(\delta \cdot 6 + \delta^2 \cdot 0 + \delta^3 \cdot 0 + ...) \approx -1$$

But, exercising on $t = 3$ sounds pretty good to the Aaron at $t = 2$, because this plan yields:

$$0 + \beta(\delta \cdot (-4) + \delta^2 \cdot 6 + \delta^3 \cdot 0 + ...) = 3\delta^2 - 2\delta \approx 0.96$$

So he delays his original plan of exercising by one more day. What happened? The definitions of "today" and "tomorrow" changed as **time advanced**. Though there was little discounting between $t = 2$ and $t = 3$ when plans were made on $t = 1$, when $t = 2$ comes around the discounting between these two time periods suddenly grows larger. Now, the β discount factor is no longer applied to time 2 utility, but it is still applied to time 3 utility.

In fact, this behavior of delaying plans continues indefinitely, to the point that Aaron never actually exercises. It is the time-inconsistency of his preference that leads to procrastination.

3 Axiomatization of time discounting (optional)

3.1 *"Axiomatization"?* In section notes 3, we discussed the relationship between preferences and utility functions. In particular, we emphasized that **preferences are primitive** while utility functions are just convenient numerical summaries that represent preferences. This philosophy applies to **any class** of utility functions. In particular, it applies to utility functions over intertemporal consumption studied in this note.

Formally speaking, $U(c_1, c_2) = u(c_1) + \delta u(c_2)$ is just the **utility representation** of some underlying preference relation $\succsim$ on $\mathbb{R} \times \mathbb{R}$, where $(x, y) \succsim (x', y')$ means the decision-maker prefers getting x in period 1 and y in period 2 over x' in period 1 and y' in period 2. At the same time, it should be clear that the functional form $U(c_1, c_2) = u(c_1) + \delta u(c_2)$ places some **restrictions** on its associated $\succsim$. For example, if $\delta < 1$, then the **impatience** exhibited by U should also be visible in $\succsim$. For any $x, y \in \mathbb{R}$ with $x > y$, it should be that $(x, y) \succ (y, x)$. The natural question to ask is whether we can find the set of **all** restrictions that the functional form imposes on preference, thus completely characterize the set of preference relations on $\mathbb{R} \times \mathbb{R}$ which are associated with discounting between two periods.

In general, when economists talk about "**axiomatizing**" a utility representation, they mean finding a set of necessary and sufficient conditions for when a preference $\succsim$ admits a utility representation in the desired functional form. These conditions on $\succsim$, often called **axioms**, help foreground assumptions about human behavior hidden within the utility function. Once a functional form is axiomatized, questioning whether it is a reasonable model of consumer behavior becomes equivalent to questioning the reasonableness of its associated axioms.

We have already seen an instance of axiomatization – Debreu's theorem! Recall that Debreu's theorem says a preference $\succsim$ on $X \subseteq \mathbb{R}^n$ is complete, transitive, and continuous if and only if there exists a continuous u that represents $\succsim$. The functional form being axiomatized is "continuous utility function" and the collection of axioms on $\succsim$ are completeness, transitivity, and continuity. Granted, continuous utility functions is a very "broad" class of functional forms. We shall see, however, that we can do the same exercise for exponential time discounting and quasi-hyperbolic time discounting.

3.2 *Axiomatization of exponential time discounting.* Suppose there are T periods numbered $\{1, 2, ..., T\}$ for $2 < T < \infty$.[13] The domain of preference is $\mathbb{R}^T$. Each member of $\mathbb{R}^T$ is viewed as a **dated consumption plan**, whose t^{th} coordinate tells us how much will be consumed in period t. (There is no money or budget in this world as we are only looking at the consumer's preference, not her intertemporal optimization problem.) Consider a preference $\succsim$ on $\mathbb{R}^T$. We need a collection of axioms on $\succsim$ which are necessary and sufficient for the existence of a utility function representation in the form of exponential time discounting:

[13] There exist unfortunate technical difficulties in the $T = 2$ case which make the axiomatization messier.

$$U(c_1, c_2, ..., c_T) = \sum_{t=1}^{T} \delta^{t-1} u(c_t) \tag{24}$$

where $\delta \in (0,1)$ and $u : \mathbb{R} \to \mathbb{R}$ is continuous and non-constant. Consider this collection of five axioms:

(**A1**) *Initial sensitivity*: There exist $a, b \in \mathbb{R}$, $\mathbf{x} \in \mathbb{R}^{T-1}$ such that $(a, x_1, x_2, ..., x_{T-1}) \succ (b, x_1, x_2, ..., x_{T-1})$.

(**A2**) *Initial separability*: For any $a, b, c, d \in \mathbb{R}$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{T-2}$,

$$(a, b, x_1, x_2, ..., x_{T-2}) \succsim (c, d, x_1, x_2, ..., x_{T-2})$$

if and only if

$$(a, b, y_1, y_2, ..., y_{T-2}) \succsim (c, d, y_1, y_2, ..., y_{T-2})$$

(**A3**) *Stationarity*: For any $a \in \mathbb{R}$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{T-1}$, $(a, x_1, x_2, ..., x_{T-1}) \succsim (a, y_1, y_2, ..., y_{T-1})$ if and only if $(x_1, x_2, ..., x_{T-1}, a) \succsim (y_1, y_2, ..., y_{T-1}, a)$.

(**A4**) *Impatience*: If $a, b \in \mathbb{R}$ with $(a, a, ..., a) \succ (b, b, ..., b)$, then $(a, c, c, ..., b) \succ (b, c, c, ..., a)$ for any $c \in \mathbb{R}$.

(**A5**) *Regularity*: $\succsim$ is rational and continuous.

To interpret, (**A1**) ensures the consumer cares about what she gets in the first period. (**A2**) says the trade-off between the first two periods is not affected by consumption in later periods. (**A3**) stipulates the choice problem facing the consumer at $t = 2$ is "the same as" the problem she faced at $t = 1$. (**A4**) captures impatience. (**A5**) contains standard assumptions. It turns out these five axioms are necessary and sufficient.

Theorem. *Preference $\succsim$ on $\mathbb{R}^T$ satisfies axioms (**A1**) through (**A5**) if and only if it admits an exponential time discounting representation with $\delta \in (0,1)$.*

Do not attempt to prove the sufficiency direction of this theorem, as the argument relies on an axiomatization of separable utility functions which we have not covered.

In light of this theorem, any consumer whose behavior is not captured by an exponential time discounting model must violate one or more of the axioms (**A1**) through (**A5**)! If we were asked to test whether a consumer really has exponential time discounting preferences, we need only test these five axioms.

Exercise A. Verify that axioms (**A1**) through (**A5**) are necessary. That is, every exponential time discounting utility function with $\delta \in (0,1)$ must satisfy these axioms.

Exercise B. Which of the axiom(s) (**A1**) through (**A5**) does a consumer with quasi-hyperbolic time discounting violate?

Exercise C. Dismount bar from rack and stand with shoulder width stance. Bend hips back while allowing knees to bend forward, keeping back straight and knees pointed same direction as feet. Descend until thighs are just past parallel to floor. Extend knees and hips until legs are straight. Return and repeat.

Exercise D. Conjecture how the axioms must be modified to characterize the case of $\delta > 1$.

Exercise E. Conjecture how the axioms must be modified and what sorts of axioms must be added to characterize the case of $T = \infty$.

3.3 *Axiomatization of quasi-hyperbolic time discounting.* David Laibson introduced quasi-hyperbolic discounting as a model of consumer intertemporal preferences in 1997[14]. This functional form has

[14] This functional form has appeared earlier as a model of inter-generational preferences, with each period representing one generation.

enjoyed enormous popularity in the economics profession, though its axiomatization long eluded us. The key obstacle relative to exponential discounting is the lack of stationarity. Indeed, the **present bias** of a quasi-hyperbolic consumer might generate the following preference:

$$(1, 0, 0, 0, ...) \succ (0, 2, 0, 0, ...)$$

and at the same time

$$(0, 1, 0, 0, 0, ...) \prec (0, 0, 2, 0, 0, ...)$$

The lack of stationarity destroys the recursive structure of the problem that allowed for simple axiomatization of exponential discounting.

Now for some very recent good news. In **2014**, a set of necessary and sufficient axioms for quasi-hyperbolic discounting was finally discovered by José Olea (NYU) and Tomasz Strzalecki (Harvard). Their characterization involves **10 axioms** and is a bit too technical for be presented here. However, it's nice to know such an axiomatization exists!

Economics 1011a ▷ Section 6 : Risk and Preferences[15] 10/14/2014

(1) How to model risk; **(2)** Risk aversion; **(3)** Paradoxes of expected utility

TF: Kevin He (he02@fas.harvard.edu)

1 How to model risk

1.1 *Abstract nonsense.* A model of **risk in economics** is often built from the following ingredients:

- A finite set S, called a "**state space**" or "**states of the world**". Since S is finite, enumerate it as $S = \{s_1, s_2, ..., s_n\}$.
- A **probability distribution** p on S, which is a list of positive numbers $p_1, p_2, ..., p_n$, where $p_k \geq 0$ for every k and $\sum_{k=1}^{n} p_k = 1$.
- A function $X : S \to \mathbb{R}$, called a **random variable**.

To interpret, S is an exhaustive, mutually exclusive enumeration of things that could happen in the world. Exactly one "state" in S must realize. For instance, in a betting game that involves rolling a 6-sided fair die, the state space might be written as:

$$S = \{\text{die shows 1, die shows 2, die shows 3, die shows 4, die shows 5, die shows 6}\}$$

Let's abbreviate $S = \{s_1, s_2, ..., s_6\}$ in the natural way.

There are usually multiple ways to write down the state space of any situation. For example, the state space of the die-rolling experiment can also be written as:

$$\hat{S} = \left\{ \begin{array}{c} \text{die shows 1, die shows 2, die shows 3 and it's snowing,} \\ \text{die shows 3 and it's not snowing, die shows 4, die shows 5, die shows 6} \end{array} \right\}$$

But for modeling purposes, $\hat{S}$ is just silly. Whether it's snowing or not does not affect the bet under consideration. In line with the mantra of always writing the simplest model possible, we want to use the smallest state space that still captures all the relevant features of the economic situation under consideration.

Next up is the probability distribution. With the state space $S = \{s_1, s_2, ..., s_n\}$ fixed, p_k represents the probability that state s_k realizes. In the example of die rolling,

$$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = \frac{1}{6}$$

Finally, a random variable is a **function** from the state space to the set of real numbers. For $s \in S$, think of $X(s)$ as the payoff when state s realizes. Exactly what the function X is depends on the

[15] Conspicuously absent this week is any discussion on the axiomatization of expected utility. This axiomatization involves two axioms and is much simpler than axiomatization of exponential discounting (5 axioms) or quasi-hyperbolic discounting (10 axioms) discussed last week. Therefore, those interested in the axiomatization of expected utility are encouraged to ask me in office hours or consult a standard textbook on economic theory, such as Chapter 5 of *Microeconomic Foundations I: Choice and Competitive Markets* by David Kreps.

bet. If the bet is "win \$100 if the die shows a prime number, win \$0 if the die shows a composite number, lose \$200 if the die shows 1", then

$$X(s_1) = -200, X(s_2) = 100, X(s_3) = 100, X(s_4) = 0, X(s_5) = 100, X(s_6) = 0$$

1.2 *How do you value a lottery?* For our purposes, a **lottery** is synonymous for a random variable. A lottery takes on different values depending on which state of the world realizes. The task of "valuing" a lottery, then, is about finding a utility function on lotteries:

$$U(X(\cdot))$$

Note the argument to function U is **another function**, namely $X : S \to \mathbb{R}$. In general one could imagine many ways to assign a "utility level" to each random variable X. For example, here are two wacky candidates:

- $U(X)$ is the smallest value that X can attain. If this is our utility function, then the lottery from betting on the die roll gets a utility level of -200.
- $U(X)$ is the difference between the largest value and the smallest value that X can attain. In this case the lottery is assigned a utility level of 300.

While they are technically utility functions over lotteries, these two candidates are "weird" for two reasons. One, they directly use the outcome of the lottery without considering how an individual "feels" about the monetary gains and losses. Two, they evaluate lotteries using a very narrow criterion and do not give a "holistic" judgment on the lottery.

The most classic theory of how an individual values a lottery is the **expected utility model**, which can be viewed as a **functional form restriction** on U:

$$\boxed{U(X) = \sum_{k=1}^{n} p_k \cdot u(X(s_k))} \tag{25}$$

where $u : \mathbb{R} \to \mathbb{R}$ is called a **Bernoulli utility function** as to avoid confusion with the utility function U over lotteries. To interpret, we apply u to transform the lottery payoff in each state into a utility level. Then we take a **weighted average** of these utility levels, where the weights are given by the probabilities of different states.

2 Risk aversion

2.1 *Who wants to be a millionaire?* Humans are typically **risk averse** over large-stake gains. To give a concrete illustration, imagine you are a contestant on the game show **Who Wants to Be a Millionaire**. You are on the final question. If you answer correctly, you win one million dollars. If your answer is wrong, you leave with nothing.[16] You have just used your final lifeline, 50:50, but you still have no idea which one of the two remaining answers is correct. You also know that you can walk away without answering the question, in which case you leave with \$500,000. Do you take a guess or do you leave?

[16] In the actual game show a contestant who answers the final question incorrectly leaves with \$32,000, not \$0.

When confronted with this choice, most people will take the \$500,000 and leave, even though taking a guess corresponds to a lottery with the same expected value (but not the same expected utility!). This phenomenon comes from **diminishing marginal utility of wealth**. The first \$500,000 can make a huge difference in someone's life, but winning one million dollars is not "twice as good" as winning \$500,000. As such, the risk is not worth it. More formally, diminishing marginal utility of wealth corresponds to concavity of the Bernoulli utility function u.

2.2 *Implications of a concave Bernoulli utility function.*

(i) Turning down fair bets. A **fair bet** is a lottery X with

$$\mathbb{E}[X] = 0$$

Starting from some initial wealth y, if u is concave then:

$$\mathbb{E}[u(y + X)] \leq u\left(\mathbb{E}[y + X]\right) = u(y)$$

where the first inequality comes from **Jensen's inequality**.

This is related to the phenomenon of **fully insuring** against losses when insurance premium is **actuarily fair**. We may consider a lottery X that takes on value $-d + pd$ with probability p and value $0 + pd$ with probability $(1 - p)$. To interpret, the lottery refers to the change in wealth that someone currently buying full insurance would experience if she were to stop buying insurance altogether. In every state of the world, she is pd richer since she no longer pays the (actuarily fair) insurance premium. However, she loses d in the bad state of the world where her house burns down. It is simple to verify that the lottery X is a fair bet. If the individual has a concave u, then she would turn down this (and every other) fair bet. But this means she prefers full insurance to no insurance.

(ii) Certainty equivalence smaller than expected value.

Recall that for a given lottery X and initial wealth level y, the **certainty equivalence** of X at y is written as $c(X, y)$ and defined implicitly by:

$$u(y + c(X, y)) = \mathbb{E}[u(y + X)]$$

That is, $c(X, y)$ is the for-sure payment that would make the decision-maker exactly as happy as facing the lottery X. When u is concave,

$$\boxed{c(X, y) \leq \mathbb{E}[X]} \tag{26}$$

which says a lottery is **"worth" less than its expected value**. (This is once again a consequence of Jensen's inequality.) This property means there exists a trade between a risk-averse individual and a risk-neutral firm that benefits both parties. In particular, the individual can sell her lottery to the firm for $c(X, y) + \epsilon$, where $\epsilon > 0$ is small. This makes the individual strictly better off than facing the lottery, since she received a for-sure payment larger than the certainty equivalence of the lottery. At the same time, the firm on average makes strictly positive profits on this trade, which is good enough if the firm's objective function is risk neutral.

2.3 *Constant absolute risk aversion (CARA).* In lecture we looked at the **CARA** Bernoulli utility function:

$$u(y) = -\exp(-\alpha y)$$

where $\alpha > 0$. An individual with this Bernoulli utility function has a risk attitude **invariant with wealth level** in the following sense. Suppose I offer you a bet based on a flip of a fair coin and you have CARA Bernoulli utility. If you [accept / turn down] this bet at your current wealth, you would continue to [accept/turn down] the same bet when your wealth is \$100,000, \$1 million, \$1 billion, or any other number. To see this, suppose the bet pays you $G > 0$ if the coin lands heads but costs you $L > 0$ if the coin lands tails. The expected utility from accepting the bet is:

$$\frac{1}{2} \cdot (-\exp(-\alpha(y+G))) + \frac{1}{2}(-\exp(-\alpha(y-L)))$$

which simplifies to

$$\frac{1}{2}(-\exp(-\alpha y)) \cdot [\exp(-\alpha G) + \exp(\alpha L)]$$

The individual accepts the bet if and only if:

$$\begin{aligned} \frac{1}{2}(-\exp(-\alpha y)) \cdot [\exp(-\alpha G) + \exp(\alpha L)] &\geq -\exp(-\alpha y) \\ \exp(-aG) + \exp(\alpha L) &\leq 2 \end{aligned}$$

which is a condition not dependent on initial wealth y. Actually, this fact holds for any kind of lottery X, not just lotteries associated with fair coin tosses. We state this fact below:

Fact. *If an expected-utility maximizing individual has CARA Bernoulli utility and X is a lottery, then she [accepts / turns down] X at wealth level y if and only if she also [accepts / turns down] X at every other wealth level y'.*

3 Paradoxes of expected utility

3.1 *Allais paradox and probability weighting function.* Recall Allais Paradox from lecture today:

Lottery A		Lottery B		Lottery C		Lottery D	
\$1,000,000	w.p. 1	\$5,000,000	w.p. 0.10	\$1,000,000	w.p. 0.11	\$5,000,000	w.p. 0.10
		\$1,000,000	w.p. 0.89	\$0	w.p. 0.89	\$0	w.p. 0.90
		\$0	w.p. 0.01				

The puzzling empirical fact is that most people prefer A over B but prefer D over C. This cannot happen if people are expected-utility maximizers.

Intuitively speaking, it is the tension between the **linearity** with which probability enters into an **expected utility** calculation and the **non-linearity** with which probability enters into our **psychological evaluation** of risk that generates Allais paradox. We might have trouble accepting the 1% risk of getting \$0 in lottery B because we have trouble visualizing just how small a chance 1% is. At the same time, we have no qualms about picking lottery D over lottery C, even though D also features an extra 1% chance of getting \$0 compared to C. We say to ourselves: there is already a very high probability (89%) of not getting any payoff in lottery C, so what's another 1% risk on top of that?

The above discussion suggests a modification of expected utility model where we replace objective probabilities with their **distorted** counterparts:

$$\boxed{U(X) = \sum_x \phi(\Pr[X = x]) \cdot u(x)} \tag{27}$$

Here, $\phi : [0, 1] \to [0, 1]$ is a **probability weighting function** so that $\phi(p)$ represents the psychological impact of an objective probability p. We usually assume $\phi(0) = 0$, $\phi(1) = 1$, ϕ is increasing. The decision-maker uses these $\phi-$distorted "probabilities", now called **decision weights**, to evaluate a lottery.

To see how a probability weighting function can help explain Allais paradox, suppose $u(y) = \ln(\epsilon+y)$ where $\epsilon = \exp(-45)$ is a very small positive number. Suppose $\phi(p) = \sqrt{p}$. Under this combination of Bernoulli utility and probability weighting function, the utility of A is:

$$\ln(\epsilon + 1000000) \cdot \sqrt{1} \approx 13.812$$

The utility of B is:

$$\begin{aligned}
& \ln(\epsilon + 5000000) \cdot \sqrt{0.10} + \ln(\epsilon + 1000000) \cdot \sqrt{0.89} + \ln(\epsilon + 0) \cdot \sqrt{0.01} \\
\approx\ & \ln(\epsilon + 5000000) \cdot 0.316 + \ln(\epsilon + 1000000) \cdot 0.943 + \ln(\epsilon) \cdot \mathbf{0.1} \\
\approx\ & 13.411
\end{aligned}$$

Therefore, pick A over B.

At the same time, the utility of C is:

$$\begin{aligned}
& \ln(\epsilon + 1000000) \cdot \sqrt{0.11} + \ln(\epsilon + 0) \cdot \sqrt{0.89} \\
\approx\ & \ln(\epsilon + 1000000) \cdot 0.332 + \ln(\epsilon) \cdot \mathbf{0.943} \\
\approx\ & -37.87
\end{aligned}$$

The utility of D is:

$$\begin{aligned}
& \ln(\epsilon + 5000000) \cdot \sqrt{0.10} + \ln(\epsilon + 0) \cdot \sqrt{0.90} \\
\approx\ & \ln(\epsilon + 5000000) \cdot 0.316 + \ln(\epsilon) \cdot \mathbf{0.949} \\
\approx\ & -37.81
\end{aligned}$$

Therefore, pick D over C.

The trick here is that **decision weights are not linear in objective probability**. In particular, a 1% change in objective probability translates into a huge increase in decision weights if the change occurs at a very low baseline level of objective probability. When we compare lotteries A and B, the 1% chance of getting \$0 in B gets a decision weight of 0.1. However, when we compare lotteries C and D, the 1% extra chance of getting \$0 in D relative to C translates only into a decision weight change of 0.006. The concavity of the weighting function $\phi(p) = \sqrt{p}$ means the marginal contribution of objective probability to decision weights is diminishing, much like how a concave Bernoulli utility function leads to diminishing marginal utility of wealth.

3.2 *Ellsberg paradox and Knightian uncertainty.* We slightly rephrase an example from lecture today. There are **two jars**, each filled with **100 balls**. All balls are either **red** or **green**. Jar A contains 50 red balls and 50 green balls. The composition of jar B is **unknown**. I draw one ball at random from each of the two jars. The relevant states of the world are:

$$S = \{RR, RG, GR, GG\}$$

which represent "red ball from both jars", "red ball from jar A, green ball from jar B", "green ball from jar A, red ball from jar B", "green ball from both jars". Consider four lotteries:

Lottery		W	X	Y	Z
Bet		red ball from A	green ball from A	red ball from B	green ball from B
Payoff	RR	\$20	\$0	\$20	\$0
	RG	\$20	\$0	\$0	\$20
	GR	\$0	\$20	\$20	\$0
	GG	\$0	\$20	\$0	\$20

The puzzling empirical fact: a decision-maker is typically indifferent between W and X, indifferent between Y and Z, but strictly prefers either one of W or X over either one of Y or Z.

The Ellsberg paradox represents a far more fundamental challenge to expected utility theory than Allais paradox. The setup of the Ellsberg experiment violates a basic tenet of our model of risk – that states have **well-defined, objective probabilities**. In a usual model of risk, we could point to any member of the state space and say what its probability is. If the state space contains the outcomes of a fair die toss, then each outcome has probability $\frac{1}{6}$. This $\frac{1}{6}$ probability is "**objective**", in the sense that it is inherent in the setup of the problem and every decision-maker agrees on it. But what is the probability of state RR in Ellsberg's experiment? We cannot say for sure since we do not know the composition of jar B. Uncertainty about the probability distribution over state space is called **Knightian uncertainty**.

We might try to resolve this problem by saying the decision-maker simply forms a **subjective belief** over the state space in the absence of **objective probabilities**. Sure, there is no objective distribution over S. But this doesn't prevent the individual from believing the following in her head:

$$\Pr[RR] = \frac{1}{4}, \Pr[RG] = \frac{1}{4}, \Pr[GR] = \frac{1}{4}, \Pr[GG] = \frac{1}{4}$$

This is the belief that would arise if the individual thinks the composition of jar B is 50 red balls, 50 green balls. Of course, this belief is subjective in the sense that another individual might think jar B actually contains 30 red balls and 70 green balls, in which case his belief about the probability distribution over S would be different.

Unfortunately, allowing subjective belief does not resolve the paradox either. In fact, the punchline of the Ellsberg experiment is that **decision-making under Knightian uncertainty is not equivalent to decision-making under a subjective belief**. Suppose the individual forms a belief over S. In that case one of following two statements must be true:

$$\Pr[RR] + \Pr[GR] \geq \frac{1}{2}$$

or

$$\Pr[RG] + \Pr[GG] \geq \frac{1}{2}$$

So one of Y or Z pays \$20 with probability no less than 50% and pays \$0 otherwise. At the same time, one of the following two statements must be true:

$$\Pr[RR] + \Pr[RG] \leq \frac{1}{2}$$

or

$$\Pr[GR] + \Pr[GG] \leq \frac{1}{2}$$

So one of W or X pays \$20 with probability no more than 50% and pays \$0 otherwise. Therefore at least one of Y or Z must be better than one of W or X, if the decision-maker really evaluates the situation using subjective beliefs. This conclusion goes against the data, which means the typical decision-maker is not acting as if she has a subjective belief over S.

3.3 *Ambiguity aversion.* Loosely speaking, our first attempt at resolving the Ellsberg paradox failed because the approach we took still smelled too much like expected utility. Allowing the decision-maker to entertain a subjective belief over the state space essentially adds one "**degree of freedom**" to the classic expected utility model. The probability distribution over states of the world is no longer determined by the economic environment, but rather by some mental process within the decision-maker. Other than this modification, the subjective belief model **works exactly like expected utility**. Its failure to explain the paradox suggests we need a radical departure from expected utility, not just incremental modifications.

In the past 20 years, economic theorists proposed and studied a large number of models of decision-making that undertake this departure. These models are collectively called models of **ambiguity aversion**. Much like how a risk-averse individual dislikes objective risk and prefers to have a deterministic outcome instead, an ambiguity-averse individual dislikes Knightian uncertainty and prefers to have objective risk instead. This captures preference for bets on a ball drawn from jar A over bets on a ball drawn from jar B in Ellsberg's experiment.

The cleanest ambiguity-averse model is called **maxmin expected utility (MEU)**. There is a state space $S = \{s_1, s_2, ..., s_n\}$ but no objective probability distribution over it. The decision-maker entertains **a set of** probability measures M, where every member $p \in M$ is a distribution on S. The utility of a lottery $X : S \to \mathbb{R}$ is computed as:

$$\boxed{U_{MEU}(X) = \min_{p \in M} \left\{ \sum_{k=1}^{n} p_k \cdot u\left(X(s_k)\right) \right\}} \tag{28}$$

To interpret, the decision-maker is **uncertain about the true probability distribution** over S. She is unwilling to name a single distribution over S as the "right" one, but thinks every distribution in some collection M of distributions is plausible. She is also very **pessimistic**, so she will evaluate any lottery you propose under its **worst-case performance**. That is, she will evaluate your lottery using the most unfavorable distribution in M, giving it the lowest expected utility possible.

The MEU model can explain Ellsberg's paradox. With the state space $S = \{RR, RG, GR, GG\} = \{s_1, s_2, s_3, s_4\}$, let M be a collection of 101 probability measures on S, enumerated as

$$M = \{p^{(0)}, p^{(1)}, p^{(2)}, ..., p^{(100)}\}$$

Each of $p^{(j)}$ is a probability distribution on S for $0 \leq j \leq 100$, where

$$p_1^{(j)} = \frac{1}{2} \cdot \frac{j}{100},\ p_2^{(j)} = \frac{1}{2} \cdot \frac{100-j}{100},\ p_3^{(j)} = \frac{1}{2} \cdot \frac{j}{100},\ p_4^{(j)} = \frac{1}{2} \cdot \frac{100-j}{100}$$

That is, $p^{(j)}$ is the correct probability distribution on S if we knew for sure that there were j red balls out of 100 in jar B. A MEU decision-maker under Knightian uncertainty treats every distribution $p^{(j)}$ as plausible. This does not mean that her M is equal to the set of all distributions on S. For instance, every $p \in M$ reflects her knowledge that half of the balls in jar A are red and that the balls drawn from jar A and jar B are statistically independent.

Let's compute $U_{MEU}(Y)$. Since Y involves betting on a red ball from jar B, the worst case scenario is when jar B contains 0 red balls. This situation is captured by the probability distribution $p^{(0)}$. Hence,

$$\begin{aligned}
U_{MEU}(Y) = \min_{p \in M} \left\{ \sum_{k=1}^{4} p_k \cdot u\left(Y(s_k)\right) \right\} &= p_1^{(0)} u(Y(s_1)) + p_2^{(0)} u(Y(s_2)) + p_3^{(0)} u(Y(s_3)) + p_4^{(0)} u(Y(s_4)) \\
&= 0 \cdot u(20) + \frac{1}{2} \cdot u(0) + 0 \cdot u(20) + \frac{1}{2} \cdot u(0) \\
&= u(0)
\end{aligned}$$

The **worst case** performance of Y is getting **\$0 with probability 1**. What about $U_{MEU}(Z)$? Well this lottery's worst case performance is just as bad as that of Y, except its expected utility is minimized by a different distribution from M, namely $p^{(100)}$. It is not hard to see that $U_{MEU}(Z) = u(0)$ also.

What about the lottery W? For a given $0 \leq j \leq 100$, the decision-maker's expected utility from W under the probability distribution $p^{(j)}$ is:

$$\begin{aligned}
p_1^{(j)} u(W(s_1)) + p_2^{(j)} u(W(s_2)) + p_3^{(j)} u(W(s_3)) + p_4^{(j)} u(W(s_4)) &= p_1^{(j)} u(20) + p_2^{(j)} u(20) + p_3^{(j)} u(0) + p_4^{(j)} u(0) \\
&= \frac{1}{2} u(20) + \frac{1}{2} u(0)
\end{aligned}$$

since $p_1^{(j)} + p_2^{(j)} = p_3^{(j)} + p_4^{(j)} = \frac{1}{2}$ for every $0 \leq j \leq 100$. Therefore, it turns out her expected utility from W is the **same** regardless of which distribution $p \in M$ she uses in computing it. This makes sense. Betting on a red ball from jar A involves only objective risk and does not depend on her belief about the composition of jar B. So, her minimum expected utility across all members of M for the lottery W is:

$$U_{MEU}(W) = \min_{p \in M} \left\{ \sum_{k=1}^{4} p_k \cdot u\left(W(s_k)\right) \right\} = \min_{0 \leq j \leq 100} \left\{ \sum_{k=1}^{4} p_k^{(j)} \cdot u\left(W(s_k)\right) \right\} = \frac{1}{2} u(20) + \frac{1}{2} u(0)$$

because the 101 expected utilities we are taking the minimum over are all equal! By a symmetric argument we can also show $U_{MEU}(X) = \frac{1}{2} u(20) + \frac{1}{2} u(0)$.

Therefore, MEU generates indifference between W and X, indifference between Y and Z, and strict preference for either W or X over either of Y or Z, as in data.

Models of ambiguity aversion have been used by behavioral economists to explain many choice phenomena. For instance, **home bias** is a puzzle in financial economics. A typical US investor's portfolio contains about 90% domestic stocks and 10% foreign stocks. But based on the weight of US market in the world equity portfolio, the optimal portfolio consists of at least 50% foreign stocks. Why such strong aversion to the foreign equity? One explanation is that the US investor understands the risks surrounding US stocks much better than the risks surrounding foreign stocks. In fact, she might view the returns on a US stock as a lottery on a state space with objective probabilities, but the returns of a foreign stock as a lottery on a state space without such objective probabilities. Said another way, investing in US stock is like betting on a draw from Ellsberg's jar A, while investing in foreign stock is like betting on a draw from Ellsberg's jar B. Ambiguity-averse decision models, such as MEU, can help us explain such behavior.

Economics 1011a ▷ Section 7 : Welfare, Market, and GE[17] —— 10/28/2014

(1) EV, CV, CS; **(2)** Economy without money; **(3)** Walrasian equilibrium

TF: Kevin He (he02@fas.harvard.edu)

— 1 Equivalent variation, compensating variation, consumer surplus —

1.1 *Integral representations of welfare measures.* Consider a consumer with wealth Y who faces n consumption goods. Initially, the prices of the n goods are $\mathbf{p}^{(0)} = (p_1^{(0)}, p_2^{(0)}, ..., p_n^{(0)})$ and the consumer achieves utility level $u^{(0)}$. However, price of the first good then **decreases** from $p_1^{(0)}$ to $p_1^{(1)}$, leading to a new price vector $\mathbf{p}^{(1)} = (p_1^{(1)}, p_2^{(0)}, ..., p_n^{(0)})$. Under $\mathbf{p}^{(1)}$, the consumer achieves utility level $u^{(1)}$ where $u^{(1)} > u^{(0)}$. How can we phrase the **welfare impact** of this price change in monetary terms? **Equivalent variation** (**EV**) measures the dollar equivalent of utility gain under old prices $\mathbf{p}^{(0)}$. That is, EV is equal to the dollar amount we must give to the consumer so that she achieves her **new utility** under the **old prices**.

$$\boxed{EV = e\left(\mathbf{p}^{(0)}, u^{(1)}\right) - Y} \tag{29}$$

Compensating variation (**CV**) measures the dollar equivalent of utility gain under new prices $\mathbf{p}^{(1)}$. That is, CV is equal to the dollar amount we must take away from the consumer so that she achieves her **old utility** under the **new prices**.

$$\boxed{CV = Y - e\left(\mathbf{p}^{(1)}, u^{(0)}\right)} \tag{30}$$

But remember we have two identities: $Y = e\left(\mathbf{p}^{(0)}, u^{(0)}\right)$ and $Y = e\left(\mathbf{p}^{(1)}, u^{(1)}\right)$. We use them to write:

$$EV = e\left(\mathbf{p}^{(0)}, u^{(1)}\right) - e\left(\mathbf{p}^{(1)}, u^{(1)}\right) = \int_{p_1^{(1)}}^{p_1^{(0)}} \frac{\partial e}{\partial p_1}(\tilde{p}_1, p_2^{(0)}, ..., p_n^{(0)}, u^{(1)})d\tilde{p}_1$$

where last equality comes from the **Fundamental Theorem of Calculus**. But $\frac{\partial e}{\partial p_1} = h_1$ by **Shephard's lemma**, so we deduce:

$$\boxed{EV = \int_{p_1^{(1)}}^{p_1^{(0)}} h_1(\tilde{p}_1, p_2^{(0)}, ..., p_n^{(0)}, u^{(1)})d\tilde{p}_1} \tag{31}$$

Analogously,

$$CV = e\left(\mathbf{p}^{(0)}, u^{(0)}\right) - e\left(\mathbf{p}^{(1)}, u^{(0)}\right) = \int_{p_1^{(1)}}^{p_1^{(0)}} \frac{\partial e}{\partial p_1}(\tilde{p}_1, p_2^{(0)}, ..., p_n^{(0)}, u^{(0)})d\tilde{p}_1$$

$$\boxed{CV = \int_{p_1^{(1)}}^{p_1^{(0)}} h_1(\tilde{p}_1, p_2^{(0)}, ..., p_n^{(0)}, u^{(0)})d\tilde{p}_1} \tag{32}$$

[17] Part 1 of this week's section materials borrows from the work of previous TFs, in particular the notes of Zhenyu Lai. Additional image credit: *Microeconomic Theory* by Mas-Colell, Whinston, and Green.

So EV and CV can be represented as integrals of the Hicksian demand over the interval of price change. The only difference is that the integrands are Hicksian demands at two different utility levels: the new utility $u^{(1)}$ level for EV and the old utility level $u^{(0)}$ for CV.

There is yet a third measure of consumer welfare change called the **consumer surplus** (**CS**). Consumer surplus is defined with an integral representation,

$$\boxed{CS = \int_{p_1^{(1)}}^{p_1^{(0)}} x_1(\tilde{p}_1, p_2^{(0)}, ..., p_n^{(0)}, Y) d\tilde{p}_1} \tag{33}$$

1.2 *Graphical interpretations of EV, CV, and CS.* Figure 8 displays EV, CV, and CS assuming good 1 is a normal good. By duality, $x_1\left(\mathbf{p}^{(0)}, Y\right) = h_1\left(\mathbf{p}^{(0)}, u^{(0)}\right)$ and $x_1\left(\mathbf{p}^{(1)}, Y\right) = h_1\left(\mathbf{p}^{(1)}, u^{(1)}\right)$. This helps us trace out the Marshallian demand, which must intersect the Hicksian demand at these two points.

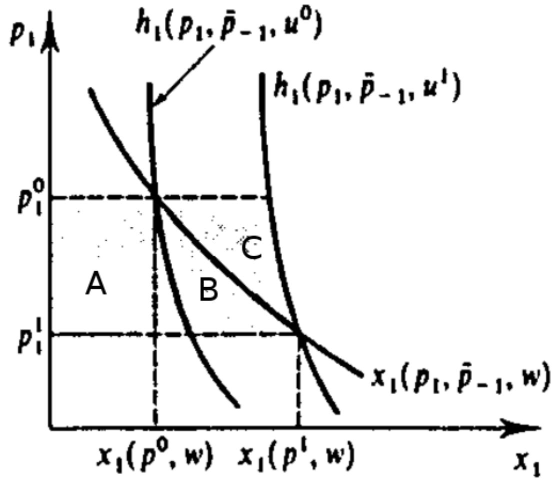


Figure 8: $CV = A$, $CS = A \cup B$, $EV = A \cup B \cup C$. Note $|CV| \leq |CS| \leq |EV|$ for a normal good whose price decreases.

2 Economy without money

Money seems ubiquitous in our study of economics thus far. The central optimization problems of the firm and the consumer prominently feature money in their objective and/or constraint (cost of capital, wealth of consumer, price of kale, etc). We have also seen that even behavior not usually analyzed in monetary terms, such as fertility choice or decision to pursue a romantic relationship, could be modeled as optimization problems involving money.

But money is merely an **economic institution invented by human society**. If we understand economics as the study of allocations of scarce resources, then economists should be interested in money only as an instrument towards "good" resource allocations, where "good" is yet undefined.

This part of the notes studies an exchange economy where the concepts of "money" and "price" were **never invented**. Without any reference to these two concepts, we define a set of **allocational outcomes** that seem **desirable**. In the next part of the notes, we shall see an equilibrium concept involving money achieves outcomes in this desirable set. This exercise provides a justification of money (and market) as a useful economic institution serving allocational ends.

2.1 *Pareto optimality in an exchange economy.* So far, our Economics 1011a models have dealt with optimization problems of a **single decision-maker**. But an **economy** almost always consists of **multiple** individuals[18]. We introduce below a model of an economy with many individuals but only a finite amount of resources.

Definition. An **exchange economy** has:

- L commodities ($\ell = 1, 2, ..., L$), so that space of consumption is $\mathbb{R}^L_+$
- I agents ($i = 1, 2, ..., I$), each with a utility function $u_i : \mathbb{R}^L_+ \to \mathbb{R}$ and an initial endowment $\omega_i \in \mathbb{R}^L_+$

Definition. An **allocation** in an exchange economy is a list of vectors $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_I)$ with each $\mathbf{x}_i \in \mathbb{R}^L_+$. Call an allocation **feasible** if $\sum_{i=1}^I \mathbf{x}_i = \sum_{i=1}^I \omega_i$.

To interpret, think of $\mathbf{x}_i$ in an allocation as the bundle of goods that agent i gets to consume. Feasible allocations are ways of splitting up the total social endowment amongst the I agents with nothing left over.

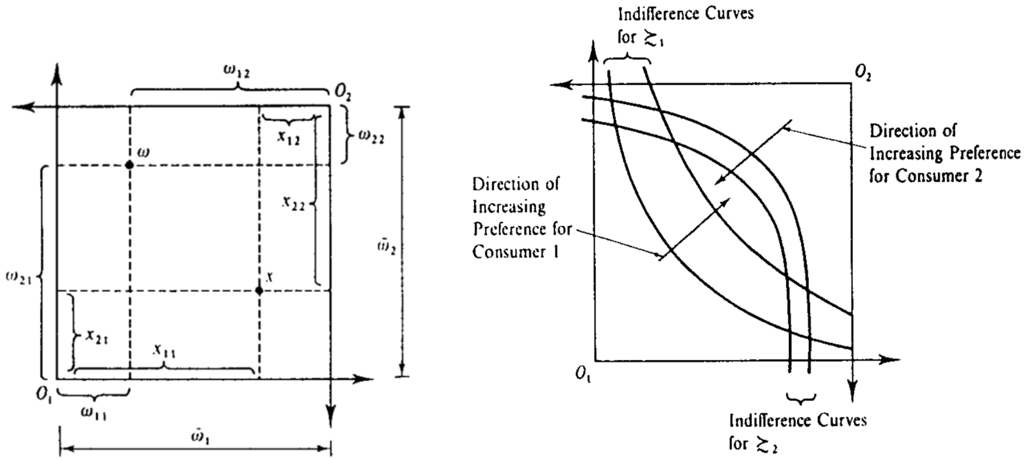
We are ready to define the first desideratum of allocation. The most important problem of trying define an "optimal" resource allocation is that there is no natural way to evaluate a change in allocation that makes some agents better off and other agents worse off. If I take away an apple from Alice and give it to Bob, is this a "better" allocation than the old one? These kinds of questions are tough to answer without specifying a way of **aggregating** all the welfare changes (with possibly **different signs**) across all agents in the economy, which we are not prepared to do. It is uncontroversial, however, that if a change in allocation improves the welfare of **every** agent in the economy, then the new allocation is certainly better than the old one. This is the idea behind Pareto efficiency.

Definition. Allocation **y Pareto dominates** allocation $\mathbf{x}$ if $u_i(\mathbf{y}_i) \geq u_i(\mathbf{x}_i)$ for every $1 \leq i \leq I$ and $u_{i^*}(\mathbf{y}_{i^*}) > u_{i^*}(\mathbf{x}_{i^*})$ for some $1 \leq i^* \leq I$. Call a feasible allocation **Pareto efficient** if it is not Pareto dominated by any other feasible allocation.

As an example of an exchange economy, suppose $L = 3$, $I = 3$, and $u_i(x_{1i}, x_{2i}, x_{3i}) = \min\{x_{1i}, x_{2i}, x_{3i}\}$. Suppose also $\omega_1 = (0, 1, 1)$, $\omega_2 = (1, 0, 1)$, $\omega_3 = (1, 1, 0)$. The **endowment** is a **feasible** allocation, but it is very **inefficient**: everyone is getting 0 utility because every agent has Leontief preferences and is missing one of the three consumption goods. Noticing that the total social endowment is $(2, 2, 2)$, we might consider the "**egalitarian allocation**" where every agent gets $(\frac{2}{3}, \frac{2}{3}, \frac{2}{3})$. This Pareto dominates the endowment allocation since every agent's utility improves from 0 to $\frac{2}{3}$. In fact, the egalitarian allocation is Pareto efficient in this example. There is no way to make any agent better off without making someone else worse off. This is not the only Pareto efficient allocation though. The allocation where agent 1 gets $(2, 2, 2)$ while agents 2 and 3 each gets $(0, 0, 0)$ is also Pareto efficient! Indeed, it is impossible to Pareto dominate this **very skewed** allocation since any attempt to assign more resources to agents 2 or 3 will entail removing resources from agent 1. One could interpret this as a criticism of Pareto efficiency. Pareto efficiency applies to even "outrageous" allocations and does not take distributional fairness into account.

To study another example of exchange economy, consider the case of $L = 2, I = 2$. There is a **graphical method** for analyzing exchange economies with 2 agents and 2 commodities called the

[18] Unless you are Robinson Crusoe. In fact, toy examples of economies that consist of only one consumer are usually called **Crusoe economies**. Fun anecdote: some graduate student apparently thought "Crusoe economy" is named after an economist, like "Leontief production function" or "Hicksian demand". This led him to ask people which university Professor Crusoe works at.



(a) An Edgeworth box. (b) Indifference curves in an Edgeworth box.

Figure 9: Edgeworth box for two agents, two commodities exchange economies.

Edgeworth box. As in Figure 9a, an Edgeworth box juxtaposes the consumption spaces of both agents on a single graph. The width of the box is $\bar{\omega}_1 := \omega_{11} + \omega_{12}$, the total social endowment of commodity 1 in the economy. The height of the box is $\bar{\omega}_2 := \omega_{21} + \omega_{22}$. Crucially, agent 1's consumption space has the "usual" axes and uses $(0, 0)$ as the origin, while the consumption space for agent 2 has **"inverted" axes** and puts the origin at $(\bar{\omega}_1, \bar{\omega}_2)$. Any point $r = (r_1, r_2)$ within the box refers to a feasible allocation. This allocation assigns r_1 of good 1, r_2 of good 2 to agent 1 and gives the remaining social endowment to agent 2.

Figure 9b draws the indifference curves of the two agents in an Edgeworth box. As we might expect, if preferences are monotonically increasing in consumption, then each agent assigns **higher utility** levels to points **further away** from her origin.

The allocation $\mathbf{x}$ in the first Edgeworth box of Figure 10 is not Pareto efficient. To see this, we draw the indifference curves of the two agents associated with the $\mathbf{x}$ allocation. Any point inside the shaded region Pareto dominates $\mathbf{x}$, since it lies on **higher indifference curves for both agents**. However, the allocation in the second Edgeworth box of Figure 10 is Pareto efficient. After drawing the indifference curves through $\mathbf{x}$, we discover that the indifference curves of the two agents are in fact **tangent at this allocation**. So, there is **no intersection** between the set of allocations that make agent 1 strictly better off and the set of allocations that leave agent 2 no worse off. The allocation $\mathbf{x}$ cannot be Pareto dominated.

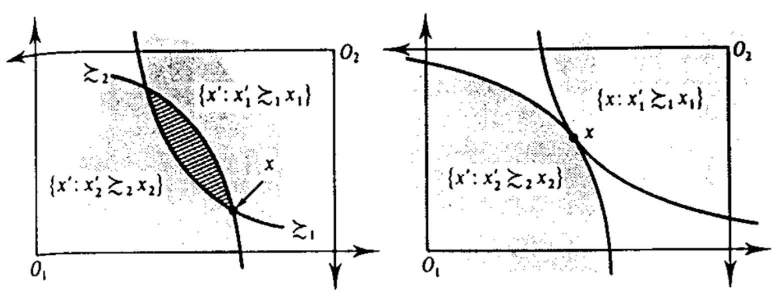


Figure 10: Allocation $\mathbf{x}$ in the Edgeworth box on the left is not Pareto efficient. Allocation $\mathbf{x}$ in the Edgeworth box on the right is Pareto efficient.

2.2 Blocking coalitions and the core. We can think of a feasible allocation in an exchange economy as a **proposal** put forward by a social planner. However, if economic agents are **rational actors**, they need not accept every proposal. Recall that agent i has endowment ω_i, which can be interpreted as the bundle she owned before resource re-allocation was possible. She will likely **refuse to participate** in the (Pareto efficient) allocation where all individual endowments are confiscated and given to, say, agent 1. She would rather run away from the rest of the economy with her endowment than take part in this proposed trade.

The definition of a blocking coalition generalizes this idea. A feasible allocation might be unacceptable to not one agent, but a **group of agents**, who would collectively do better if they **declared independence** from the rest of the economy and traded only amongst themselves. If such a group exists, then the proposed allocation is probably unstable and will not arise under voluntary market participation.

Definition. A **coalition** is a nonempty subset $S \subseteq \{1, 2, ..., I\}$. A coalition S **blocks** allocation $\mathbf{x} = (\mathbf{x}_1, ..., \mathbf{x}_I)$ if there exists bundles $\{\mathbf{y}_i\}_{i \in S}$ such that (i) $\sum_{i \in S} \mathbf{y}_i = \sum_{i \in S} \omega_i$; (ii) $u_i(\mathbf{y}_i) \geq u_i(\mathbf{x}_i)$ for every $i \in S$; (iii) $u_{i^*}(\mathbf{y}_i) > u_{i^*}(\mathbf{x}_i)$ for some $i \in S$.

In words, S is a blocking coalition for the allocation $\mathbf{x}$ if we can find a feasible re-allocation of endowments amongst the agents in S that Pareto dominates $\mathbf{x}$ within S.

Definition. The **core** of an exchange economy is the set of feasible allocations not blocked by any coalition.

Note that an allocation not blocked by the grand coalition of everyone in the economy must be Pareto efficient. Therefore the core[19] of the economy is a subset of the Pareto efficient outcomes.

As an example, let's consider again an exchange economy with $L = 3$, $I = 3$, and $u_i(x_{1i}, x_{2i}, x_{3i}) = \min\{x_{1i}, x_{2i}, x_{3i}\}$. The endowments are $\omega_1 = (0, 1, 1)$, $\omega_2 = (1, 0, 1)$, $\omega_3 = (1, 1, 0)$. The allocation $\mathbf{x}$ that assigns $(2, 2, 2)$ to agent 1 and $(0, 0, 0)$ to agents 2 and 3 is **Pareto efficient**. However, it is **blocked** by the coalition $S = \{2, 3\}$. Indeed, consider the allocation $\mathbf{y}_2 = (1, \frac{1}{2}, \frac{1}{2})$, $\mathbf{y}_3 = (1, \frac{1}{2}, \frac{1}{2})$ amongst the two coalition members. This is a possible way of splitting up the total endowments of only agents 2 and 3. Further, it makes the two agents strictly better off than accepting the allocation that assigns $(0, 0, 0)$ to each of them. Therefore, $\mathbf{x}$ is not in the core.

3 Walrasian equilibrium

3.1 Money as an instrument towards efficiency and stability. We have identified two allocational desiderata in an exchange economy without any reference to money. The first desideratum is **efficiency**, as formalized by Pareto efficiency. An allocation that can be improved in a way that makes everyone better off is surely undesirable. The second desideratum is **stability**, as formalized by the core. Any proposed allocation would lead to some blocking coalition declaring their economic independence is surely a bad outcome.

We shall see that the introduction of money, or more precisely the introduction of a **market economy that attaches prices to all the commodities**, leads to equilibrium outcomes that satisfy these two properties. When commodity prices are introduced into our exchange economy, allocational decision becomes **de-centralized**. Every agent in the economy receives some wealth by selling her entire

[19] You might remember another concept called the "core" in your previous studies of cooperative games (or from the 2012 Nobel Memorial Prize in Economics). What we are defining here is in fact a special case of that concept. The problem of trading resources in an exchange economy can be phrased as a cooperative game. The "core" of that game is exactly the core as defined in this set of notes.

endowment to the market. She then solves the usual **utility maximization problem**, where she demands a consumption bundle that maximizes her welfare subject to her budget constraint. Notice the total supply of commodities is **fixed** in the economy. In fact, absent firms and production, the total supply is simply the total social endowment. The price-based market economy is in equilibrium if the sum of consumer demands is a feasible allocation.

Definition. A **Walrasian equilibrium** is an allocation $\mathbf{x}^* = (\mathbf{x}_1^*, ..., \mathbf{x}_I^*)$ and a price vector[a] $\mathbf{p}^* \in \mathbb{R}^L_{++}$ such that:
(i) [individual maximization] $\mathbf{x}_i^* \in \arg\max_{\mathbf{x}_i \in \mathbb{R}^L_+} u_i(\mathbf{x}_i)$ s.t. $\mathbf{p}^* \cdot \mathbf{x}_i \leq \mathbf{p}^* \cdot \omega_i$ for every $1 \leq i \leq I$.
(ii) [market clearing] $\sum_{i=1}^I \mathbf{x}_i^* = \sum_{i=1}^I \omega_i$

[a]The notation $\mathbf{p}^* \in \mathbb{R}^L_{++}$ means $\mathbf{p}^*$ is an L-dimensional real vector and $p_\ell > 0$ for every $1 \leq \ell \leq L$.

As an example, consider the Edgeworth box visualization of a two agents, two commodities economy. No matter what price $\mathbf{p}^* \in \mathbb{R}^2_{++}$ is announced, every agent can always afford her endowment. So, if we draw the **budget line** through the total social endowment, we **partition** the Edgeworth box into two regions: the set of consumptions bundles affordable to agent 1 and the set of consumption bundles affordable to agent 2, as in Figure 11a.

Utility maximization is associated with the usual picture, shown in Figure 11b. Each agent finds an indifference curve tangent to her budget line. The **point of tangency** is her Marshallian demand. In general, for an arbitrary price vector the Marshallian demands of the two agents will both be points on the budget line, but they **need not be the same point**. This mismatch means markets did not clear at the given prices.

A Walrasian equilibrium is visualized in the Edgeworth box as a price vector such that the Marshallian demands of the two agents refer to the **same allocation**. See Figure 11c.

A Walrasian equilibrium is efficient and stable. More precisely, provided **mild regularity conditions** hold regarding the consumers' utility functions and we can find a price vector that clears the market, the de-centralized allocation that arise as the Marshallian demands in a Walrasian equilibrium is Pareto efficient and in the core.

Here is one version of the "mild regularity condition": call a utility function u_i **strongly monotonic** if whenever $\mathbf{x}_i, \hat{\mathbf{x}}_i$ are such that $\hat{x}_{\ell i} > x_{\ell i}$ for every $1 \leq \ell \leq L$, we have $u_i(\hat{\mathbf{x}}_i) > u_i(\mathbf{x}_i)$. This is a very weak condition. In particular, we do not need to assume u_i is concave or differentiable (or even continuous).

Theorem. *(First welfare theorem) If u_i is strongly monotone for every $1 \leq i \leq n$, then any Walrasian equilibrium allocation is Pareto efficient.*

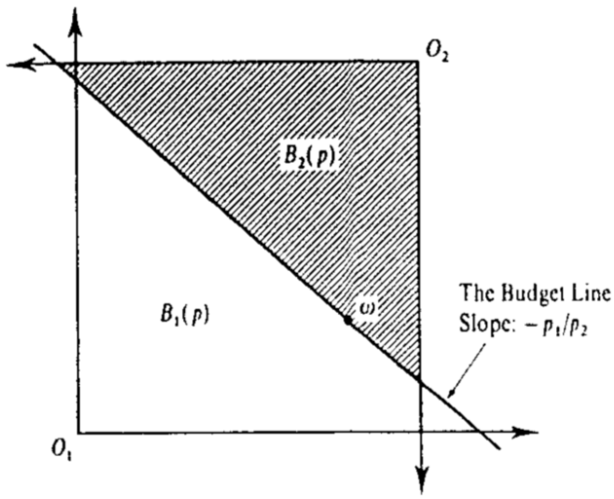
Proof. See Thursday's lecture. □

Theorem. *If u_i is strongly monotone for every $1 \leq i \leq n$, then any Walrasian equilibrium allocation is in the core.*

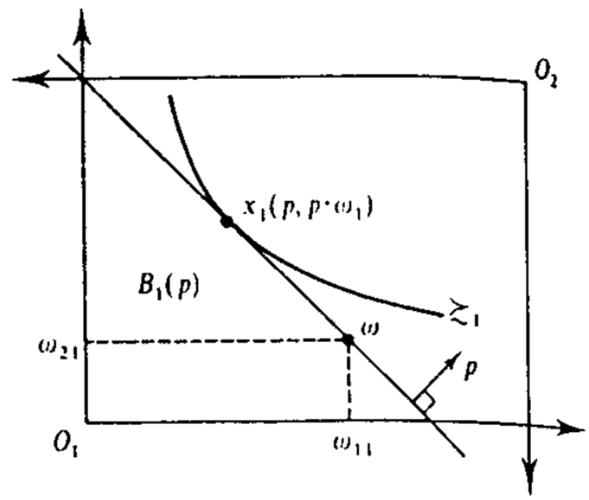
Proof. Left as an exercise. □

Exercise A: Prove this result. As in lecture, your argument should not involve taking derivatives, since u_i may not even be continuous.

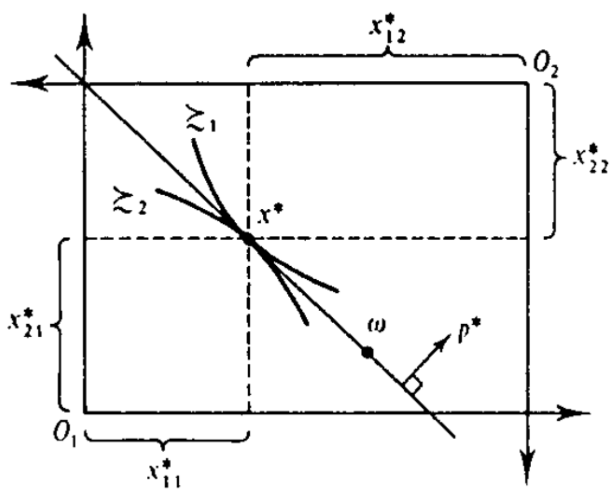
3.2 *Where do market clearing prices come from?* We have established that a Walrasian equilibrium achieves desirable allocational properties. But the next natural question is how an economy settles



(a) Under prices $\mathbf{p}$, a budget line through the endowment point ω partitions the Edgeworth box into two regions. $B_1(\mathbf{p})$ is affordable to agent 1 while $B_2(\mathbf{p})$ is affordable to agent 2.



(b) Utility maximization problem (**UMP**) in an Edgeworth box.



(c) Walrasian equilibrium $\mathbf{x}^*$ in an Edgeworth box

Figure 11: Visualizing Walrasian equilibrium in an Edgeworth box.

on a set of market-clearing prices. Walras gave one explanation in **1874**, which was later termed the "**Walrasian auctioneer**". Walras was inspired by real auctioneers working at the Paris stock exchange in the 19th century, who mediated the market by calling out different prices for a stock until they found a price that equated supply with demand. In the context of an exchange economy, imagine a fictitious auctioneer whose sole objective is to clear the market. The auctioneer observes which goods are in **excess demand** and which goods are in **excess supply** in the economy. He then adjusts prices, raising the prices of goods in excess demand and lowering the prices of goods in excess supply. This process is repeated, with the magnitude of price adjustment diminishing over time as the economy converges towards market clearing. Through this process of trial and error ("**tâtonnement**"), the auctioneer eventually finds a market clearing price.

One modern interpretation of the Walrasian auctioneer is a **root-finding algorithm** in computer science. To understand this strange analogy, consider an economy with $L = 2$ commodities. Write $Z(p)$ for the excess demand for commodity 2 when the price vector is $\mathbf{p} = (1, p)$. That is,

$$\boxed{Z(p) := \sum_{i=1}^{I} x_{2,i}(\mathbf{p}, \mathbf{p} \cdot \omega_i) - \sum_{i=1}^{I} \omega_{2,i}} \tag{34}$$

where $x_{2,i}(\mathbf{p}, Y)$ represents consumer i's Marshallian demand for commodity 2 under prices $\mathbf{p}$ and income Y. For small value of p, we might have $Z(p) > 0$, i.e. positive excess demand of commodity 2, as the cheap price induces significant substitution towards commodity 2. For large value of p, we might have $Z(p) < 0$, i.e. negative excess demand (also called excess supply) of commodity 2, as consumers respond to the high relative price of commodity 2 by selling off their endowment of commodity 2 and substituting to commodity 1. The job of the Walrasian auctioneer is to find some number p^* as to clear the market, that is to say $Z(p^*) = 0$. But this is just a root-finding problem from numerical analysis. In fact, the sequence of trial roots produced by a root-finding algorithm might be interpreted as a tâtonnement process converging on the true, market clearing price.

```
N ← 1
While N ≤ NMAX
   c ← (a + b)/2
   If f(c) = 0 or (b − a)/2 < TOL then
     Output(c)
     Stop
   EndIf
   N ← N + 1
   If sign(f(c)) = sign(f(a)) then a ← c else b ← c
EndWhile
Output("Method failed.") # max number of steps exceeded
```

Figure 12: Two ways to think about the Walrasian auctioneer.

Economics 1011a ▷ Section 8 : Externalities and Public Goods - 11/04/2014

(1) Loose ends in GE; **(2)** Computing Walrasian eqm; **(3)** Bystander effect and public goods

TF: Kevin He (he02@fas.harvard.edu)

1 Loose ends in general equilibrium theory

The first task is to tie up some loose ends in our study of Walrasian equilibrium in an exchange economy. Recall our setup:

Definition. An **exchange economy** has:

- L commodities ($\ell = 1, 2, ..., L$), so that space of consumption is $\mathbb{R}^L_+$
- I agents ($i = 1, 2, ..., I$), each with a utility function $u_i : \mathbb{R}^L_+ \to \mathbb{R}$ and an initial endowment $\omega_i \in \mathbb{R}^L_+$

Definition. A **Walrasian equilibrium** is an allocation $\mathbf{x}^* = (\mathbf{x}^*_1, ..., \mathbf{x}^*_I)$ and a price vector $\mathbf{p}^* \in \mathbb{R}^L_{++}$ such that:

(i) [individual maximization] $\mathbf{x}^*_i \in \underset{\mathbf{x}_i \in \mathbb{R}^L_+}{\arg\max}\ u_i(\mathbf{x}_i)$ s.t. $\mathbf{p}^* \cdot \mathbf{x}_i \leq \mathbf{p}^* \cdot \omega_i$ for every $1 \leq i \leq I$.

(ii) [market clearing] $\sum_{i=1}^I \mathbf{x}^*_i = \sum_{i=1}^I \omega_i$

Last week, we established that Walrasian equilibria satisfy some very nice properties. In lecture, we saw that every Walrasian equilibrium is **Pareto efficient**. As an exercise, you also proved a stronger result: every Walrasian equilibrium is in the **core**. But statements of the form "every element of set E satisfies property P" are not very interesting unless we know the set E is **nonempty**! If E is empty, then such statements are **vacuously true**. In particular, if an exchange economy has no Walrasian equilibria, then any statement of the form "every Walrasian equilibrium in this exchange economy is ______" is true but unhelpful.

The following result states a sufficient condition for the existence of at least one Walrasian equilibrium.

> **Theorem.** *(Existence of Walrasian equilibrium) Suppose for every agent i, (i) the utility function u_i is continuous, strongly monotone, and strictly concave; (ii) the endowment ω_i is in $\mathbb{R}^L_{++}$. Then there exists at least one Walrasian equilibrium in the exchange economy.*

Compared to the conditions we stated for first welfare theorem, the conditions for the existence of Walrasian equilibrium are more stringent. In particular, we need to assume utility functions are **continuous** and **strictly concave**, which were not required for first welfare theorem. We also need the technical assumption that every individual is endowed with a nonzero amount of every commodity in the economy. This is because the "corner case" where agent i does not own any commodity ℓ turns out to cause all kinds of pathological behaviors. An outline of the existence proof is included at the end of this week's notes.

In Thursday's lecture, we gave an "almost proof" of the second welfare theorem. For a given Pareto efficient allocation $\mathbf{x}^* = (\mathbf{x}^*_1, ..., \mathbf{x}^*_I)$, the argument goes, consider an exchange economy where the endowment of agent i is exactly $\mathbf{x}^*_i$. Provided this economy has a Walrasian equilibrium, we argued that the equilibrium allocation has to be $\mathbf{x}^*$. But now that we have a set of sufficient conditions for the existence of a Walrasian equilibrium, we obtain as a corollary a precise statement of the second welfare theorem:

Theorem. *(second welfare theorem) Suppose for every agent i, the utility function u_i is continuous, strongly monotone, and strictly concave. Suppose also $\mathbf{x}^* = (\mathbf{x}_1^*, ..., \mathbf{x}_I^*)$ is a Pareto efficient allocation such that $\mathbf{x}_i^* \in \mathbb{R}_{++}^L$ for every i. Then there exists a set of new endowments $\{\omega_i'\}_{1 \le i \le I}$ with the property $\sum_{i=1}^I \omega_i' = \sum_{i=1}^I \omega_i$ such that if agent i is re-assigned the endowment ω_i' for every $1 \le i \le I$, then there exists a Walrasian equilibrium with allocation $\mathbf{x}^*$ under this new set of endowments.*

One criticism of the second welfare theorem as stated above is that if the government can enforce a redistribution of endowments, then there is no need for a market to begin with. The government should just eliminate the free market and enforce its desired Pareto efficient allocation $\mathbf{x}^*$ through redistribution of consumption bundles. There is a more interesting version of the second welfare theorem that prohibits the government from redistributing endowments but allows it to make budget-balanced money transfers between individuals. That is to say, to reach a desired Pareto efficient outcome, it suffices to tax some individuals and subsidize others without changing anyone's endowment bundle ω_i.

2 Computing Walrasian equilibrium

2.1 *Walras' law.* There are two related results in general equilibrium that both go by the name of "Walras' law". The individual Walras' law says a consumer with a strongly monotone utility function will spend all of her wealth in solving **UMP**.

Proposition. *(individual Walras' law) If utility function u_i is strongly monotone, then for any price vector $\mathbf{p}$ and income Y, Marshallian demand satisfies: $\mathbf{p} \cdot \mathbf{x}_i(\mathbf{p}, Y) = Y$.*

The market Walras' law is an implication of the individual Walras' law. Roughly speaking, it says if all but one commodity markets clear, then the last commodity market must also clear.

Proposition. *(market Walras' law) Consider an exchange economy with L commodities and I individuals where each individual has a strongly monotone utility function. If price vector $\mathbf{p} \in \mathbb{R}_{++}^L$ is such that $\sum_{i=1}^I x_{\ell i}(\mathbf{p}, \omega_i \cdot \mathbf{p}) = \sum_{i=1}^I \omega_{\ell i}$ for $L-1$ of the commodities $\ell \in \{1, 2, ..., L\}$, then in fact $\sum_{i=1}^I x_{\ell i}(p, \omega_i \cdot p) = \sum_{i=1}^I \omega_{\ell i}$ for every $\ell \in \{1, 2, ..., L\}$.*

Proof. Suppose $\sum_{i=1}^I x_{\ell i}(\mathbf{p}, \omega_i \cdot \mathbf{p}) = \sum_{i=1}^I \omega_{\ell i}$ for $1 \le \ell \le L-1$ (other cases are similar). By individual Walras' law, for every consumer i we have:

$$\mathbf{p} \cdot \mathbf{x}_i(\mathbf{p}, \omega_i \cdot \mathbf{p}) = \mathbf{p} \cdot \omega_i$$

Expanding the dot product,

$$\sum_{\ell=1}^L p_\ell \cdot x_{\ell i}(\mathbf{p}, \omega_i \cdot \mathbf{p}) = \sum_{\ell=1}^L p_\ell \cdot \omega_{\ell i}$$

But this equality holds for every individual i. Summing both sides across i obtains:

$$\sum_{i=1}^I \left(\sum_{\ell=1}^L p_\ell \cdot x_{\ell i}(\mathbf{p}, \omega_i \cdot \mathbf{p}) \right) = \sum_{i=1}^I \left(\sum_{\ell=1}^L p_\ell \cdot \omega_{\ell i} \right)$$

Exchanging order of the double summation,

$$\sum_{\ell=1}^{L}\left(\sum_{i=1}^{I} p_\ell \cdot x_{\ell i}(\mathbf{p}, \omega_i \cdot \mathbf{p})\right) = \sum_{\ell=1}^{L}\left(\sum_{i=1}^{I} p_\ell \cdot \omega_{\ell i}\right)$$

$$\sum_{\ell=1}^{L} p_\ell \cdot \left(\sum_{i=1}^{I} x_{\ell i}(\mathbf{p}, \omega_i \cdot \mathbf{p})\right) = \sum_{\ell=1}^{L} p_\ell \cdot \left(\sum_{i=1}^{I} \omega_{\ell i}\right)$$

But $\sum_{i=1}^{I} x_{\ell i}(\mathbf{p}, \omega_i \cdot \mathbf{p}) = \sum_{i=1}^{I} \omega_{\ell i}$ for $1 \leq \ell \leq L-1$, so canceling out the first $L-1$ summands on each side leaves us with:

$$p_L \cdot \left(\sum_{i=1}^{I} x_{Li}(\mathbf{p}, \omega_i \cdot \mathbf{p})\right) = p_L \cdot \left(\sum_{i=1}^{I} \omega_{Li}\right)$$

Since $p_L > 0$, we divide both sides by p_L to obtain:

$$\sum_{i=1}^{I} x_{Li}(\mathbf{p}, \omega_i \cdot \mathbf{p}) = \sum_{i=1}^{I} \omega_{Li}$$

that is to say the market for commodity L also clears. $\square$

The market Walras' law is useful when explicitly solving for the Walrasian equilibrium price in an exchange economy. Provided every consumer has a strongly monotone utility function, it suffices to check that supply equals demand for the first $L-1$ commodities in order to conclude that all markets clear.

2.2 How to find the Walrasian equilibrium. Consider an economy with $I = 2, L = 2$. The utility of agent i is $u_i(x_{1i}, x_{2i}) = \ln(x_{1i}) + \ln(x_{2i})$. Endowments are $\omega_1 = (1, 2)$ and $\omega_2 = (2, 1)$.

We wish to find a Walrasian equilibrium price vector. That is to say, a price vector that clears all markets. Remember, the price vector in a general equilibrium model serves **dual purpose**:

- It determines how much **wealth** each agent has, through selling her endowments. Since agent 1 is endowed with more of commodity 2 and agent 2 is endowed with more commodity 1, a higher relative price on commodity 2 will make agent 1 relatively wealthier than agent 2.
- It determines the **prices** that consumers face in their **UMP**.

Here are the steps you should take to solve for a Walrasian equilibrium:

Step 1: Write down each agent's Marshallian demand as a function of $\mathbf{p}$.

It might be useful to think of each agent as having wealth Y and solve the problem as a standard **UMP**. In this economy, the **UMP** of agent i is:

$$\max_{x_{1i}, x_{2i}} \ln(x_{1i}) + \ln(x_{2i}) \text{ s.t. } p_1 x_{1i} + p_2 x_{2i} \leq Y$$

Consider the monotonic transformation $h(v) = \sqrt{\exp(v)}$. Applying this transformation to u_i obtains:

$$h(u_i(x_{1i}, x_{2i})) = \sqrt{\exp(\ln(x_{1i})) \cdot \exp(\ln(x_{2i}))} = \sqrt{x_{1i} \cdot x_{2i}} = x_{1i}^{\frac{1}{2}} x_{2i}^{\frac{1}{2}}$$

Therefore, u_i represents the same preference as a Cobb-Douglass utility function with $\alpha = \beta = \frac{1}{2}$. Marshallian demand is the same for all all utility functions that represent the same preference, so we conclude:

$$x_{11}(\mathbf{p}, Y) = \frac{(1/2)Y}{p_1}, x_{21}(\mathbf{p}, Y) = \frac{(1/2)Y}{p_2}, x_{12}(\mathbf{p}, Y) = \frac{(1/2)Y}{p_1}, x_{22}(\mathbf{p}, Y) = \frac{(1/2)Y}{p_2}$$

But the wealth of each consumer is equal to the worth of her endowment under price vector $\mathbf{p}$. So in fact,

$$x_{11}(\mathbf{p}) = \frac{(1/2)(\omega_1 \cdot \mathbf{p})}{p_1}, x_{21}(\mathbf{p}) = \frac{(1/2)(\omega_1 \cdot \mathbf{p})}{p_2}, x_{12}(\mathbf{p}) = \frac{(1/2)(\omega_2 \cdot \mathbf{p})}{p_1}, x_{22}(\mathbf{p}) = \frac{(1/2)(\omega_2 \cdot \mathbf{p})}{p_2}$$

Step 2: Impose market-clearing condition on $L-1$ markets.

Provided utility functions are strongly monotone (they are in this example), the market version of Walras' law allows us to conclude the market-clearing condition for both commodities from the market-clearing condition of just one commodity. Let's write down the market-clearing condition of commodity 1. We need to find some $\mathbf{p}$ such that:

$$\underbrace{x_{11}(\mathbf{p}) + x_{12}(\mathbf{p})}_{\text{total demand}} = \underbrace{\omega_{11} + \omega_{12}}_{\text{total supply}}$$

Expanding $x_{11}(\mathbf{p})$ and $x_{12}(\mathbf{p})$ gets us:

$$\frac{(1/2)(p_1\omega_{11} + p_2\omega_{21})}{p_1} + \frac{(1/2)(p_1\omega_{12} + p_2\omega_{22})}{p_1} = \omega_{11} + \omega_{12}$$

Step 3: Solve for equilibrium price.

The Walrasian equilibrium price is not unique. We might choose to normalize $p_1 = 1$, in which case the above equality simplifies:

$$\frac{1}{2}(\omega_{11} + p_2\omega_{21}) + \frac{1}{2}(\omega_{12} + p_2\omega_{22}) = \omega_{11} + \omega_{12}$$

$$p_2\omega_{21} + p_2\omega_{22} = \omega_{11} + \omega_{12}$$

$$p_2 = \frac{\omega_{11} + \omega_{12}}{\omega_{21} + \omega_{22}} = \frac{3}{3} = 1$$

Therefore, $(1, 1)$ is a Walrasian equilibrium price vector.

2.3 *The First Welfare Theorem may fail in the presence of externality.* An implicit assumption in the exchange economy model is that each agent i derives utility from only her **private consumption**. That is, u_i only depends on $\mathbf{x}_i$. However, this assumption can be unrealistic for several reasons:

- There can be **public goods** that everyone gets to enjoy, which are hard to write as private allocations to different agents. Clean air is a textbook example. This is modeled as $u_i(\mathbf{x}_i, m)$ where m is the consumption of public good (which does not carry an index for consumer i since it is not private consumption but common across all consumers in the same economy).

- In a behavioral economics setting, a consumer's utility for a good might depend on the **price she paid**. For instance, buying a pair of shoes at a steep discount might make a consumer happy for **psychological** reasons (and lead to an accumulation of unworn shoes bought at various sales...) This is modeled as $u_i(\mathbf{x}_i, \mathbf{p})$ where $\mathbf{p}$ is the price vector.
- In a model with firms and production, consumers may have preference over how the firms run their **production process**. For example, consumers may find a consumption good less enjoyable if they know it was produced in an unethical way. In an economy with one firm, this is modeled as $u_i(\mathbf{x}_i, \mathbf{y})$ where $\mathbf{y} \in \mathbb{R}^L$ is the production vector chosen by the firm.[20]
- A growing literature on happiness suggests it is not the absolute level of consumption that makes people happy, but rather **comparisons** of own consumption to consumptions of **social peers**. This can be modeled by assuming that consumers in the economy live on the vertices of some graph $\mathcal{G}$, whose edges represent connections in a social network. The utility of consumer i could look like $u_i\left(\mathbf{x}_i, (\mathbf{x}_j)_{\{j \text{ s.t. } (i,j) \text{ are neighbors in } \mathcal{G}\}}\right)$.

In each of these modifications to the general equilibrium model, first welfare theorem fails. To work through a concrete example, suppose we are back in the economy with $I = L = 2$, endowments $\omega_1 = (1,2)$ and $\omega_2 = (2,1)$. However, consumer 1 has "**social preference**" with respect to the consumption of the first commodity, so that her utility is actually:

$$\ln(x_{11} + x_{12}) + \ln(x_{21})$$

The utility of consumer 2 is unchanged. When a price is announced, consumer 2 first solves his Marshallian demands (x_{12}, x_{22}). Consumer 1 takes this x_{12} as given and maximizes her utility over choice of x_{11} and x_{21}.

Since the utility function of consumer 2 is unchanged, his demand is as before:

$$x_{12}(\mathbf{p}) = \frac{(1/2)(\omega_2 \cdot \mathbf{p})}{p_1}, x_{22}(\mathbf{p}) = \frac{(1/2)(\omega_2 \cdot \mathbf{p})}{p_2}$$

Consumer 1 optimizes by equating the ratio of marginal utilities from x_{11} and x_{21} with the ratio of their prices, that is to say she finds $\hat{x}_{11}$ and $\hat{x}_{21}$ so that:

$$\frac{1/(\hat{x}_{11} + x_{12}(\mathbf{p}))}{1/(\hat{x}_{21})} = \frac{p_1}{p_2}$$

Rearranging gives

$$\hat{x}_{21} = \frac{p_1}{p_2}(\hat{x}_{11} + x_{12}(\mathbf{p}))$$

Normalizing $p_2 = 1$, this says $\hat{x}_{21} = p_1(\hat{x}_{11} + x_{12}(\mathbf{p}))$. Making use of the budget constraint for consumer 1,

$$p_1\hat{x}_{11} + 1 \cdot (p_1(\hat{x}_{11} + x_{12}(\mathbf{p}))) = p_1 + 2$$

$$\Rightarrow x_{11}(\mathbf{p}) = \frac{p_1 + 2 - p_1 x_{12}(\mathbf{p})}{2p_1}$$

[20] This was the topic of my senior thesis.

Imposing market-clearing condition in commodity 1,

$$x_{11}(\mathbf{p}) + x_{12}(\mathbf{p}) = 3 \Rightarrow \frac{p_1 + 2 - p_1 x_{12}(\mathbf{p})}{2p_1} + x_{12}(\mathbf{p}) = 3$$

$$\frac{p_1 + 2 - (1/2)(2p_1 + 1)}{2p_1} + \frac{(1/2)(2p_1 + 1)}{p_1} = 3$$

$$\overset{\text{algebra}}{\Rightarrow} p_1 = \frac{3}{8}$$

This implies a Walrasian equilibrium allocation with $x_{12}^* = x_{12}(\frac{3}{8}, 1) = \frac{2(\frac{3}{8})+1}{2(\frac{3}{8})}$ and $x_{11}^* = \frac{\frac{3}{8}+2-x_{12}^*}{2(\frac{3}{8})}$. But $x_{11}^* > 0$, so we can consider a **modification** to the Walrasian equilibrium allocation where consumer 1 gives all of the commodity 1 in her Walrasian equilibrium allocation to consumer 2. This makes consumer 1 no worse off since she treats consumption of commodity 1 by either agent as **perfect substitutes**. However, this change makes consumer 2 strictly better off. Therefore, the Walrasian equilibrium allocation is not Pareto efficient.

3 Bystander effect and public goods

3.1 *The bystander effect.* When a person experiences **grave physical danger** (heart attack, assault, etc.), a lone nearby **bystander** often offers help (performing CPR, calling 911, etc.) An important observation in social psychology is that a bystander becomes **less likely** to help a nearby victim when other bystanders are present. In fact, the likelihood of each individual offering help could diminish so quickly with the number of bystanders that the probability of at least one person helping the victim actually **decreases** with the number of bystanders. A particularly well-known example is the murder of **Kitty Genovese** in 1964. An attacker stabbed Genovese in her apartment. The crime took place over a 30 minute interval, during which Genovese's screams for help could be heard by any of the 38 neighbors in the apartment complex. However, no neighbor offered help.

3.2 *Bystander effect as a public goods problem.* The logic of the bystander effect is similar to the **free-rider problem** for public goods. Each bystander cares, at least to small degree, about the welfare of the victim. The welfare of the victim is thus a public good in the eyes of the bystanders. However, contributing to the public good (i.e. helping the victim) is costly and only one contributor (helper) is needed. So every bystander would rather if someone else contributed to the public good so that they can simply enjoy the benefit without incurring a private cost.

More formally, suppose every bystander gets utility $-z$ if the victim dies, utility 0 if the victim lives, where $z > 0$. The victim lives if and only if at least one person helps. Every helper incurs a **private cost** of $c > 0$, which carries different interpretations depending on the victim's situation. In the case of assault, c might refer to the **physical danger** of confronting an armed criminal. In the case of a heart attack, c might refer to the **time and effort cost** of performing CPR or the worry of inflicting unintentional injury on the victim and facing **future lawsuits**. Suppose $z > c$, so that a lone bystander would always help. We show a strong form of free-riding: there is an "equilibrium"[21] where the probability that the victim lives strictly decreases in the number of bystanders.

Fix $n > 1$, the number of bystanders. Every bystander chooses a probability of helping, $\bar{p}(n) \in [0, 1]$, which maximizes their own expected utility given that every other bystander symmetrically picks the same $\bar{p}(n)$.That is:

[21] More precisely, a symmetric, mixed Nash equilibrium. We will formally introduce game theory next week.

$$\bar{p}(n) \in \arg\max_{p\in[0,1]} \{Q_n(p)\cdot 0 + (1 - Q_n(p))\cdot(-z) - pc\}$$

where $Q_n(p)$ is the probability that at least 1 bystander helps the victim when $n-1$ bystanders each helps with probability $\bar{p}(n)$ while one bystander helps with probability p. This means $Q_n(p) = 1 - [(1-p(n))^{n-1}\cdot(1-p)]$.
Notice that:

- $\bar{p}(n) \neq 1$, else $Q_n(p) = 1$ for every choice of p. This creates the incentive for every bystander to choose $p = 0$ instead and fully free-ride on the efforts of others, so it is not an equilibrium.
- $\bar{p}(n) \neq 0$, else every bystander has the incentive to choose $p = 1$ instead since $z > c$ and they have no one to free-ride on. Again, this is not an equilibrium.

So the optimization problem must admits an argmax in the **open interval** $(0,1)$. However, the objective

$$(1 - Q_n(p))\cdot(-z) - pc = -\left[(1-\bar{p}(n))^{n-1}\cdot(1-p)\right]\cdot z - pc$$

is **linear** in p. The only time when an interior argmax exists is when the two extreme points $p = 0$ and $p = 1$ attain the **same value** in the objective function. That is to say, every bystander is **exactly indifferent** between helping and not helping. This requires:

$$-c = -(1-\bar{p}(n))^{n-1}\cdot z$$

$$\bar{p}(n) = 1 - (c/z)^{1/(n-1)}$$

As we expect, $\bar{p}(n)$ is strictly decreasing in n due to free-riding. What is more, the probability that at least one out of n bystander helps the victim is:

$$1 - (1-\bar{p}(n))^n = 1 - (c/z)^{n/(n-1)}$$

which is also **strictly decreasing** in n and has a limit of $1 - (c/z)$. In particular, if say $c = 7$, $z = 10$, then:

- There is 100% chance that the victim receives help when there is 1 bystander
- There is 51% chance that the victim receives help when there are 2 bystanders
- There is 41% chance that the victim receives help when there are 3 bystanders
- There is 31% chance that the victim receives help when there are 38 bystanders

What can policy do to mitigate the bystander effect? One idea is to decrease c, the cost of helping. For instance, **Good Samaritan laws** offer legal protection to those who give reasonable assistance to people in peril. The Good Samaritan Act of Ontario, Canada reads:

> Despite the rules of common law, a person described in subsection (2) who voluntarily and without reasonable expectation of compensation or reward provides the services described in that subsection is not liable for damages that result from the person's negligence in acting or failing to act while providing the services, unless it is established that the damages were caused by the gross negligence of the person. 2001, c. 2, s. 2 (1).

Such laws are meant to dissuade the fear of future legal retaliation against a helper who unintentionally injures the victim, which enters into the cost calculation of a bystander.

4 Technical appendix (optional)

We give a proof outline for the existence of Walrasian equilibrium in an exchange economy. The idea is to construct a "**Walrasian auctioneer on steroids**", who changes commodity prices in the following way:

- Step 1: Observe which of the L commodities suffer the **highest level of excess demand**. Call this subset of commodities $\Lambda \subseteq \{1, 2, ..., L\}$.
- Step 2: Change the price vector such that any commodity **outside** of Λ has a price of 0 (!!!).
- Step 3: goto Step 1

So, a typical sequence of prices generated by this auctioneer in an economy with $L = 3$ commodities might look like:

$$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}) \xrightarrow{\Lambda=\{1\}} (1, 0, 0) \xrightarrow{\Lambda=\{2,3\}} (0, \frac{2}{5}, \frac{3}{5}) \xrightarrow{\Lambda=\{1\}} ...$$

It seems clear that starting with the price vector $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, this crazy auctioneer will **never converge** on a Walrasian equilibrium price, since he will always assigns a price of 0 to some commodity. However, this behavior depends on the initial condition. If there is a price vector where all 3 commodities are "equally in excess demand", then initializing the auctioneer at this price vector leads to **stable** behavior:

$$(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}) \xrightarrow{\Lambda=\{1,2,3\}} (\frac{1}{2}, \frac{1}{3}, \frac{1}{6}) \xrightarrow{\Lambda=\{1,2,3\}} (\frac{1}{2}, \frac{1}{3}, \frac{1}{6}) \xrightarrow{\Lambda=\{1,2,3\}} ...$$

Furthermore, it turns out any price vector that equalizes the excess demand of all commodities must actually clear the market (i.e. the excess demand of every good is the same and equals to 0). So the problem of finding a market-clearing price boils down to the problem of finding a **stable initial condition** for this auctioneer.

To be more precise, we first recognize since Marshallian demand is homogeneous of degree 0 in $(\mathbf{p}, Y)$, it is without loss to assume the price vector of the economy lies in the unit simplex:

$$\Delta := \left\{ \mathbf{p} \in \mathbb{R}^L_+ : \sum_{\ell=1}^{L} p_\ell = 1 \right\}$$

Technically, only those vectors that lie in the interior of Δ, that is to say

$$\text{int}(\Delta) := \left\{ \mathbf{p} \in \mathbb{R}^L_{++} : \sum_{\ell=1}^{L} p_\ell = 1 \right\}$$

are legitimate price vectors, since we do not formally allow any commodity to have a price of zero. However, the crazy auctioneer might change the price so that some commodities have a price of exactly 0, so that we do need to work with Δ.

Consider the excess demand function that outputs, for every price vector in the interior of Δ, an L-dimensional vector of excess demands:

$$Z : \text{int}(\Delta) \to \mathbb{R}^L$$

$$Z(\mathbf{p}) := \sum_{i=1}^{I} \mathbf{x}_i(\mathbf{p}, \omega_i \cdot \mathbf{p}) - \sum_{i=1}^{I} \omega_i$$

where $\mathbf{x}_i(\mathbf{p}, Y)$ is the Marshallian demand of consumer i under prices $\mathbf{p}$ and income Y. The set of "commodities in most excess demand" is a function of the price vector, defined as:

$$\Lambda(\mathbf{p}) := \begin{cases} \{\ell : Z_\ell(\mathbf{p}) = \max_k Z_k(\mathbf{p})\} & \text{for } \mathbf{p} \in \text{int}(\Delta) \\ \{\ell : p_\ell = 0\} & \text{for } \mathbf{p} \in \text{boundary}(\Delta) \end{cases}$$

To interpret, for a "usual" price vector in $\mathbb{R}^L_{++}$, the most excessively demanded commodities are those with the highest excess demand according to Z. But for those "price vectors" that include 0's, the set of most excessively demand commodities are precisely those commodities with 0 price. (Presumably every consumer demands an **infinite amount** of such free commodities.)

The price adjustment process of the extremist Walrasian auctioneer is thus a correspondence[22]:

$$\Phi : \Delta \rightrightarrows \Delta$$

$$\Phi(\mathbf{p}) := \{\mathbf{q} \in \Delta : q_\ell = 0 \text{ for every } \ell \notin \Lambda(\mathbf{p})\}$$

One can show that any **fixed-point** of Φ, that is to say any $\mathbf{p}^*$ with the property that $\mathbf{p}^* \in \Phi(\mathbf{p}^*)$, must be a Walrasian equilibrium of the economy.

Exercise A: Show that if $\mathbf{p}^* \in \Delta$ satisfies $\mathbf{p}^* \in \Phi(\mathbf{p}^*)$, then in fact $\mathbf{p}^* \in \text{int}(\Delta)$.

Exercise B: Show that $\mathbf{p}^* \in \Delta$ satisfies $\mathbf{p}^* \in \Phi(\mathbf{p}^*)$, then $Z(\mathbf{p}^*) = \mathbf{0}$.

Therefore, if Φ has a fixed point, then the exchange economy has a Walrasian equilibrium. The task, then, is to demonstrate Φ has at least one fixed point.

Your first guess might be to apply Brouwer's fixed-point theorem from differential topology. Indeed, Brouwer's result guarantees the existence of a fixed point for a function $f : A \to A$, under regularity conditions. However, the Φ we are dealing with is not a function, but a correspondence. Therefore the theorem we need is a generalization of Brower's theorem to correspondences, called **Kakutani fixed-point theorem**.

[22] In general, the notation $h : A \rightrightarrows B$ is equivalent to $h : A \to 2^B$. That is, h evaluated at any element of A returns a **subset** of B (instead of an element from B as in the case of a function). We call h a **correspondence** from A to B. In our case, $\Phi(\mathbf{p})$ returns the set of prices that the crazy Walrasian auctioneer considers acceptable when the current price is $\mathbf{p}$. Since his only constraint is that any commodity outside of $\Lambda(\mathbf{p})$ get a price of 0, there typically exist multiple acceptable new prices.

Theorem. *(Kakutani fixed-point theorem) Let $A \subseteq \mathbb{R}^n$ be non-empty, compact, and convex. Suppose $f : A \rightrightarrows A$ is such that $f(a)$ is nonempty and convex for every $a \in A$. Also, suppose*

$$\mathit{Graph}(f) := \{(a, a') \in A \times A | a' \in f(a)\}$$

is a closed subset of $A \times A$. Then there exists $a^* \in A$ such that $a^* \in f(a^*)$.

To finish the proof, take $A = \Delta$, $f = \Phi$ in the statement of Kakutani's theorem. Then we just need to verify that if the exchange economy satisfies the regularity conditions in the hypotheses of the Walrasian equilibrium existence theorem, then auctioneer's price adjustment process satisfies the hypotheses in Kakutani fixed-point theorem. This involves some tedious analysis but the steps are fairly standard.

Economics 1011a ▷ Section 9 : The Game[23] 11/11/2014

(1) Payoff matrix; **(2)** Normal-form game; **(3)** How to solve for NE; **(4)** NE existence

TF: Kevin He (he02@fas.harvard.edu)

1 Two-player game in a payoff matrix

1.1 *Interpreting the payoff matrix.* Here is the familiar **payoff matrix** representation of a two-player game.

	L	R
T	1,1	0,0
B	0,0	2,2

Player 1 (**P1**) chooses a row (**T**op or **B**ottom) while player 2 (**P2**) chooses a column (**L**eft or **R**ight). Each cell contains the payoffs to the two players when the corresponding pair of strategies is played. The first number in the cell is the payoff to P1 while the second number is the payoff to P2. (By the way, this game is called the **game of assurance**.)

Two important things to keep in mind:

- The two players choose their strategies **simultaneously**. That is, P2 cannot observe which row P1 picks when choosing his column.
- The terminology "payoff matrix" is slightly misleading. The numbers that appear in a payoff matrix are actually **Bernoulli utilities**, not monetary payoffs. In particular, in computing the expected utility of each player under a mixed-strategy profile, we simply take a weighted average of the matrix entries – there is no need to apply a "utility function" to the entries before taking the average as they are already denominated in utils.

1.2 *Nash equilibrium from a prediction point-of-view.* To motivate why we use Nash equilibrium to analyze games, it is important to know that there are actually two area of research that go by the name of "game theory". The full names of these two areas are "**combinatorial game theory**" and "**equilibrium game theory**". Despite the similarity in name, these two versions of game theory have quite different research agendas. The most salient difference is that combinatorial game theory studies well-known board games like chess and Go where there exists (theoretically) a "**winning strategy**" for one player. Combinatorial game theorists aim to find these winning strategies, thereby solving the game. On the other hand, no "winning strategies" (usually called **"dominant strategies**" in our lingo) exist for most games studied by equilibrium game theorists[24]. In the payoff matrix we just examined, due to the simultaneous-move condition, there is no one strategy that guarantees P1 always "wins" regardless of how P2 plays, in contrast to the existence of such guaranteed wins (or ties) in, say, tic-tac-toe.

[23] I just lost. [http://en.wikipedia.org/wiki/The_Game_(mind_game)]
"How to solve for NE" and "Examples of 2x2 Games with Different Nash Equilibria" written by Zhenyu Lai, a previous TF. "Periodic Table of 2x2 Games" was compiled by Bryan Bruns.

[24] The prisoner's dilemma is an exception here.

If a game has a dominant strategy for one of the players, then it is straight-forward to predict its outcome under optimal play. The player with the dominant strategy will employ this strategy and the other player will do the best they can to minimize their losses. However, predicting outcome in a game without dominant strategies requires the analyst to make assumptions. These assumptions are usually called **equilibrium assumptions** and give "equilibrium game theory" its name.

The most famous equilibrium assumption is the Nash equilibrium. Intuitively speaking,

"Definition". A **Nash equilibrium** is a strategy profile where no player has a strictly profitable **unilateral** deviation.

Said another way, a Nash equilibrium is where no single player can improve upon her own payoff, taking as given the actions of others. It seems reasonable to rule out non-Nash equilibrium outcomes as the predictions of what would happen in the game under "optimal" play, because such outcomes are **not stable**. For example, at strategy profile (T, R) in a game of assurance, both players are getting a payoff of 0. But, a unilateral deviation from either guarantees the deviator a new payoff strictly larger than 0. So, analogous to how the existence of a blocking coalition make non-core allocations unreasonable predictions of what happens in an exchange economy, so too the existence of unilateral deviations make non-Nash strategy profiles unreasonable predictions of the outcome in a game.

This is not to say all Nash equilibria are "good" predictions of a game's outcome, but it seems uncontroversial that non-Nash outcomes are not reasonable predictions. So being a Nash equilibrium is at least a **necessary condition** for a reasonable prediction. There exist other equilibrium assumptions that make different predictions about what happens in a one-shot game. We do not have the time to cover them in this class, but the predictions they deliver are generallyh (unsurprisingly) subsets of Nash equilibria.

1.3 *Some examples.* In the game of assurance, we readily verify that both (T, L) and (B, R) are Nash equilibria. Note one of these two Nash equilibria Pareto dominates the other. In general, Nash equilibria **need not be Pareto efficient**. This is because the definition of a Nash equilibrium only accounts for the absence of profitable **unilateral** deviations. Indeed, starting from the strategy profile (T, L), if P1 and P2 agree to simultaneously change their strategies, then they will both be better off. However, these sorts of simultaneous deviations by a "coalition" are not allowed.

But wait, there's more! Suppose P1 plays $\frac{2}{3}T \oplus \frac{1}{3}B$.[25] Suppose P2 plays $\frac{2}{3}L \oplus \frac{1}{3}R$. This strategy profile is also a (mixed) Nash equilibrium! When P1 is playing $\frac{2}{3}T \oplus \frac{1}{3}B$, P2 gets an expected payoff of $\frac{2}{3}$ from playing L and an expected payoff of $\frac{2}{3}$ from playing R. Therefore, P2 has no profitable unilateral deviation because every strategy he could play, pure or mixed, would give the same payoff of $\frac{2}{3}$. Similarly, P2's mixed strategy $\frac{2}{3}L \oplus \frac{1}{3}R$ means P1 gets an expected payoff of $\frac{2}{3}$ whether she plays T or B, so P1 does not have a profitable deviation either.

More games of economic relevance appear in the appendix "Examples of 2x2 Games with Different Nash Equilibria". You are encouraged to study them. The second appendix, "Periodic Table of 2x2 Games", contains a complete enumeration of all strict ordinal 2x2 games, up to isomorphism. You do not need to study this periodic table as it is only meant to give you an idea of the variety of 2x2 games.

2 Normal-form game

2.1 *General definition of a two-player normal-form game.* The payoff matrix representation of a game is convenient, but it is not sufficiently general. In particular, it seems unclear how we can represent

[25] This is shorthand for playing T with $\frac{2}{3}$ probability and playing B with $\frac{1}{3}$ probability.

games in which players have **infinitely** many possible strategies, such as a **Cournot duopoly**, in a finite payoff matrix. We must therefore turn to the general definition.

Definition. A **two-player normal-form game** $\mathcal{G}$ consists of:

- Strategy sets S_1, S_2
- Utility functions $u_1 : S_1 \times S_2 \to \mathbb{R}$, $u_2 : S_1 \times S_2 \to \mathbb{R}$

We will often write $\mathcal{G} = (S_1, S_2, u_1, u_2)$.

To interpret, S_i is the **strategy set** of player i, that is to say the set the actions they can take in the game. When each player chooses a strategy simultaneously from their own strategy set, we get a **strategy profile** $(s_1, s_2) \in S_1 \times S_2$. Players derive payoff by applying their respective utility functions to the strategy profile.

As an aside, you will sometimes see the notation "$-i$", as in s_{-i} or S_{-i}. This means "other than i". In a game with two players, "player -1" really means "player 2". Similarly, S_{-2} really means S_1, i.e. the strategy set of P1.

The payoff matrix representation of a game is a **specialization** of this definition. In a payoff matrix, the elements of S_1 and S_2 are written as the names of the rows and columns, while the values of u_1 and u_2 at different members of $S_1 \times S_2$ are written in the cells. If $S_1 = \{s_1^A, s_1^B\}$ and $S_2 = \{s_2^A, s_2^B\}$, then the game $\mathcal{G} = (S_1, S_2, u_1, u_2)$ can be written in a payoff matrix:

	s_2^A	s_2^B
s_1^A	$u_1(s_1^A, s_2^A), u_2(s_1^A, s_2^A)$	$u_1(s_1^A, s_2^B), u_2(s_1^A, s_2^B)$
s_1^B	$u_1(s_1^B, s_2^A), u_2(s_1^B, s_2^A)$	$u_1(s_1^B, s_2^B), u_2(s_1^B, s_2^B)$

Conversely, the game of assurance can be converted into the standard definition by taking $S_1 = \{T, B\}$, $S_2 = \{L, R\}$, $u_1(T, L) = 1$, $u_1(B, R) = 2$, $u_1(T, R) = u_1(B, L) = 0$, $u_2(T, L) = 1$, $u_2(B, R) = 2$, $u_2(T, R) = u_2(B, L) = 0$,

The general definition allows us to write down games with infinite strategy sets. In a **duopoly** setting where firms choose own production quantity, their choices are not taken from a finite set of possible quantities, but are in principle allowed to be any positive real number. So, consider a game with $S_1 = S_2 = [0, \infty)$,

$$u_1(s_1, s_2) = p(s_1 + s_2) \cdot s_1 - C(s_1)$$

$$u_2(s_1, s_2) = p(s_1 + s_2) \cdot s_2 - C(s_2)$$

where $p(\cdot)$ and $C(\cdot)$ are the inverse demand function and cost function, respectively. Interpreting s_1 and s_2 as the quantity choices of firm 1 and firm 2, this is Cournot competition phrased as a normal-form game.

2.2 *Best response and Nash equilibrium.* Our heuristic definition of the Nash equilibrium involved the absence of profitable unilateral deviations. Another way to say this is that every player is playing optimally conditional on the strategy of the opponent. We now formalize this idea of conditional optimality.

Definition. The **individual best-response correspondences** $BR_1 : S_2 \rightrightarrows S_1$, $BR_2 : S_1 \rightrightarrows S_2$ are defined as:

$$\boxed{\begin{cases} BR_1(s_2) := \arg\max\limits_{\hat{s}_1 \in S_1} u_1(\hat{s}_1, s_2) \\ BR_2(s_1) := \arg\max\limits_{\hat{s}_2 \in S_2} u_2(s_1, \hat{s}_2) \end{cases}} \tag{35}$$

The **best-response correspondence** $BR : (S_1 \times S_2) \rightrightarrows (S_1 \times S_2)$ simply involves putting BR_1 and BR_2 into a vector:

$$\boxed{BR(s_1, s_2) := (BR_1(s_2), BR_2(s_1))} \tag{36}$$

To interpret, the individual best-response correspondences return the argmax of each player's utility function when the opponent plays some known strategy. Depending on P2's strategy, P1 may have multiple maximizers, all yielding the same utility. As a result, we must allow BR_1 and BR_2 to be **correspondences** rather than functions. To interpret the BR correspondence, suppose P1 and P2 play the game not just once, but for **multiple rounds**. In each round, each player plays the strategy that **would have been optimal** against their opponent's strategy from last round. Then $BR(s_1, s_2)$ is the strategy profile that would be played next round if (s_1, s_2) is played this round.

Example. Consider the game of rock-paper-scissors, written in payoff matrix form:

	rock	paper	scissors
rock	O	**II**	**I**
paper	**I**	O	**II**
scissors	**II**	**I**	O

where we abbreviate O$= (0,0)$, **I**$= (1,-1)$, **II**$= (-1,1)$. Then for $i \in \{1,2\}$, $BR_i(\text{rock}) = \text{paper}$, $BR_i(\text{paper}) = \text{scissors}$, $BR_i(\text{scissors}) = \text{rock}$. Starting with a particular strategy profile, say (rock, paper), repeated applications of the best response correspondence delivers a cyclic result:

$$\begin{pmatrix} \text{rock} \\ \text{paper} \end{pmatrix} \overset{BR}{\to} \begin{pmatrix} \text{scissors} \\ \text{paper} \end{pmatrix} \overset{BR}{\to} \begin{pmatrix} \text{scissors} \\ \text{rock} \end{pmatrix} \overset{BR}{\to} \begin{pmatrix} \text{paper} \\ \text{rock} \end{pmatrix} \overset{BR}{\to} \begin{pmatrix} \text{paper} \\ \text{scissors} \end{pmatrix} \overset{BR}{\to} \begin{pmatrix} \text{rock} \\ \text{scissors} \end{pmatrix} \overset{BR}{\to} \begin{pmatrix} \text{rock} \\ \text{paper} \end{pmatrix} \overset{BR}{\to} \ldots$$

P1 plays rock initially, which leads P2 to best respond with paper, which in turn leads P1 to best respond with scissors, etc. After 3 applications of the BR correspondence, the rock→paper→scissors→rock cycle is complete, but now rock appears as the strategy of P2 (since the length of the cycle is odd).

$$\begin{pmatrix} \textbf{ROCK} \\ \text{paper} \end{pmatrix} \overset{BR}{\to} \begin{pmatrix} \text{scissors} \\ \textbf{PAPER} \end{pmatrix} \overset{BR}{\to} \begin{pmatrix} \textbf{SCISSORS} \\ \text{rock} \end{pmatrix} \overset{BR}{\to} \begin{pmatrix} \text{paper} \\ \textbf{ROCK} \end{pmatrix} \overset{BR}{\to} \ldots$$

The best response correspondence turns out to be a crucial building block for Nash equilibrium.

Definition. A Nash equilibrium is a strategy profile $(s_1^*, s_2^*) \in S_1 \times S_2$ such that $(s_1^*, s_2^*) \in BR(s_1^*, s_2^*)$.

This definition views Nash equilibrium as a **fixed-point** of the best response correspondence. That is to say, imagine again a pair of players who play multiple rounds and who best respond to opponent strategy from last round. Then a Nash equilibrium is a profile (s_1^*, s_2^*) that the players could keep playing forever. Another way to say this that each player holds **correct belief** about what their opponent is going to play and best responds to this belief.

This definition relates to the previous, intuitive definition of a Nash equilibrium as the absence of profitable unilateral deviation. If s_1^* is a best response to s_2^*, then s_1^* is already the argmax in the optimization problem:

$$\max_{\hat{s}_1 \in S_1} \; u_1(\hat{s}_1, s_2^*)$$

P1 cannot do better by unilaterally choosing a different strategy, by the definition of argmax. Same remark holds for P2.

Exercise A. A strategy s_1 for P1 is **strictly dominated** there exists another strategy s_1' such that $u_1(s_1, s_2) < u_1(s_1', s_2)$ for every $s_2 \in S_2$. Explain why no player can play a strictly dominated strategy in a Nash equilibrium.

Does a Nash equilibrium always exist? The answer is no! As we have defined it, a Nash equilibrium requires both players to use pure strategies (i.e. mixed strategy not allowed). The game of rock-paper-scissors clearly does not have a pure strategy Nash equilibrium, since at any profile of pure strategies at least one of the players can do strictly better by deviating to the strategy that beats their opponent's strategy.

2.3 *Mixed best response and mixed Nash equilibrium.* We begin by defining a mixed strategy and extending the domain of our payoff functions to such mixed strategies. (As an aside, if we do not add qualifier to the phrases "Nash equilibrium" or "strategy profile", we are implicitly talking about the pure versions.)

Definition. In a game $\mathcal{G} = (S_1, S_2, u_1, u_2)$ where S_1, S_2 are finite, a **mixed strategy** σ_i for player i is a probability distribution on S_i. We write $\sigma_i(s_i)$ for the probability that the distribution σ_i assigns to $s_i \in S_i$. We also write $\Delta(S_i)$ for the set of all mixed strategies of player i.

In the game of assurance, $\frac{2}{3}T \oplus \frac{1}{3}B$ and $\frac{2}{3}L \oplus \frac{1}{3}R$ are examples of mixed strategies.

A player who plays a mixed strategy is **intentionally** introducing **randomness** into their actions. Instead of choosing a single strategy from S_i, player i flips a (biased) coin to determine which strategy will be played. In particular, with probability $\sigma_i(s_i)$ the strategy $s_i \in S_i$ will be played.

Two remarks:

- Pure strategies also count as mixed strategies – they are just **degenerate** distributions that put probability 1 on some s_i and probability 0 everywhere else.
- In the event that both players play mixed strategies, the randomizations of the two players are **independent**.

Next, we define the "extended payoff function" which returns the expected payoff of each player when a profile of mixed strategies is played.

Definition. The **extended payoff functions** $\tilde{u}_1 : \Delta(S_1) \times \Delta(S_2) \to \mathbb{R}$ and $\tilde{u}_2 : \Delta(S_1) \times \Delta(S_2) \to \mathbb{R}$ are defined by:

$$\boxed{\begin{aligned}\tilde{u}_1(\sigma_1,\sigma_2) &:= \sum_{s_1\in S_1}\sum_{s_2\in S_2} u_1(s_1,s_2)\cdot\sigma_1(s_1)\cdot\sigma_2(s_2)\\ \tilde{u}_2(\sigma_1,\sigma_2) &:= \sum_{s_1\in S_1}\sum_{s_2\in S_2} u_2(s_1,s_2)\cdot\sigma_1(s_1)\cdot\sigma_2(s_2)\end{aligned}}$$

That is, when the profile of mixed strategies (σ_1, σ_2) is played, $\tilde{u}_i(\sigma_1, \sigma_2)$ computes the expected payoff to player i by taking a weighted average across all possible $u_i(s_1, s_2)$, where the weights are given by

$$\sigma_1(s_1)\cdot\sigma_2(s_2)$$

the probability that pure strategy profile (s_1, s_2) is played when the two players mix independently.

Exercise B: For a convex set $V \subseteq \mathbb{R}^n$, call $f : V \to \mathbb{R}$ **affine**[26] if whenever $v_1, v_2, ..., v_r \in V$ and $\sum_{k=1}^r p_k = 1$ with $p_k \geq 0$, we have $f\left(\sum_{k=1}^r p_k v_k\right) = \sum_{k=1}^r p_k \cdot f(v_k)$. When $V, W \subseteq \mathbb{R}^n$ are both convex, call $g : V \times W \to \mathbb{R}$ **bi-affine** if $v \mapsto g(v, \bar{w})$ is affine for every fixed $\bar{w} \in W$ and $w \mapsto g(\bar{v}, w)$ is affine for every fixed $\bar{v} \in V$. Show that $\tilde{u}_1$ and $\tilde{u}_2$ are both bi-affine.

Mixed best response and mixed Nash equilibria are straight-forward generalizations of their pure counterparts, using $\tilde{u}_1$ and $\tilde{u}_2$ instead of u_1 and u_2.

Definition. The **mixed individual best response correspondences** $\overline{BR_1} : \Delta(S_2) \rightrightarrows \Delta(S_1)$, $\overline{BR_2} : \Delta(S_1) \rightrightarrows \Delta(S_2)$ are defined as:

$$\boxed{\begin{cases}\overline{BR_1}(\sigma_2) := \underset{\hat{\sigma}_1\in\Delta(S_1)}{\arg\max}\ \tilde{u}_1(\hat{\sigma}_1,\sigma_2)\\ \overline{BR_2}(\sigma_1) := \underset{\hat{\sigma}_2\in\Delta(S_2)}{\arg\max}\ \tilde{u}_2(\sigma_1,\hat{\sigma}_2)\end{cases}} \tag{37}$$

The **mixed best response correspondence** $\overline{BR} : (\Delta(S_1) \times \Delta(S_2)) \rightrightarrows (\Delta(S_1) \times \Delta(S_2))$ is defined as:

$$\boxed{\overline{BR}(\sigma_1,\sigma_2) := (\overline{BR_1}(\sigma_2), \overline{BR_2}(\sigma_1))} \tag{38}$$

Definition. A **mixed Nash equilibrium** is a mixed strategy profile $(\sigma_1^*, \sigma_2^*) \in \Delta(S_1) \times \Delta(S_2)$ such that $(\sigma_1^*, \sigma_2^*) \in \overline{BR}(\sigma_1^*, \sigma_2^*)$.

Unlike pure Nash equilibrium, a mixed Nash equilibrium always exists in a finite, two player game. This is (a version of) the main result of John Nash's doctoral dissertation.

Theorem. *(Existence of mixed Nash equilibrium). Let $\mathcal{G} = (S_1, S_2, u_1, u_2)$ be a two-player normal-form game where S_1, S_2 are finite. Then $\mathcal{G}$ has at least one mixed Nash equilibrium.*

Proof. See part 4. □

The following result is useful for finding mixed Nash equilibria in practice:

[26] This is related to the idea of a linear transformation. Indeed, every linear transformation is also affine, as you can verify. However, an affine transformation need not be linear, since linear transformation requires $f(\alpha v) = \alpha f(v)$ for every $\alpha \in \mathbb{R}$. When V is only assumed to be convex (and not necessarily a vector space), it might be the case that $v \in V$ but $2v \notin V$, so that $f(\alpha v)$ need not even be defined for every α.

Proposition. *Suppose* (σ_1^*, σ_2^*) *is a mixed Nash equilibrium. Then for any* $s_1 \in S_1$ *such that* $\sigma_1^*(s_1) > 0$*, we have* $\tilde{u}_1(s_1, \sigma_2^*) = \tilde{u}_1(\sigma_1^*, \sigma_2^*)$*. Similarly, for any* $s_2 \in S_2$ *such that* $\sigma_2^*(s_2) > 0$*, we have* $\tilde{u}_2(\sigma_1^*, s_2) = \tilde{u}_2(\sigma_1^*, \sigma_2^*)$*.*

Proof. Suppose we may find $s_1 \in S_1$ so that $\sigma_1^*(s_1) > 0$ but $\tilde{u}_1(s_1, \sigma_2^*) \neq \tilde{u}_1(\sigma_1^*, \sigma_2^*)$. In the event that $\tilde{u}_1(s_1, \sigma_2^*) > \tilde{u}_1(\sigma_1^*, \sigma_2^*)$, we contradict the optimality of σ_1^* in the maximization problem $\arg\max\ \tilde{u}_1(\hat{\sigma}_1, \sigma_2^*)$, for we should have just picked $\hat{\sigma}_1 = s_1$ instead. In the event that $\tilde{u}_1(s_1, \sigma_2^*) < \tilde{u}_1(\sigma_1^*, \sigma_2^*)$, we enumerate $S_1 = \{s_1^{(1)}, ..., s_1^{(r)}\}$ and use the fact that $\tilde{u}_1$ is bi-affine (from **Exercise B**) to expand:

$$\tilde{u}_1(\sigma_1^*, \sigma_2^*) = \sum_{k=1}^{r} \sigma_1^*(s_1^{(k)}) \cdot \tilde{u}_1(s_1^{(k)}, \sigma_2^*)$$

The term $\tilde{u}_1(s_1, \sigma_2^*)$ appears in the summation on the right with a strictly positive weight, so if $\tilde{u}_1(s_1, \sigma_2^*) < \tilde{u}_1(\sigma_1^*, \sigma_2^*)$ then there must exist another $s_1' \in S_1$ such that $\tilde{u}_1(s_1', \sigma_2^*) > \tilde{u}_1(\sigma_1^*, \sigma_2^*)$. But now we have again contradicted the fact that σ_1^* is a best mixed response to σ_2^*. □

Corollary. *In a mixed Nash equilibrium, each player is indifferent between any two strategies that they play with strictly positive probability.*

This "indifference condition" allows us to quickly determine the set of mixed Nash equilibria in a game, as we shall see.

3 How to solve for pure and mixed Nash equilibria

When given a payoff matrix, follow these steps to find all of its pure and mixed Nash equilibria.

1. Identify **dominated strategies**. These will never be part of a Nash equilibrium, pure or mixed. They can be safely **ignored**.
2. Find all the **pure-strategy Nash equilibria** by considering all cells in the payoff matrix.
3. Look for a **mixed** Nash equilibrium where one player is playing a pure strategy while the other is strictly mixing.
4. Look for a **mixed** Nash equilibrium where **both** players are strictly mixing.

(By strictly mixing I mean playing a mixed strategy that puts nonzero probabilities on at least two pure strategies.)

Let's apply these steps to the following payoff matrix.

	L	R	Y
T	$2, 2$	$-1, 2$	$1, 1$
B	$-1, -1$	$0, 1$	$-3, 0$
X	$1, -3$	$-2, 2$	$0, 1$

Step 1: Strategy X for P1 is strictly dominated by T. Strategy Y for P2 is strictly dominated by R. Hence we can restrict attention to the smaller, 2x2 game in the upper left corner.

Step 2: (T, L) is a pure Nash equilibrium as no player has a profitable unilateral deviation. (The deviation $L \to R$ does not **strictly** improve the payoff of P2, so it doesn't break the equilibrium.) At (T, R), P1 deviates $T \to B$, so it is not a pure strategy Nash equilibrium. At (B, L), P2 deviates $L \to R$. At (B, R), no player has a profitable unilateral deviation, so it is a pure strategy Nash equilibrium. In summary, the game has two pure-strategy Nash equilibria: (T, L) and (B, R).

Step 3: Now we look for mixed Nash equilibria where one player is using a pure strategy while the other is using a strictly mixed strategy. By corollary from before, if a player strictly mixes between two pure strategies, then they must be getting the **same payoff** from playing either of these two pure strategies.

Using this **indifference condition**, we quickly realize it cannot be the case that P2 is playing a pure strategy while P1 strictly mixes. Indeed, if P2 plays L then $u_1(T, L) > u_1(B, L)$. If P2 plays R then $u_1(B, R) > u_1(T, R)$.

Similarly, if P1 is playing B, then the indifference condition cannot be sustained for P2 since $u_2(R, B) > u_2(L, B)$.

Now suppose P1 plays T. Then $u_2(T, L) = u_2(T, R)$. This indifference condition ensures that any strictly mixed strategy of P2 $pL \oplus (1-p)R$ for $p \in (0, 1)$ is a mixed best response to P1's strategy. However, remember to ensure this is a mixed Nash equilibrium, we must **also make sure** P1 does not have any profitable unilateral deviation. This requires:

$$\tilde{u}_1(T, pL \oplus (1-p)R) \geq \tilde{u}_1(B, pL \oplus (1-p)R)$$

that is to say,

$$\begin{aligned} 2p + (-1) \cdot (1-p) &\geq (-1) \cdot p + 0 \cdot (1-p) \\ 4p &\geq 1 \\ p &\geq \frac{1}{4} \end{aligned}$$

Therefore, $(T, pL \oplus (1-p)R)$ is a strictly mixed Nash equilibrium where P2 strictly mixes when $p \in [\frac{1}{4}, 1)$.

Step 4: There are no mixed Nash equilibria where both players are strictly mixing. To see this, notice if $\sigma_1^*(B) > 0$, then

$$u_2(\sigma_1^*, L) = 2 \cdot (1 - \sigma_1^*(B)) + (-1) \cdot (\sigma_1^*(B)) < 2 \cdot (1 - \sigma_1^*(B)) + (1) \cdot (\sigma_1^*(B)) = u_2(\sigma_1^*, R)$$

So it cannot be the case that P2 is also strictly mixing, since P2 is not indifferent between L and R.

In total, the game has two pure Nash equilibria, (T, L) and (B, R), and infinitely many strictly mixed Nash equilibria, $(T, pL \oplus (1-p)R)$ where $p \in [\frac{1}{4}, 1)$.

4 Existence of mixed NE in finite games (optional)

Let's prove the main result from John Nash's doctoral dissertation!

Remember, Nash equilibrium is simply a **fixed-point** of the BR correspondence, much like how a Walrasian equilibrium is a fixed-point of the price adjustment correspondence of the "crazy auctioneer", Φ. Unsurprisingly, the proof involves, yet again, Kakutani fixed-point theorem.

Theorem. *(Kakutani fixed-point theorem) Let $A \subseteq \mathbb{R}^n$ be non-empty, compact, and convex. Suppose $f : A \rightrightarrows A$ is such that $f(a)$ is nonempty and convex for every $a \in A$. Also, suppose*

$$\mathit{Graph}(f) := \{(a, a') \in A \times A | a' \in f(a)\}$$

is a closed subset of $A \times A$. Then there exists $a^* \in A$ such that $a^* \in f(a^*)$.

We proceed to establish the existence result.

Theorem. *(Existence of mixed Nash equilibrium). Let $\mathcal{G} = (S_1, S_2, u_1, u_2)$ be a two-player normal-form game where S_1, S_2 are finite. Then $\mathcal{G}$ has at least one mixed Nash equilibrium.*

Proof. Let $A = \Delta(S_1) \times \Delta(S_2)$ and $f = BR$ in the statement of Kakutani fixed-point theorem. If $|S_1| = n$ and $|S_2| = m$, then $\Delta(S_1) = \{\mathbf{p} \in \mathbb{R}^n_+ : \sum_{k=1}^n p_k = 1\}$ and $\Delta(S_2) = \{\mathbf{p} \in \mathbb{R}^n_+ : \sum_{k=1}^m p_k = 1\}$. Evidently each of $\Delta(S_1), \Delta(S_2)$ is non-empty, compact, and convex, so that their Cartesian product $\Delta(S_1) \times \Delta(S_2)$ inherits these properties as well.

To establish that BR is nonempty, convex and has closed graph, it suffices to establish these properties for BR_1 and BR_2. (In fact it suffices to establish this for BR_1 since the other case is completely analogous.) Recall $BR_1(\sigma_2)$ is the argmax in the optimization problem

$$\max_{\hat{\sigma}_1 \in \Delta(S_1)} \tilde{u}_1(\hat{\sigma}_1, \sigma_2)$$

But $\hat{\sigma}_1 \mapsto \tilde{u}_1(\hat{\sigma}_1, \sigma_2)$ is affine by **Exercise B**, so it is *a fortiori* continuous. Further, the domain of optimization $\Delta(S_1)$ is compact. By the extreme-value theorem, the continuous function $\hat{\sigma}_1 \mapsto \tilde{u}_1(\hat{\sigma}_1, \sigma_2)$ on a compact domain must attain its maximum value. Therefore, the argmax exists. To see that $BR_1(\cdot)$ is convex-valued, suppose both $\tilde{\sigma}_1$ and $\tilde{\sigma}_1'$ are both in $BR_1(\sigma_2)$ for some $\sigma_2 \in \Delta(S_2)$. Then they are both global maximizers of $\tilde{u}_1(\cdot, \sigma_2)$, so in particular any convex combination of them $\lambda\tilde{\sigma}_1 + (1 - \lambda)\tilde{\sigma}_1'$ for $\lambda \in [0, 1]$ is also a global maximizer of $\tilde{u}_1(\cdot, \sigma_2)$, using the fact that $\tilde{u}_1(\cdot, \sigma_2)$ is affine.

It remains to establish the closed-graph property for B_1. This turns out to be a straight-forward implication of Berge's maximum theorem, which roughly states that the argmax of an optimization problem has closed-graph provided the objective function is jointly continuous in both the argument and the parameter.

Theorem. *(Berge's theorem of the maximum) Suppose $X \subseteq \mathbb{R}^n$ is compact. Suppose also $\Theta \subseteq \mathbb{R}^m$, $g : X \times \Theta \to \mathbb{R}$ is jointly continuous in its two arguments. Consider the argmax correspondence $x^* : \Theta \rightrightarrows X$ where*

$$x^*(\theta) := \arg\max_{x \in X} g(x; \theta)$$

Then x^* has closed graph.

Letting $X = \Delta(S_1)$, $\Theta = \Delta(S_2)$, and $g = \tilde{u}_1$ delivers the desired closed-graph property, where we make use of the bi-affinity of $\tilde{u}_1$ to imply its joint continuity in its two arguments. □

Exercise C. A two-player normal-form game $\mathcal{G} = (S_1, S_2, u_1, u_2)$ is called **symmetric** if $S_1 = S_2 = S$ and $u_1(s_1, s_2) = u_2(s_2, s_1)$ for every $s_1, s_2 \in S$. A mixed Nash equilibrium (σ_1^*, σ_2^*) is called symmetric if $\sigma_1^* = \sigma_2^*$. By modifying the proof of the existence of Nash equilibrium just given, show that a symmetric two-player normal-form game with finite S_1, S_2 has at least one symmetric mixed Nash equilibrium.

Appendix. Examples of 2x2 Games with Different Nash Equilibria

Prisoner's Dilemma

	Loyal	Defect
Loyal	2,2	0,3
Defect	3,0	1,1

Story: Two suspects in a crime are held in separate cells. If both stay quiet, each will be convicted of a minor offense. If one defects and agrees to act as a witness against the other, he will be freed and the other will be convicted of a major crime. If both defect, their testimony will be less valuable and each gets a moderate sentence.

Features: u_1(defect, loyal) > u_1(loyal, loyal) > u_1(defect, defect) > u_1(loyal, defect)

Nash Equilibrium: (Defect, Defect)

Intuition: The incentive to "free-ride" eliminates the possibility that the mutually desirable outcome (loyal, loyal) occurs. Regardless of what your opponent does, it is always optimal to defect.

Pure Coordination Game

	Early	Late
Early	1,1	0,0
Late	0,0	1,1

Story: Two teammates arrange to meet to work on a project. Unfortunately, neither can remember the meeting time. Both are required to be present for the task to be completed. Each person thus chooses between arriving early or late in the hope that the other person would also be present.

Features: u(early, early) = u (late, late) are preferred to u(late, early) and u(early, late)

Nash Equilibria: (Early, Early), (Late, Late), 1 mixed strategy NE

Intuition: Players desire to cooperate, but are unsure what action opponents will take. Payoffs are symmetric in both same-action outcomes and what matters is that players manage to coordinate.

Game of Assurance

	Movie	TV
Movie	2,2	0,0
TV	0,0	1,1

Story: Two people wish to coordinate on what to do in their free time. While they would be happy either both watching a movie or TV, each of them knows that the other party prefers watching a movie to watching TV.

Features: u(movie, movie) > u(TV, TV) are preferred to u(movie, TV) and u(TV, movie)

Nash Equilibria: (Movie, Movie), (TV, TV), 1 mixed strategy NE

Intuition: Players have common interest in a pareto superior equilibrium. Given sufficient common belief in opponent's action, the mutually preferred outcome exists as a focal point.

Battle of the Sexes

	Football	Opera
Football	2,1	0,0
Opera	0,0	1,2

Story: Two people wish to go out together. One person prefers football, the other prefers opera. But both are most unhappy if they end up going out alone.

Features: u_1(football, football) > u_1(opera, opera)
u_2(opera, opera) > u_2(football, football)

Nash Equilibrium: (Football, Football), (Opera, Opera), 1 mixed strategy NE

Intuition: Players agree that it is better to cooperate but disagree about best outcome.

Hawk-Dove (Chicken)

	Hawk	Dove
Hawk	-1,-1	2,0
Dove	0,2	1,1

Story: Two animals are fighting over some prey. Each can be passive (dove) or aggressive (hawk). Each prefers to be a hawk if the opponent is a dove because the hawk will always get the prey over the dove. Two doves will share the prey. Conversely, two hawks will fight each other to the death, leading to the least preferred outcome.

Features: u_1(hawk, dove) > u_1(dove, dove) > u_1(dove, hawk) > u_1(hawk, hawk)

Nash Equilibria: (Hawk, Dove), (Dove, Hawk), 1 mixed strategy NE

Intuition: It is mutually beneficial for players to play different strategies. Opposite of coordination game. Here, players share a rival, non-excludable resource (prey), and sharing comes at a cost (reduced food). Yet, this cost has to be balanced against likelihood of fighting to the death.

Matching Pennies

	Heads	Tails
Heads	1,-1	-1,1
Tails	-1,1	1,-1

Story: Two people each have a penny and simultaneously choose whether to show heads or tails. If they show the same side, person 2 pays person 1 a dollar. Conversely, if they show different sides, person 1 pays person 2 a dollar. Each person cares only about the amount of money received and prefers to receive more rather than less.

Features: P1 prefers (heads, heads) or (tails, tails).
P2 prefers (heads, tails) or (tails, heads).

Nash Equilibria: No pure strategy NE due to conflicting preferences. Mixed strategy NE exists.

Intuition: Given that players know what each other are going to do, there is always one player who is better off deviating.[1]

[1] Chiappori, Levitt, Groseclose (2002), and Palacios-Huerta (2003) find that how soccer players and goalies take penalty kicks resemble a mixed strategy equilibrium.

Zhenyu Lai: zlai@fas.harvard.edu

Periodic Table of 2x2 Games

Strict ordinal games
Symmetric games on diagonal axis

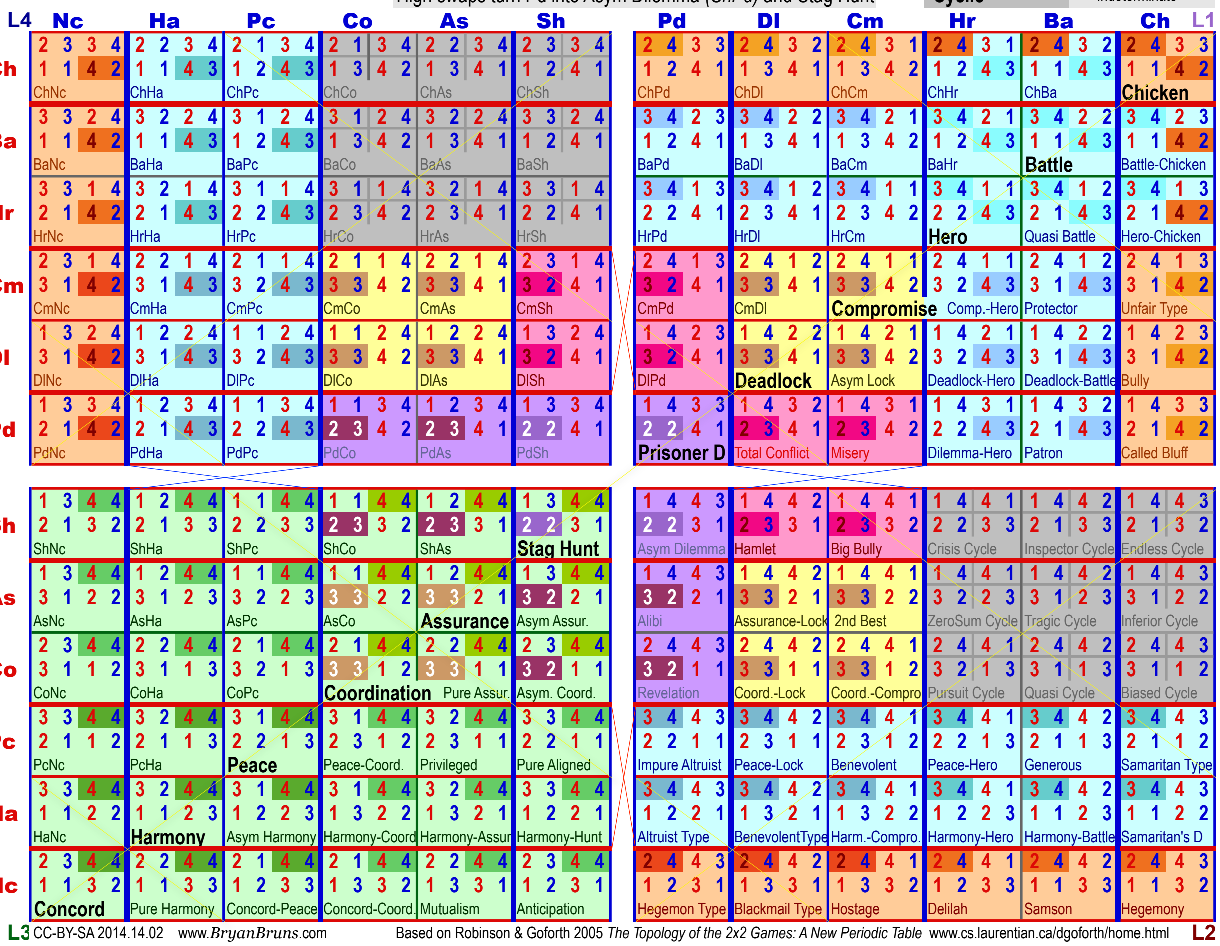
	Left		Right	
Up	1	4	3	3
Down	2	2	4	1

Payoffs: Row Column
Nash equilibrium
Pareto-inferior

Prisoner's Dilemma

Payoff swaps link neighboring games
1↔2 Low swaps form tiles of 4 games — Tile
2↔3 Middle swaps join tiles into 4 layers
3↔4 High swaps cross layers, bonding bands of tiles
Scrolling Prisoner's Dilemma to center shows relationships
Layers and table (toruses) wrap side-to-side & top-to-bottom
Layers differ by alignment of best payoffs
Payoffs from symmetric games form asymmetric games
High swaps turn Pd into Asym Dilemma (*ShPd*) and Stag Hunt

Payoff Families		
Win-win 4,4	Harmonious	
	Stag Hunt	
Biased 4,3	Battle	
Self-serving	Benevolent	Samaritan
Second Best 3,3		
Unfair 4,2	Winner	Loser
Inferior	Sad 3,2	
Dilemma 2,2	Alibi 3,2	
Cyclic	Indeterminate	



 www.BryanBruns.com Based on Robinson & Goforth 2005 *The Topology of the 2x2 Games: A New Periodic Table* www.cs.laurentian.ca/dgoforth/home.html

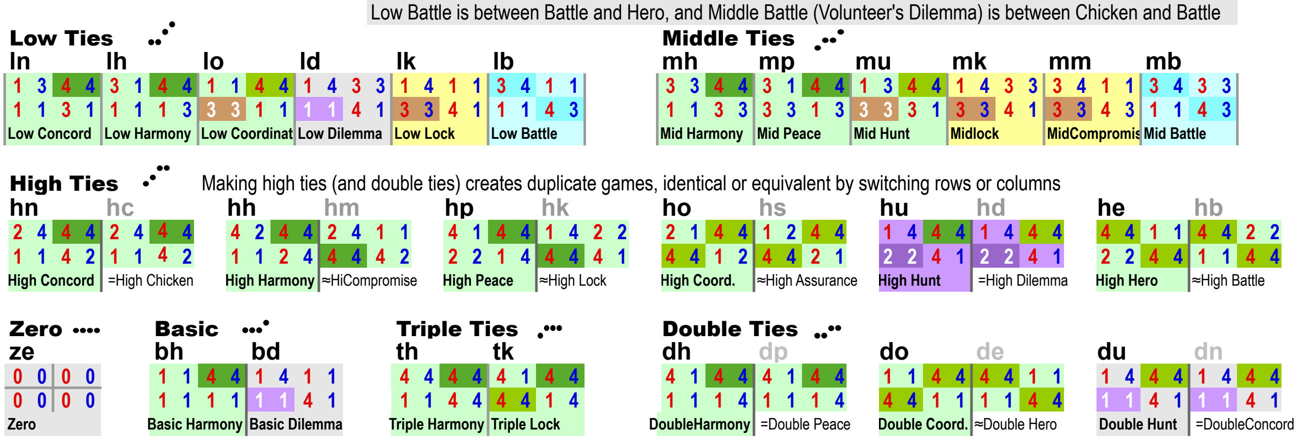
For more diagrams, explanations, and references, see *Changing Games: An Atlas of Conflict and Cooperation in 2x2 Games* www.2x2atlas.org

To find a game: Make ordinal 1<2<3<4. Put column with Row's 4 right row with Column's 4 up; find layer by alignment of 4s; find symmetric games with Row & Column payoffs.

Symmetric Games with Ties

Games with ties lie between strict ordinal games, linked by half-swaps that make or break ties. For example, Low Battle is between Battle and Hero, and Middle Battle (Volunteer's Dilemma) is between Chicken and Battle

Making high ties (and double ties) creates duplicate games, identical or equivalent by switching rows or columns



(1) extensive-form games and SPE; **(2)** Backwards induction; **(3)** Signaling game

TF: Kevin He (he02@fas.harvard.edu)

Today we will move away from normal-form games and study two new classes of games: **extensive-form games** and **signaling game**. The focus this week is more on problem-solving and less on developing a rigorous theory. As such, intuitive definitions will substitute for formal ones.

1 Extensive-form games and subgame-perfect equilibrium

1.1 *Game trees and extensive-form games.* An extensive-form game can be represented by a **game tree**, as shown in Figure 13.

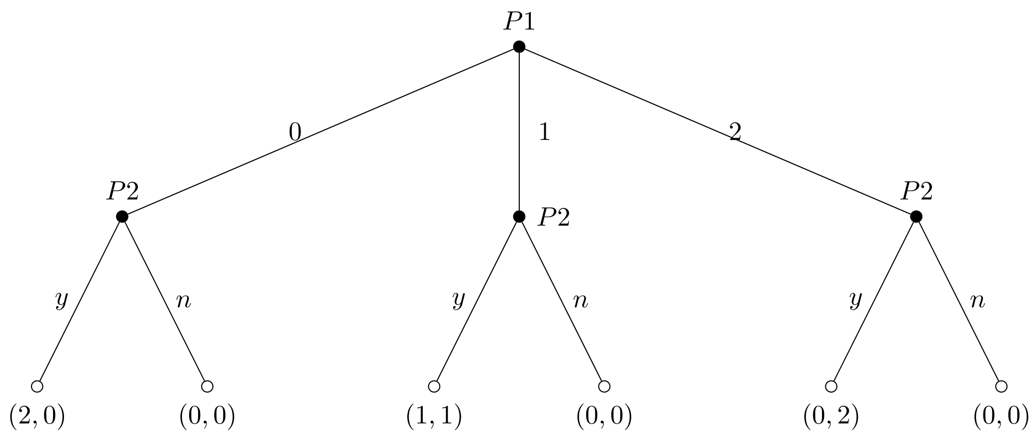


Figure 13: A game tree representation of the ultimatum game.

A game tree's nodes can be **terminal** (drawn as ∘ in this diagram) or **non-terminal** (drawn as • in this diagram). The game starts at the **root** of the tree. Every non-terminal node is labeled with the player who is to move. The **pure strategies** available to the mover correspond to the different **edges** below the node. Each edge is labeled with the name of the pure strategy. Every terminal node is labeled with a vector of **payoffs**, one for each player. Once a terminal node is reached, the game ends and the associated payoffs are realized. As in normal-form games, the payoffs are written in **Bernoulli utilities**, not monetary amounts.

The ultimatum game models an interaction between P1 and P2 who must split two identical, indivisible goods. P1 proposes an allocation. Then P2 approves or rejects the allocation. If the allocation is accepted, it is implemented. If it is rejected, then neither player gets any of the good. The player to move at the root of the game tree is P1. Her strategy set at the root is $\{0, 1, 2\}$, which correspond to giving 0, 1, 2 units of the good to P2. Regardless of which strategy P1 chooses, the game moves to a node where it is P2's turn to play. His pure strategy set at each of his three decision nodes is $\{y, n\}$, corresponding to accepting and rejecting the proposed allocation.

A **complete contingent strategy** (or "**strategy**" for short) for player i specifies which pure strategy i will pick at **every** node in the game tree where it is their turn to play[27]. Of course, for a typical

[27] We can also consider mixed strategies in an extensive-form game setting. However, for the purposes of this class, we restrict attention to pure strategies when dealing with extensive-form games.

strategy profile in an extensive-form game, not every node will be visited. For instance, if P1 in the ultimatum game decides to play 0, then P2's second and third decision nodes would never be reached. The reason why we require that P2's strategy specifies what he would do at every one of his decision nodes is that P2 does not *a priori* know what kind of strategy P1 will use, so every of P2's decision nodes could in principle be reached and every part of his plan is in principle relevant.

1.2 *Subgames and subgame-perfect equilibrium.* In the tree diagram representation of an extensive-form game, take any non-terminal node and consider the **subtree** rooted at this node. This subtree defines a **subgame** in the natural way. The ultimatum game in Figure 13 has **4 subgames**: 3 of the subgames start with a decision node of P2, while the entire game is a subgame of itself too.

A **subgame-perfect equilibrium** (SPE) is a strategy profile (i.e. a profile of complete contingent strategies) such that no player has a strictly profitable unilateral deviation in any subgame. For instance, the strategy profile where P1 plays 1 and P2 rejects if offered 0, accepts if offered 1, accepts if offered 2 is an SPE. To verify this, evidently P2 does not have any profitable unilateral deviation in any of 3 subgames that start with his decision node. Further, no deviation can yield P2 a higher payoff in the entire extensive game, provided that P1 is playing 1. At the same time, P1 has no profitable unilateral deviations either. Playing 0 yields P1 a payoff of 0 since P2's strategy plays "n" at that node. Similarly, playing 2 also yields P1 a payoff of 0. The offer would be accepted since P2's strategy plays "y" when offered 2, but P1 would be left with nothing after donating everything to P2.

What happens in the SPE? P1 will offer 1 and P2 will accept. Even though the subgame where P1 offers 0 is **never reached** along the equilibrium path, the strategy that P2 **would have played** in that subgame serves a crucial role in sustaining the SPE. The threat of rejection prevents P1 from keeping everything for herself. Further, the threat is credible, in the sense that rejection after receiving 0 makes P2 no worse off than acceptance.

1.3 *Nash equilibrium in an extensive-form game.* In contrast to the SPE requirement of no profitable unilateral deviation in any subgame, a **Nash equilibrium** (NE) in an extensive-form game only requires the absence of profitable unilateral deviations in the largest subgame, namely the entire game itself. The strategy profile where P1 plays 2, P2 plays "n" when P1 plays 0 or 1, plays "y" when P1 plays 2 is a NE. Certainly P2 has no profitable deviations since he is already getting the highest payoff possible. As for P1, she also has no profitable unilateral deviations, since offering 0 or 1 to P2 leads to rejection. Importantly, the behavior of P2 is **not optimal** at the subgame that would be reached if P1 plays 1. Indeed, P2 can get a higher payoff in this subgame by playing "y" instead. Yet this "mistake" does not translate into a profitable unilateral deviation in the game, since the un-optimized decision node was never reached along the equilibrium path. Even if P2 changes his behavior to always accept an offer of 1, this does not affect his payoff since P1 never offers him 1 anyway.

In some sense, this NE where P2 gets 2 is artificially sustained by an **non-credible threat**. P2 threatens to reject the proposal if P1 offers 1, despite the fact he has no incentive to carry out this threat if P1 really makes this offer. If P2 could, like Odysseus, tie himself to the mast and commit to rejecting the offer of 1 before the game starts, then the NE makes sense as a prediction of the game's outcome. But in the absence of such commitment devices, P2 behaves more like a quasi-hyperbolic discounter, who will fall to the temptation of the even-split offer, just as the discounter falls to the temptation of delaying physical exercise by yet another day. Indeed, ruling out such non-credible threats is the main motivation for preferring SPE over NE as a prediction of extensive-form game outcome.

2 Backwards induction

Backwards induction is an algorithm for finding an SPE payoff in an extensive-form game. The idea is to **replace subgames** with terminal nodes corresponding to the **SPE payoffs** in the deleted subgames. Suppose you know from previous calculations that a subgame in the game tree has a unique SPE payoff. But that means, once this subgame is reached in any SPE of the original game, there is only one way for the game to end. So, we might as well forget about this subgame and replace it with its SPE payoff. Successive replacements will eventually reduce the original extensive-form game into a single, terminal node.

If one knows the unique SPE payoff of some subgame, one should directly perform the replacement. Alternatively, there is a mechanical procedure to do backwards induction from scratch. Start with a non-terminal node **furtherest away** from the root of the game. Call this node v and suppose player i moves at v. Since we have picked the deepest non-terminal node, all of i's pure strategies at this node must lead to terminal nodes. So, in any SPE of the original game, i must pick the move that maximizes her own payoff at v.[28] Therefore, we have figured out the unique SPE payoff of the subgame starting at v. We may now replace the subgame at v with the selected payoff vector, keeping track of i's optimal action if we wish to find the SPE strategies in addition to the SPE payoffs. Repeat this procedure, working backwards from the nodes further away from the root of the game. Eventually, the game tree will be reduced to a single terminal node, whose payoff is the unique SPE payoff of the extensive-form game.

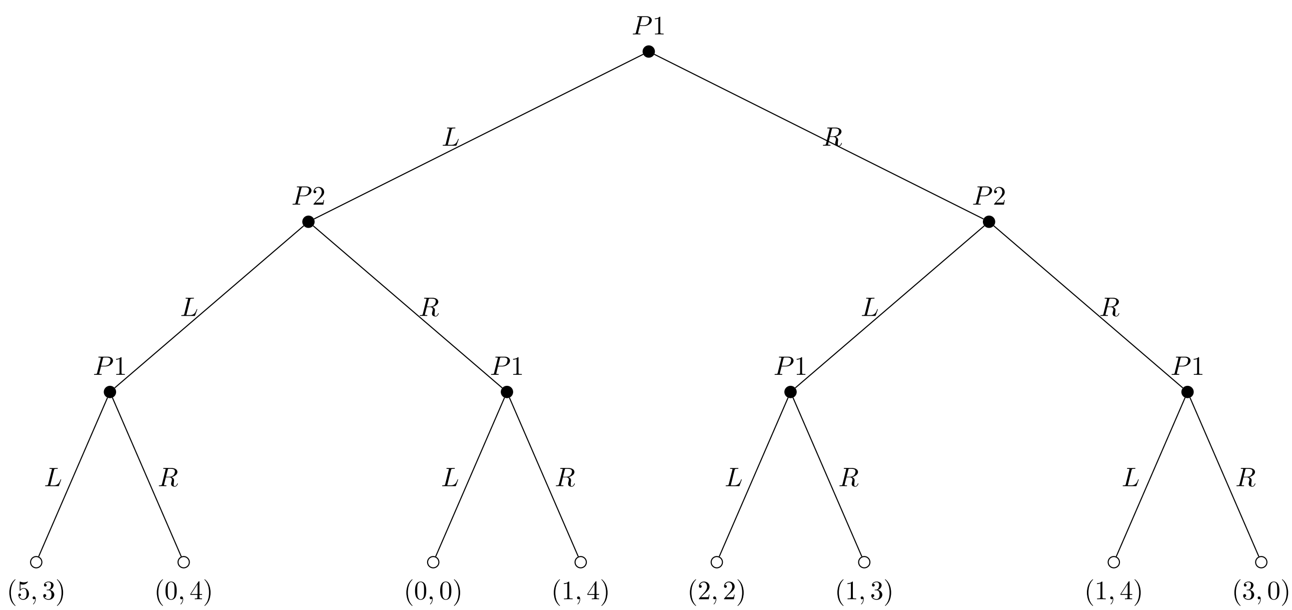


Figure 14: A extensive-form game in the game tree representation.

[28] Assume the backwards induction algorithm never encounters a situation where the player to move is indifferent between two pure strategies that both lead to the same private payoff.

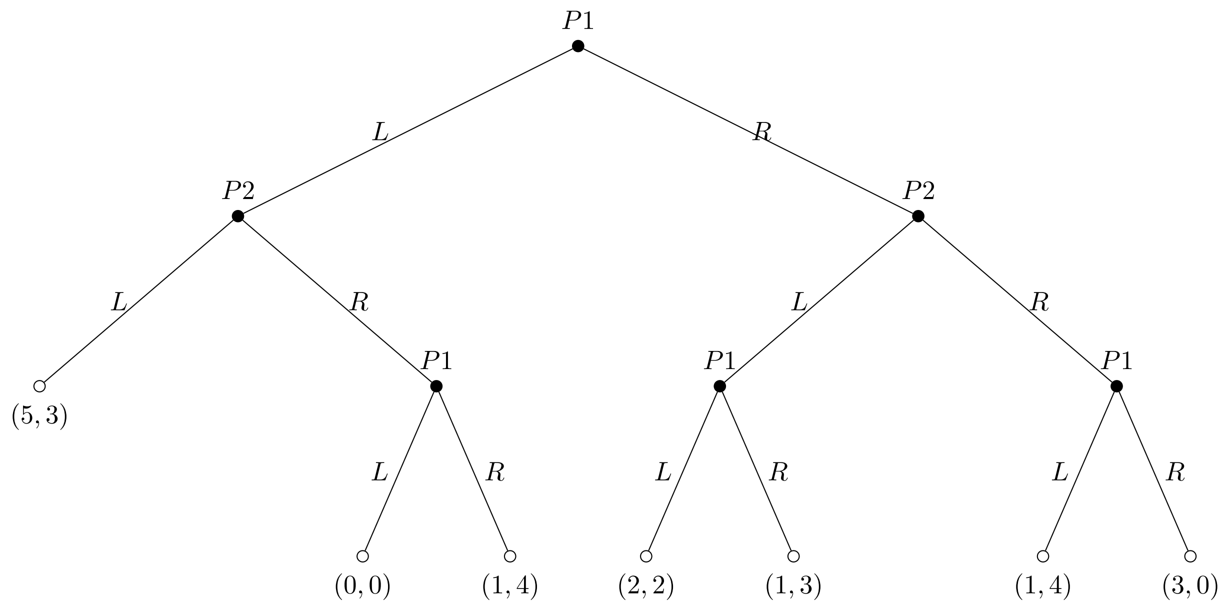


Figure 15: Backwards induction replaces subgames with terminal nodes associated with the SPE payoffs in those subgames. Here is the resulting game tree after one step of backwards induction.

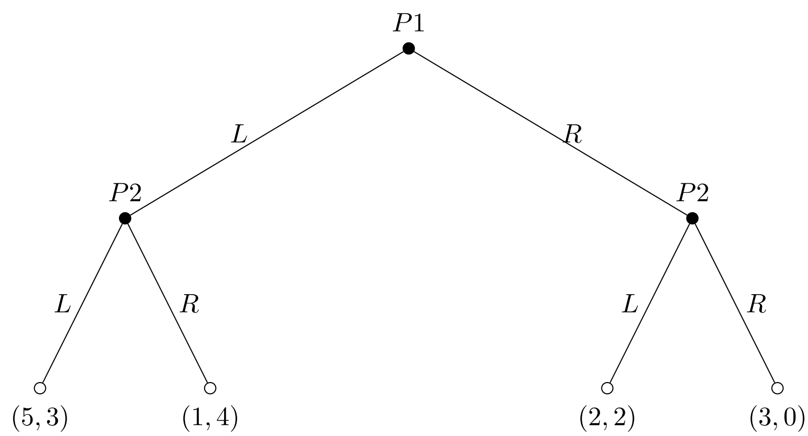


Figure 16: Backwards induction in progress. All nodes at depth 3 in the original tree have been eliminated.

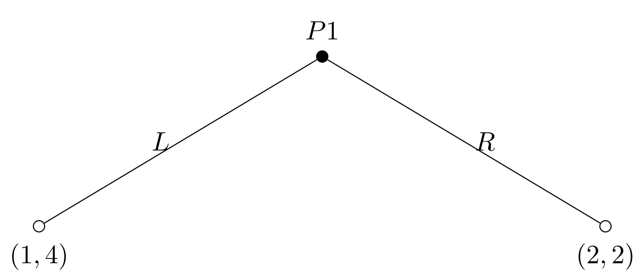


Figure 17: Backwards induction in progress. Only nodes with depth 1 remain.

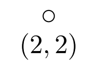


Figure 18: Backwards induction finds the unique SPE payoff in this game.

3 Signaling game

3.1 *Setup of a signaling game.* So far we have considered games of complete information. In signaling games, there are two complications:

- P1 now has **two "types"**, which can affect the payoff of **both** players
- P1 knows her type, but P2 does not know P1's type. This **asymmetry of information** often leads to inefficient outcomes.

Formally speaking, a signaling game (for the purposes of Economics 1011a) consists of:

- a set of two types $\Theta = \{\theta_L, \theta_H\}$ for P1
- A prior distribution π on Θ
- S_1, set of actions for P1[29]
- S_2, set of actions for P2
- Utility functions $u_i : S_1 \times S_2 \times \Theta \to \mathbb{R}$ for $i = 1, 2$

The type of P1 reflects some **hidden characteristic** that only P1 knows about, which nonetheless may affect the payoffs of both players in the game. The distribution π reflects P2's prior belief about the type of P1 at the start of the game. The distribution π is taken as a primitive in the description of the signaling game and its origin is usually outside of the model. Three possible sources are **past experience** (what fraction of P1's that P2 has met before turned out to have each type), **statistical evidence** (the distribution of types in the population where the current P1 is drawn), or **introspection** (so that P2 generates his prior belief in the same way a Bayesian statistician generates his Bayesian prior).

At the start of the game, P1 learns her own type, then picks an action (possibly dependent on her type) from S_1. Then, P2 observes this action and picks an action from S_2 in response. Finally, payoffs are realized according to $u_i(s_1, s_2, \theta)$.

3.2 *Some examples.* In the **market for lemons**, P1 is the seller and P2 is the buyer. The two types of P1 refer to whether the seller has a good car (θ_H) or a bad car (θ_L). P1 knows whether her car is good or not, but P2 only has a prior belief π on Θ, which assigns 50-50 chance to P1's type being either θ_H or θ_L. We may think of $S_1 = \{0, 1\}$ as a binary decision of whether or not sell the car, $S_2 = \mathbb{R}_+$ as the price that the seller offers, with utilities defined as:

- For any $p \in S_2$, $u_1(0, p, \theta_H) = G$, $u_1(0, p, \theta_L) = B$ with $G > B$. Also, $u_2(0, p, \theta) = 0$ for any $p \in S_2, \theta \in \Theta$. In words, if P1 decides not to sell then P1 gets G or B depending on her type (i.e. whether the car is good). P2 gets 0 payoff.

[29] We refer to S_1, S_2 as "actions" instead of "strategies" to avoid confusion, since a "strategy" in a signaling game has a different meaning than in a normal-form game.

- $u_1(1, p, \theta) = p$ for any $p \in S_2$, $\theta \in \Theta$. Also, $u_2(1, p, \theta_H) = G + g - p$, $u_2(1, p, \theta_L) = B + b - p$. If P1 sells, then P1 gets the price offered by P2 while P2 gets utility of $G + g$ or $B + b$ from the car, depending on P1's type, minus the price p.

Notice the **type of P1** enters into **P2's utilities** in the event $s_1 = 1$. This is a key feature of asymmetric information models, where the private information held by one player affects another player's payoffs.

Next let's look at the **college major choice** example from lecture, slightly rephrased. P1 is a student and P2 is an employer. The two types of P1 refer to whether the student has high ability (θ_H) or low ability (θ_L). P1 knows her own ability, but P2 only has a prior belief π on Θ, which assigns 50-50 chance to P1's type being either θ_H or θ_L. The set of actions for P1 is $S_1 = \{P, E\}$, corresponding to majoring in either Physics or Economics. The set of actions of P2 are $S_2 = \mathbb{R}_+$, interpreted as the wage offered to the student. The utilities are defined as:

- $u_1(P, w, \theta_H) = w - c_h$, $u_1(P, w, \theta_L) = w - c_l$. That is, the utility from getting paid wage w after majoring in Physics is w minus the cost of studying physics, where the cost spends on P1's type with $c_h < c_l$.
- $u_1(E, w, \theta) = w$ for all $\theta \in \Theta$, i.e. studying Economics is costless to all types.
- $u_2(s_1, w, \theta_H) = H - w$ and $u_2(s_1, w, \theta_L) = L - w$ for any $s_1 \in S_1$. P2's payoff is **unaffected** by whether P1 majored in Physics or Economics. However, the employer's payoff does depend on the ability type of the student, which translates into productivity level on the job.

3.3 *Perfect-Bayesian Equilibrium.* We cannot apply NE or SPE to a signaling game due to the presence of **incomplete information**. Roughly speaking, both NE and SPE would require that P2's action is "optimal" in some sense. But P2 does not know P1's type, which affects how P2's payoff depends on his action $s_2 \in S_2$. Therefore whether an action that P2 took was "**optimal**" (given the limitations imposed by information) depends on P2's **beliefs** about P1's type. As such, any reasonable notion of "equilibrium" in the presence of incomplete information must involve just not the actions that players take, but also the beliefs that the uninformed player (P2) holds. Furthermore, since P2 moves only after seeing $s_1 \in S_1$ from P1, the relevant beliefs are probably not the prior belief π, since the action that P1 took might convey some information about her type. For instance, the fact that a seller decides to sell her car might tell the buyer something about the quality of the car. The fact that a student majors in Physics might tell the employer something about his ability.

One possible equilibrium concept here is **perfect-Bayesian Equilibrium** (PBE) equilibrium. Heuristically speaking, this is just SPE but with the addition of posterior beliefs for P2 and some restrictions that require the posterior beliefs to be "correct". Formally, a PBE consists of:

- A strategy $\sigma_1 : \Theta \to S_1$ for P1
- A strategy $\sigma_2 : S_1 \to S_2$ for P2
- No type of P1 has a profitable unilateral deviation
- A family of posterior beliefs $\{\pi(\cdot|s_1) : s_1 \in S_1\}$
- $\pi(\cdot|s_1)$ is generated by Bayes' rule whenever possible
- P2's action maximizes his expected utility after any s_1 by P1, according to his posterior belief $\pi(\cdot|s_1)$

To interpret, since P1 learns her own type, the action that she plays is now **type-dependent**. P2 plays only after observing the action of P1, so that his strategy is a mapping from S_1 to S_2. Whereas SPE says P1 must have no profitable unilateral deviation in any subgame, PBE requires no **type** of P1 has such deviations. This is because "learning that her type is θ_H" and "learning that her type is θ_L" can be interpreted as two subgames of the larger game for P1.

The PBE further requires that P2 has a family of posterior beliefs, $\{\pi(\cdot|s_1) : s_1 \in S_1\}$. This is **a collection of distributions** on Θ, one for each $s_1 \in S_1$. Said another way, $\pi(\theta_L|s_1) + \pi(\theta_H|s_1) = 1$ for any $s_1 \in S_1$. The distribution $\pi(\cdot|s_1)$ represents the new belief that P2 holds about P1's type after observing P1 taking the action s_1. The "**Bayesian**" part of the equilibrium concept comes from the requirement that, for any $s_1 \in \text{range}(\sigma_1)$, PBE requires:

$$\boxed{\pi(\theta|s_1) = \frac{\mathbb{P}[\text{P1 is type } \theta \text{ and plays } s_1]}{\mathbb{P}[\text{P1 plays } s_1]}} \tag{39}$$

That is, we are imposing the restriction that P2 acts like a Bayesian whenever possible and derives posterior belief $\pi(\theta|s_1)$ using the Bayesian updating formula. The emphasis on "**whenever possible**" comes from the fact that not every $s_1 \in S_1$ needs to be played by P1. For instance, if there are more available actions to P1 than there are types of P1 ($|S_1| > |\Theta|$) and every type plays a pure action, then there will always be some actions that never get played. Even when there are as many actions for P1 as types of P1, it is possible that all types of P1 play the same action, leaving some members of S_1unplayed. In any event, $\pi(\cdot|s_1)$ cannot be derived from Bayes rule when s_1 is not played by P1, since we would be dividing by 0. PBE calls those $\pi(\cdot|s_1)$ for s_1 that are never played by P1 "off-equilibrium beliefs" and places **no restrictions** on what they can be. This is not to say they are unimportant. Indeed, much like how off-equilibrium parts of a complete contingent strategy can play a key part in sustaining a SPE, so too the off-equilibrium beliefs in a PBE can play a key part in preventing deviations by P1. In the college major choice, for example, depending on parameters there can be a PBE where every type of P1 chooses to be a Physics major. This major choice is sustained by P2's belief that anyone who majors in Economics has low ability, i.e $\pi(\theta_L|E) = 1$, $\pi(\theta_H|E) = 0$. But this is in fact an off-equilibrium belief, since with no probability does any type of P1 major in Economics.

3.4 How to solve for PBEs. Make sure you look for two kinds of PBEs, namely:

(1) Look for a **separating equilibrium**, where two types of P1 play different actions.

(2) Look for a **pooling equilibrium**, where two types of P1 play the same action.

Let's take a concrete example. Consider a plaintiff (P1) and a defendant (P2) in a **civil lawsuit**. Plaintiff knows whether she has a strong case (θ_H) or weak case (θ_L), but the defendant does not. Defendant has prior belief that $\pi(\theta_H) = \frac{1}{3}$, $\pi(\theta_L) = \frac{2}{3}$. The plaintiff can ask for a low settlement or a high settlement ($S_1 = \{1, 2\}$). The defendant accepts or refuses, $S_2 = \{y, n\}$. If the defendant accepts a settlement offer of s_1, the two players **settle out-of-court** with payoffs $(s_1, -s_1)$. If defendant refuses, the case goes to trial. If the case is strong ($\theta = \theta_H$), plaintiff wins for sure and the payoffs are $(3, -4)$. If the case is weak ($\theta = \theta_L$), the plaintiff loses for sure and the payoffs are $(-1, 0)$.

Separating equilibrium: Remember, if $|\Theta| = 2$ and $|S_1| = 2$, then in any separating PBE P2's posterior belief will **exactly reveal** P1's true type. Typically, there are **multiple** potential separating equilibria, depending on what action each type of P1 plays. Be sure to **check all of them**.

- **Separating equilibrium, version 1**. Can there be a PBE where $\sigma_1(\theta_H) = 2$, $\sigma_1(\theta_L) = 1$? If so, in any such PBE we must have $\pi(\theta_H|2) = 1$, $\pi(\theta_L|1) = 1$, $\sigma_2(2) = y$, $\sigma_2(1) = n$. But this means type θ_L gets -1 in PBE and has a profitable unilateral deviation by playing $\hat{\sigma}_1(\theta_L) = 2$

instead. Asking for the high settlement makes P2 think P1 has a strong case, so that P2 will settle and the weak P1 will get 2 instead of -1. Therefore no such PBE exists.

- **Separating equilibrium, version 2**. Can there be a PBE where $\sigma_1(\theta_H) = 1$, $\sigma_1(\theta_L) = 2$? (It seems very counterintuitive that the plaintiff with a strong case asks for a lower settlement than the plaintiff with a weak case, but this is still a candidate for a separating PBE in the formal setup, so we cannot ignore it.) If so, in any such PBE we must have $\pi(\theta_L|2) = 1$, $\pi(\theta_H|1) = 1$, $\sigma_2(2) = n$, $\sigma_2(1) = y$. But this means type θ_H gets 1 in PBE and has a profitable unilateral deviation by playing $\hat{\sigma}_1(\theta_H) = 2$ instead. Asking for the high settlement makes P2 think P1 has a weak case, so that P2 will let the trial go to court. But this is great when P1 has a strong case, giving her a payoff of 3 instead of 1. Therefore no such PBE exists.

Pooling equilibrium: In a pooling equilibrium all types of P1 play the same action. When this "pooled" action s_1^* is observed, P2's posterior belief is the **same as the prior**, $\pi(\theta|s_1^*) = \pi(\theta)$, since the action carries no additional information about P1's type. When any other action is observed (i.e. an off-equilibrium action is observed), PBE allows P2's belief to be **arbitrary** as to avoid division by 0. Every member of S_1 could serve as a pooled action, so we need to **check for all of them systematically**.

- **Pooling on low settlement**. Can there be a PBE where $\sigma_1(\theta_H) = \sigma_1(\theta_L) = 1$? If so, in any such PBE we must have $\pi(\theta_H|1) = \frac{1}{3}$. Under this posterior belief, P2's expected payoff to $s_2 = n$ is $\frac{1}{3}(-4) + \frac{2}{3}(0) = -\frac{4}{3}$, while playing $s_2 = y$ always yields -1. Therefore in any such PBE we must have $\sigma_2(1) = y$. But then the θ_H type of P1 has a profitable unilateral deviation of $\hat{\sigma}_1(\theta_H) = 2$, regardless of what $\sigma_2(2)$ is! If $\sigma_2(2) = y$, that is P2 always accepts the high settlement, then type θ_H P1's deviation gives her a payoff of 2 rather than 1. If $\sigma_2(2) = n$, that is P2 always refuses the high settlement, then this is even better for the type θ_H P1 as she will get a payoff of 3 when the case goes to court. Therefore no such PBE exists.

- **Pooling on high settlement.** Can there be a PBE where $\sigma_1(\theta_H) = \sigma_1(\theta_L) = 2$? If so, in any such PBE we must have $\pi(\theta_H|2) = \frac{1}{3}$. Under this posterior belief, P2's expected payoff to $s_2 = n$ is $\frac{1}{3}(-4)+\frac{2}{3}(0) = -\frac{4}{3}$, while playing $s_2 = y$ always yields -2. Therefore in any such PBE we must have $\sigma_2(2) = n$. In order to prevent a deviation by type θ_L, we must ensure $\sigma_2(1) = n$ as well. Else, if P2 accepts the low settlement offer, θ_L would have a profitable deviation where offering the low settlement instead of following the pooling action of high settlement yields her a payoff of 1 instead of -1. But whether $\sigma_2(1) = n$ is optimal for P2 **depends on the belief**, $\pi(\theta_H|1)$. Fortunately, this is an **off-equilibrium belief** and PBE allows such beliefs to be arbitrary. Suppose $\pi(\theta_H|1) = \lambda \in [0, 1]$. Then P2's expected payoff to playing $\sigma_2(1) = n$ is $\lambda(-4) + (1-\lambda)(0) = -4\lambda$, while deviating to $\sigma_2(1) = y$ yields -1 for sure. Therefore to ensure P2 does not have a profitable deviation at the "subgame" of seeing a low settlement offer, we need $\lambda \leq \frac{1}{4}$. If P2's off-equilibrium belief is that P1 has a strong case with probability less than $\frac{1}{4}$ upon seeing a low-settlement offer, then it is optimal for P2 to reject such low-settlement offers and θ_L will not have a profitable deviation. At the same time, since P2 rejects all offers, the strong type θ_H of P1 does not have a profitable deviation either, since whichever offer she makes the case will go to court. In summary, there is a (family of) pooling equilibrium where $\sigma_1(\theta_H) = \sigma_1(\theta_L) = 2$, $\sigma_2(1) = \sigma_2(2) = n$, $\pi(\theta_H|2) = \frac{1}{3}$, $\pi(\theta_H|1) = \lambda$ where $\lambda \in [0, \frac{1}{4}]$. Crucially, it is the judicious choice of **off-equilibrium belief** $\pi(\cdot|1)$ that sustains the off-equilibrium action of $\sigma_2(1) = n$, which in turn sustains the pooling equilibrium.

Sections and office hours are canceled for all TFs this week. Happy (American) Thanksgiving!

TF: Kevin He (he02@fas.harvard.edu)

For those who miss Ec1011a, here are some fun exercises. (These are some of the problems that didn't make the cut to this year's midterm / final.)

1. At the start of a game of Ultimate, the two team captains each flips a Frisbee disc in the air to determine which team will "kick off". Team 1 wins the flip if both discs land on the same side. Team 2 wins the flip if two discs land on different sides. Unlike fair coins, Frisbee discs are asymmetric. In particular, assume every disc has some q chance of landing "heads", where $0 \le q \le 1$, $q \neq \frac{1}{2}$. By relating this situation to a game of matching pennies (or otherwise), explain why it is strictly more likely that Team 1 wins the flip.

2. Write $\mathbb{Z}\backslash\{0\}$ for the set of nonzero integers. That is, $\mathbb{Z}\backslash\{0\} = \{..., -3, -2, -1, 1, 2, 3, ...\}$. Let $\succsim$ be a preference relation on $\mathbb{Z}\backslash\{0\}$, where $a \succ b$ if and only if either (i) $|a| > |b|$, or (ii) $|a| = |b|$ but $a > 0$ while $b < 0$. Also, $a \sim b$ if and only if $a = b$. (Here, $|a|$ refers to the absolute value of a.) Construct a utility function $u : \mathbb{Z}\backslash\{0\} \to \mathbb{R}$ to represent $\succsim$.

3. Write $\mathbb{R}\backslash\{0\}$ for the set of nonzero real numbers. Let $\succsim$ be a preference relation on $\mathbb{R}\backslash\{0\}$, where $a \succ b$ if and only if either (i) $|a| > |b|$, or (ii) $|a| = |b|$ but $a > 0$ while $b < 0$. Also, $a \sim b$ if and only if $a = b$. (Here, $|a|$ refers to the absolute value of a.) Is it possible to construct a utility function $u : \mathbb{R}\backslash\{0\} \to \mathbb{R}$ to represent $\succsim$? If so, exhibit one such construction. (Discontinuous utility functions are allowed.) If not, explain why the task is impossible.

4. Most Broadway musicals have two (or fewer) acts. Give an example of a musical with three or more acts.

5. $S = \{s_1, ..., s_n\}$ is a finite state space equipped with a probability distribution $(q_i)_{i=1}^n$, where $\sum_{i=1}^n q_i = 1$ and $q_i \ge 0$ for each i. Consider an expected-utility individual whose utility only depends on his consumption of vests tomorrow. The only way to get vests tomorrow is by purchasing "futures contracts" on vests today. More precisely, there are a total of n different contracts available, where one share of the i^{th} contract delivers one vest in the event that the state tomorrow is s_i, 0 vests otherwise. The individual can buy any number of shares of each contract, however he cannot hold a "negative number of shares" in any contract (i.e. no short-selling). When the prices of the n futures contracts are given by a price vector $\mathbf{p} \in \mathbb{R}^n_{++}$ and the individual has wealth Y, the relevant optimization problem is: $\max_{\mathbf{x}\in\mathbb{R}^n_+} \sum_{i=1}^n q_i \cdot v(x_i)$ s.t. $\mathbf{p}\cdot\mathbf{x} \le Y$, where x_i is the number of shares of the i^{th} futures contract that he buys and v is his increasing, strictly concave utility function over vests. An econometrician has collected a dataset $(\mathbf{x}^{(k)}, \mathbf{p}^{(k)})_{k=1}^K$ with K observations. In the k^{th} observation, the individual demanded the bundle of contracts $\mathbf{x}^{(k)}$ when faced with prices $\mathbf{p}^{(k)}$. Under the assumption that the consumer spends all of his wealth in every observation, we must have for every $1 \le k \le K$,

$$\mathbf{x}^{(k)} = \arg\max_{\mathbf{x}\in\mathbb{R}^n_+} \sum_{i=1}^n q_i \cdot v(x_i) \text{ s.t. } \mathbf{p}^{(k)} \cdot \mathbf{x} \le \mathbf{p}^{(k)} \cdot \mathbf{x}^{(k)} \tag{40}$$

Now, call a sequence of contract pairs $(x_{i_j}^{(k_j)}, x_{i'_j}^{(k'_j)})_{j=1}^J$ "balanced" if $(i_j)_{j=1}^J$ is a permutation of $(i'_j)_{j=1}^J$ and $(k_j)_{j=1}^J$ is a permutation of $(k'_j)_{j=1}^J$. For instance,

$$\Big((x_1^{(3)}, x_9^{(4)}), (x_2^{(1)}, x_5^{(3)}), (x_5^{(4)}, x_1^{(1)})(x_9^{(1)}, x_2^{(1)})\Big)$$

is balanced. For a balanced sequence that also satisfies $x_{i_j}^{(k_j)} > x_{i'_j}^{(k'_j)}$ for all j, show that the product of prices satisfies

$$\prod_{j=1}^{J} \frac{p_{i_j}^{(k_j)}}{p_{i'_j}^{(k'_j)}} \leq 1$$

Suggested approach: use Lagrange multiplier to optimize (40) and express $p_i^{(k)}$ in terms of $v'(x_i^{(k)})$, q_i, and the Lagrange multiplier $\lambda^{(k)}$. (Remember, the value of the Lagrange multiplier is potentially different across the K optimization problems...)

(1) Math tools; **(2)** Firm's problem; **(3)** Preferences and utility; **(4)** Consumer's problem; **(5)** Time; **(6)** Risk; **(7)** Markets and welfare; **(8)** normal-form games; **(9)** Other games: extensive form, signaling

TF: Kevin He (he02@fas.harvard.edu)

1 Mathematical tools

Definition 1. Call a function **smooth** if it has continuous derivatives of order n for every $n = 1, 2, 3, \ldots$.

Definition 2. For $f : \mathbb{R}^n \to \mathbb{R}$, say f is **concave** if

$$f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) \geq \alpha f(\mathbf{x}) + (1-\alpha)f(\mathbf{y})$$

for all $\alpha \in (0,1)$, $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.
Conversely, call f **convex** if

$$f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) \leq \alpha f(\mathbf{x}) + (1-\alpha)f(\mathbf{y})$$

for all $\alpha \in (0,1)$, $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

Theorem 3. *(Young's theorem) If $f(x,y)$ has continuous second derivatives, then $f_{xy} = f_{yx}$.*

Proposition 4. *(Chain rule) Suppose $f : \mathbb{R}^n \to \mathbb{R}$ and $x_k : \mathbb{R} \to \mathbb{R}$ for $1 \leq k \leq n$. Then*

$$\begin{aligned}&\frac{d}{ds}f(x_1(s), x_2(s), \ldots, x_n(s))\\ =\ &\frac{\partial f}{\partial x_1}\cdot\frac{dx_1}{ds} + \frac{\partial f}{\partial x_2}\cdot\frac{dx_2}{ds} + \ldots + \frac{\partial f}{\partial x_n}\cdot\frac{dx_n}{ds}\end{aligned}$$

Proposition 5. *(FOC) Suppose $f : \mathbb{R} \to \mathbb{R}$ is differentiable. If $x^* \in \mathbb{R}$ is a local maximizer or a local minimizer for f, then $f'(x^*) = 0$.*

Proposition 6. *(SOC) Suppose $f : \mathbb{R} \to \mathbb{R}$ is twice differentiable and $x^* \in \mathbb{R}$. If $f'(x^*) = 0$ and $f''(x^*) < 0$, then x^* is a local maximizer for f. If $f'(x^*) = 0$ and $f''(x^*) > 0$, then x^* is a local minimizer for f.*

Remark 7. FOC and SOC characterize local optima, not global optima.

Remark 8. If the optimization problem is not over all of $\mathbb{R}$, there could exist corner solution local optima that fail FOC. For instance, if the maximization problem is over $[0, \infty)$, 0 could be a global maximizer and yet $f'(0) \neq 0$.

Proposition 9. *(Multivariate FOC) Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable. If $x^* \in \mathbb{R}^n$ is a local maximizer or local minimizer for f, then $f_k(x^*) = 0$ for each $1 \leq k \leq n$, where f_k refers to the partial derivative of f with respect to its k^{th} argument.*

Proposition 10. *If $f : \mathbb{R}^n \to \mathbb{R}$ is concave, then any $x^* \in \mathbb{R}^n$ that satisfies the multivariate FOC is a local maximizer of f.*

Proposition 11. *(SOC for two variables) Suppose $f : \mathbb{R}^2 \to \mathbb{R}$, $x^* \in \mathbb{R}^2$. If f satisfies multivariate FOC at x^* and $f_{11} < 0$, $f_{22} < 0$, $f_{11}f_{22} - (f_{12})^2 > 0$ (where all partials are evaluated at x^*), then x^* is a local maximizer for f.*

Definition 12. Given a constrained maximization problem $\max_{x,y} f(x,y)$ subject to $g(x,y) = c$, the associated **Lagrangian** is $\mathcal{L}(x,y,\lambda) := f(x,y) + \lambda(c - g(x,y))$. Under regularity conditions, if (x^*, y^*, λ^*) satisfies the FOC of the Lagrangian, then (x^*, y^*) is a constrained maximizer in the original problem.

Proposition 13. *(unconstrained envelope theorem) In an unconstrained optimization problem $\max_{\mathbf{x}\in\mathbb{R}^n} V(\mathbf{x}; \mathbf{z})$ or $\min_{\mathbf{x}\in\mathbb{R}^n} V(\mathbf{x}; \mathbf{z})$, we have $\frac{dV^*}{dz_k}(\mathbf{z}) = V_{z_k}(\mathbf{x}^*(\mathbf{z}); \mathbf{z})$.*

Proposition 14. *(constrained envelope theorem) In a constrained optimization problem $\max_{\mathbf{x}\in\mathbb{R}^n} V(\mathbf{x}; \mathbf{z})$ s.t. $g(\mathbf{x}; \mathbf{z}) = 0$ or $\min_{\mathbf{x}\in\mathbb{R}^n} V(\mathbf{x}; \mathbf{z})$ s.t. $g(\mathbf{x}; \mathbf{z}) = 0$, we have*

$$\frac{\partial V^*}{\partial z_k}(\mathbf{z}) = \frac{\partial \mathcal{L}}{\partial z_k}(\mathbf{x}^*(\mathbf{z}), \lambda^*(\mathbf{z}); \mathbf{z})$$

where $\mathcal{L}(\mathbf{x}, \lambda; \mathbf{z})$ is the Lagrangian associated with the optimization problem.

2 Firm's problem

Definition 15. The firm has two problems.

- **Profit maximization**:

  $$\max_{K,L} \{p \cdot f(K,L) - rK - wL\}$$

  Its argmax is $(K^*_\pi(p,r,w), L^*_\pi(p,r,w))$, the **profit-maximizing factor demand**. Its value function is $\pi(p,r,w)$, the **profit function**.

- **Cost minimization**:

  $$\min_{K,L} rK + wL \text{ s.t. } f(K,L) = q$$

  Its argmin is $(K^*_C(q,r,w), L^*_C(q,r,w))$, the **cost-minimizing factor demand**. Its value function is $C(q,r,w)$, the **cost function**.

Proposition 16. *K^*_π and L^*_π are homogeneous of degree 0 in (p,r,w). K^*_C and L^*_C are homogeneous of degree 0 in (r,w).*

Proposition 17. *π is homogeneous of degree 1 in (p,r,w). C is homogeneous of degree 1 in (r,w).*

Proposition 18. *(Hotelling's lemma) $\frac{\partial \pi}{\partial p} = f(K^*_\pi, L^*_\pi)$.*

Proposition 19. *$\frac{\partial \pi}{\partial r} = -K^*_\pi$, $\frac{\partial \pi}{\partial w} = -L^*_\pi$.*

Proposition 20. *$\pi(p,r,w)$ is convex in (p,r,w).*

Proposition 21. *(Shephard's lemma) $\frac{\partial C}{\partial r} = K^*_C$, $\frac{\partial C}{\partial w} = L^*_C$.*

Proposition 22. *$C(q,r,w)$ is concave in (r,w).*

Definition 23. Some common functional forms for production function:

- **Separable**. $f(x_1, x_2, ..., x_n) = g_1(x_1) + g_2(x_2) + ... + g_n(x_n)$.
- **Leontief**. $f(x,y) = \min\left\{\frac{x}{a_x}, \frac{y}{a_y}\right\}$ for $a_x, a_y > 0$.
- **Linear**. $f(x,y) = a_x x + a_y y$ with $a_x, a_y > 0$.

Definition 24. A production function exhibits **increasing returns to scale** if $f(\lambda \mathbf{x}) > \lambda f(\mathbf{x})$ for $\lambda > 1$, **decreasing returns to scale** if $f(\lambda \mathbf{x}) < \lambda f(\mathbf{x})$ for $\lambda > 1$, and **constant returns to scale** if $f(\lambda \mathbf{x}) = \lambda f(\mathbf{x})$ for $\lambda > 0$.

Proposition 25. *Suppose $f(\mathbf{0}) = 0$. If f is strictly concave, then it exhibits decreasing returns to scale. If f is strictly convex, then it exhibits increasing returns to scale.*

3 Preferences and utility

Definition 26. A **preference relation** $\succsim$ on a set X is a binary relation on X. That is, there is some subset $B \subseteq X \times X$ and $x_1 \succsim x_2$ if and only if $(x_1, x_2) \in B$.

Definition 27. Call $\succsim$ **complete** if for every pair $x, y \in X$, either $x \succsim y$ or $y \succsim x$ (or both).

Definition 28. Call $\succsim$ **transitive** if for $x, y, z \in X$, $[x \succsim y$ and $y \succsim z]$ implies $x \succsim z$.

Definition 29. Call $\succsim$ **rational** if it is complete and transitive.

Example 30. Let $X = \{p, q, r, s\}$. $\succsim_1$ is incomplete but transitive. $\succsim_2$ is complete but intransitive. $\succsim_3$ is complete and transitive.

$\succsim_1$	p	q	r	s
p	•	•	•	•
q		•		
r			•	
s				•

$\succsim_2$	p	q	r	s
p	•	•		•
q		•	•	•
r	•		•	
s		•	•	•

$\succsim_3$	p	q	r	s
p	•	•	•	•
q		•	•	•
r			•	
s		•	•	•

Joke 31. (Love is irrational) Let X be the finite set of all humans in the world. For $a, b \in X$, define $a \succsim b$ if and only if a loves b. Then $\succsim$ is

not a rational preference – in fact it fails on both counts. First, $\succsim$ is not complete, because there exist many pairs of people (a, b) where neither a loves b nor b loves a. Transitivity fails even more spectacularly.

Definition 32. Say a utility function $u : X \to \mathbb{R}$ **represents** a preference relation $\succsim$ on X if for any $x, y \in X$, $u(x) \geq u(y)$ if and only if $x \succsim y$.

Proposition 33. *Suppose X is either finite or countably infinite. A preference relation $\succsim$ on X is rational if and only if it admits a utility representation $u : X \to \mathbb{R}$.*

Fact 34. *The lexicographic preference on $\mathbb{R} \times \mathbb{R}$ is rational, yet it does not admit **any** utility representation.*

Definition 35. Suppose $X \subseteq \mathbb{R}^n$. A preference $\succsim$ on X is called **continuous** if for every sequence $a_n \to a$ and fixed alternative b, where $a_n, a, b \in X$, we have (i) $a_n \succsim b$ for every n implies $a \succsim b$, and (ii) $b \succsim a_n$ for every n implies $b \succsim a$.

Theorem 36. *(Debreu's theorem) Suppose $X \subseteq \mathbb{R}^n$. A preference relation $\succsim$ on X is rational and continuous if and only if it admits a continuous utility representation $u : X \to \mathbb{R}$.*

Proposition 37. *Suppose utility function $u : X \to \mathbb{R}$ represents a preference $\succsim$ on X. If $h : \mathbb{R} \to \mathbb{R}$ is strictly monotonic on the range of u, then $h \circ u : X \to \mathbb{R}$ also represents $\succsim$.*

4 Consumer's problem

Definition 38. The consumer has two problems.

- **UMP**: $\max_{\mathbf{x}} u(\mathbf{x})$ s.t. $\mathbf{p} \cdot \mathbf{x} \leq Y$. Its argmax is $\mathbf{x}(\mathbf{p}, Y)$, the **Marshallian demand**. Its value function is $v(\mathbf{p}, Y)$, the **indirect utility**.

- **EMP**: $\min_{\mathbf{x}} \mathbf{p} \cdot \mathbf{x}$ s.t. $u(\mathbf{x}) \geq \bar{u}$. Its argmin is $\mathbf{h}(\mathbf{p}, \bar{u})$, the **Hicksian demand**. Its value function is $e(\mathbf{p}, \bar{u})$, the **expenditure function.**

Proposition 39. *Marshallian demand is homogeneous of degree 0 in $(\mathbf{p}, Y)$. Hicksian demand is homogeneous of degree 0 in $\mathbf{p}$.*

Proposition 40. *Indirect utility is homogeneous of degree 0 in $(\mathbf{p}, Y)$. The expenditure function is homogeneous of degree 1 in $\mathbf{p}$.*

Proposition 41. *(Roy's identity) For any $1 \leq i \leq n$,*

$$-\frac{v_{p_i}(\mathbf{p}, Y)}{v_Y(\mathbf{p}, Y)} = x_i(\mathbf{p}, Y)$$

Proposition 42. *$v(\mathbf{p}, Y)$ is quasi-convex in $(\mathbf{p}, Y)$.*

Proposition 43. *$v(\mathbf{p}, Y)$ is weakly increasing in Y.*

Proposition 44. *(Shephard's lemma) For any $1 \leq i \leq n$,*

$$e_{p_i}(\mathbf{p}, \bar{u}) = h_i(\mathbf{p}, \bar{u})$$

Proposition 45. *$e(\mathbf{p}, \bar{u})$ is concave in $\mathbf{p}$.*

Proposition 46. *$e(\mathbf{p}, \bar{u})$ is weakly increasing in $\bar{u}$.*

Proposition 47. *(Duality) $\mathbf{h}(\mathbf{p}, v(\mathbf{p}, Y)) = \mathbf{x}(\mathbf{p}, Y)$, $\mathbf{x}(\mathbf{p}, e(\mathbf{p}, \bar{u})) = \mathbf{h}(\mathbf{p}, \bar{u})$.*

Proposition 48. *(Slutsky equation) For any $1 \leq i, j \leq n$,*

$$\frac{\partial x_i}{\partial p_j} = \frac{\partial h_i}{\partial p_j} - \frac{\partial x_i}{\partial Y} \cdot x_j$$

Definition 49. $\eta_i := \frac{p_i x_i}{Y}$ is the **budget share** of good i. $\varepsilon^i_{p_j} := \frac{\partial x_i}{\partial p_j} \frac{p_j}{x_i}$ is the **elasticity of demand** for good i with respect to the price of good j. $\varepsilon^i_Y := \frac{\partial x_i}{\partial Y} \frac{Y}{x_i}$ is the **elasticity of demand** for good i with respect to income Y.

Proposition 50. $\sum_{i=1}^n \eta_i \varepsilon^i_Y = 1$.

Proposition 51. *$\eta_j = -\sum_{i=1}^n \eta_i \varepsilon^i_{p_j}$ for every $1 \leq j \leq n$.*

Definition 52. An **inferior good** has negative demand elasticity with respect to income. A **Giffen good** has positive demand elasticity with respect to its own price.

Proposition 53. *Every Giffen good is also an inferior good.*

5 Time and preferences

Definition 54. Suppose an individual lives for T periods and has **exponential discounting**. At time $1 \leq s \leq T$, she evaluates a consumption plan $(c_1, c_2, ..., c_T)$ as:

$$U_s(c_1, c_2, ..., c_T) = \sum_{t=s}^{T} \delta^{t-s} \cdot u(c_t)$$

where $\delta \in (0, 1)$, $u : \mathbb{R} \to \mathbb{R}$ is concave.

Fact 55. *We interpret δ in exponential discounting either as psychological impatience or the complement to chance of death.*

Proposition 56. *The **Euler equation***

$$u'(c_t^*) = \delta^s (1 + r)^s u'(c_{t+s}^*)$$

characterizes optimal consumption in an intertemporal optimization problem.

Definition 57. Present bias refers to an individual discounting more heavily between "today" and "tomorrow" than between "tomorrow" and "the day after tomorrow".

Definition 58. Time inconsistency refers to an individual making a set of plans about future consumption but failing to follow her own plans if given the chance to change her mind later.

Definition 59. Suppose an individual lives for T periods and has **quasi-hyperbolic discounting**. At time $1 \leq s \leq T$, she evaluates a consumption plan $(c_1, c_2, ..., c_T)$ as:

$$U_s(c_1, c_2, ..., c_T) = u(c_s) + \beta \sum_{t=s+1}^{T} \delta^{t-s} u(c_t)$$

where $\delta \in (0, 1)$ and $\beta \in (0, 1]$, $u : \mathbb{R} \to \mathbb{R}$ is concave.

Fact 60. *Exponential discounting does not exhibit either present bias or time inconsistency. Quasi-hyperbolic discounting exhibits both present bias and time inconsistency.*

6 Risk and preferences

Definition 61. If X is a lottery, then an individual with **expected utility** evaluates the lottery as:

$$U(X) = \mathbb{E}\left[u(X)\right] = \sum_x \Pr(X = x) \cdot u(x)$$

where $u : \mathbb{R} \to \mathbb{R}$ is called a **Bernoulli utility function**.

Fact 62. *Humans are typically risk averse over large stake gains. This is modeled through a concave Bernoulli utility function.*

Theorem 63. *(Jensen's inequality) Suppose X is a real-valued random variable, $\varphi : \mathbb{R} \to \mathbb{R}$ is convex and $\phi : \mathbb{R} \to \mathbb{R}$ is concave. Then $\mathbb{E}\left[\varphi(X)\right] \geq \varphi\left(\mathbb{E}\left[X\right]\right)$ and $\mathbb{E}\left[\phi(X)\right] \leq \phi\left(\mathbb{E}\left[X\right]\right)$.*

Proposition 64. *An expected utility consumer with concave Bernoulli utility will turn down every fair bet and fully insure against risk when the insurance premium is actuarily fair.*

Definition 65. For lottery X and initial wealth level y, the **certainty equivalence** of X at y is written as $c(X, y)$ and defined implicitly by:

$$u(y + c(X, y)) = \mathbb{E}[u(y + X)]$$

Proposition 66. *An expected utility consumer with concave Bernoulli utility has $c(X, y) \leq \mathbb{E}[X]$.*

Definition 67. Suppose $u(\cdot)$ is a Bernoulli utility function. Then its **coefficient of absolute risk aversion** at y is defined as $-\frac{u''(y)}{u'(y)}$.

Definition 68. Suppose $u(\cdot)$ is a Bernoulli utility function. Then its **coefficient of relative risk aversion** at y is defined as $-y\frac{u''(y)}{u'(y)}$.

Definition 69. The Bernoulli utility function $u(y) = -\exp(-\alpha y)$ for $\alpha > 0$ is called constant absolute risk aversion (**CARA**).

Proposition 70. *If an expected-utility maximizing individual has CARA Bernoulli utility and X is a lottery, then she [accepts / turns down] X at wealth level y if and only if she also [accepts / turns down] X at every other wealth level y'.*

Fact 71. *The Allais paradox shows a violation of expected utility. It can be resolved using probability weighting functions.*

Fact 72. *The Ellsberg paradox shows a violation of expected utility. It can be resolved using ambiguity-averse models of decision-making, such as maxmin expected utility (MEU).*

7 Markets and welfare

Definition 73. Consider an individual with n consumption goods and wealth Y. Suppose price vector changes from $\mathbf{p}^{(0)} = (p_1^{(0)}, p_2^{(0)}, ..., p_n^{(0)})$ to $\mathbf{p}^{(1)} = (p_1^{(1)}, p_2^{(0)}, ..., p_n^{(0)})$ and abbreviate $u^{(0)} = v\left(\mathbf{p}^{(0)}, Y\right)$, $u^{(1)} = v\left(\mathbf{p}^{(1)}, Y\right)$. **Equivalent variation** (EV) is defined as

$$EV := e\left(\mathbf{p}^{(0)}, u^{(1)}\right) - Y$$

Compensating variation (CV) is defined as:

$$CV := Y - e\left(\mathbf{p}^{(1)}, u^{(0)}\right)$$

Consumer surplus (CS) is defined as:

$$CS := \int_{p_1^{(1)}}^{p_1^{(0)}} x_1(\tilde{p}_1, p_2^{(0)}, ..., p_n^{(0)}, Y) d\tilde{p}_1$$

Proposition 74.

$$EV = \int_{p_1^{(1)}}^{p_1^{(0)}} h_1(\tilde{p}_1, p_2^{(0)}, ..., p_n^{(0)}, u^{(1)}) d\tilde{p}_1$$

Proposition 75.

$$CV = \int_{p_1^{(1)}}^{p_1^{(0)}} h_1(\tilde{p}_1, p_2^{(0)}, ..., p_n^{(0)}, u^{(0)}) d\tilde{p}_1$$

Definition 76. An **exchange economy** has:

- L commodities ($\ell = 1, 2, ..., L$), so that space of consumption is $\mathbb{R}_+^L$
- I agents ($i = 1, 2, ..., I$), each with a utility function $u_i : \mathbb{R}_+^L \to \mathbb{R}$ and an initial endowment $\omega_i \in \mathbb{R}_+^L$

Definition 77. An **allocation** in an exchange economy is a list of vectors $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_I)$ with each $\mathbf{x}_i \in \mathbb{R}_+^L$. Call an allocation **feasible** if $\sum_{i=1}^I \mathbf{x}_i = \sum_{i=1}^I \omega_i$.

Definition 78. Allocation **y Pareto dominates** allocation $\mathbf{x}$ if $u_i(\mathbf{y}_i) \geq u_i(\mathbf{x}_i)$ for every $1 \leq i \leq I$ and $u_{i^*}(\mathbf{y}_{i^*}) > u_{i^*}(\mathbf{x}_{i^*})$ for some $1 \leq i^* \leq I$. Call a feasible allocation **Pareto efficient** if it is not Pareto dominated by any other feasible allocation.

Definition 79. A **Walrasian equilibrium** is an allocation $\mathbf{x}^* = (\mathbf{x}_1^*, ..., \mathbf{x}_I^*)$ and a price vector $\mathbf{p}^* \in \mathbb{R}_{++}^L$ such that:

(i) [individual maximization]

$$\mathbf{x}_i^* \in \arg\max_{\mathbf{x}_i \in \mathbb{R}_+^L} u_i(\mathbf{x}_i) \text{ s.t. } \mathbf{p}^* \cdot \mathbf{x}_i \leq \mathbf{p}^* \cdot \omega_i$$

for every $1 \leq i \leq I$.

(ii) [market clearing] $\sum_{i=1}^I \mathbf{x}_i^* = \sum_{i=1}^I \omega_i$

Definition 80. Call a utility function u_i **strongly monotone** if whenever $\mathbf{x}_i, \hat{\mathbf{x}}_i$ are such that $\hat{x}_{\ell i} > x_{\ell i}$ for every $1 \leq \ell \leq L$, we have $u_i(\hat{\mathbf{x}}_i) > u_i(\mathbf{x}_i)$.

Theorem 81. *(First welfare theorem) If u_i is strongly monotone for every $1 \leq i \leq n$, then any Walrasian equilibrium allocation is Pareto efficient.*

Remark 82. One implicit assumption of the first welfare theorem (and of the exchange economy model more generally) is the absence of externality. If there is externality, the first welfare theorem might not hold.

Theorem 83. *(second welfare theorem) Suppose for every agent i, the utility function u_i is continuous, strongly monotone, and strictly concave. Suppose also $\mathbf{x}^* = (\mathbf{x}_1^*, ..., \mathbf{x}_I^*)$ is a Pareto efficient allocation such that $\mathbf{x}_i^* \in \mathbb{R}_{++}^L$ for every i. Then there exists a set of new endowments $\{\omega_i'\}_{1 \leq i \leq I}$ with the property $\sum_{i=1}^I \omega_i' = \sum_{i=1}^I \omega_i$ such that if agent i is re-assigned the endowment ω_i' for every $1 \leq i \leq I$, then there exists a Walrasian equilibrium with allocation $\mathbf{x}^*$ under this new set of endowments.*

Proposition 84. *(individual Walras' law) If utility function u_i is strongly monotone, then for any price vector* $\mathbf{p}$ *and income* Y*, Marshallian demand satisfies:*

$$\mathbf{p} \cdot \mathbf{x}_i(\mathbf{p}, Y) = Y$$

Proposition 85. *(market Walras' law) Consider an exchange economy with L commodities and I individuals where each individual has a strongly monotone utility function. If price vector* $\mathbf{p} \in \mathbb{R}^L_{++}$ *is such that* $\sum_{i=1}^I x_{\ell i}(\mathbf{p}, \omega_i \cdot \mathbf{p}) = \sum_{i=1}^I \omega_{\ell i}$ *for $L-1$ of the commodities $\ell \in \{1, 2, ..., L\}$, then in fact* $\sum_{i=1}^I x_{\ell i}(\mathbf{p}, \omega_i \cdot \mathbf{p}) = \sum_{i=1}^I \omega_{\ell i}$ *for every $\ell \in \{1, 2, ..., L\}$.*

Definition 86. A **Pigouvian tax** on an action is equal to the costs of the action that are not internalized by the agent taking the action.

"Theorem" 87. (Coase theorem) If there are no transaction costs then bargaining leads to a Pareto efficient outcome regardless of the initial endowments. If there are no transaction costs and no income effects, then bargaining leads to the surplus-maximizing outcome and initial endowments only influence the division of surplus between the parties.

Definition 88. The **monopolist's problem** is

$$\max_{q \geq 0} \{p(q) \cdot q - C(q)\}$$

where $p(\cdot)$ is the market's inverse demand function and $C(\cdot)$ is the monopolist's cost function.

Proposition 89. *If p^* is the optimal monopoly price and $q(p)$ is the demand function, then* $\left| q'(p^*) \cdot \frac{p^*}{q(p^*)} \right| > 1$.

Definition 90. The **monopsonist's problem** is:

$$\max_{w \geq 0} \{pf(L(w)) - wL(w)\}$$

where $f(\cdot)$ is the monopsonist's production function and $L(\cdot)$ is the market's labor supply curve.

8 Normal-form games

Definition 91. A **two-player normal-form game** $\mathcal{G}$ consists of:

- Strategy sets S_1, S_2
- Utility functions $u_1 : S_1 \times S_2 \to \mathbb{R}$, $u_2 : S_1 \times S_2 \to \mathbb{R}$

Example 92. A **payoff matrix** is a representation of a two-player normal-form game. The matrix

	L	R
T	1,1	0,0
B	0,0	2,2

corresponds to a normal-form game $\mathcal{G} = (S_1, S_2, u_1, u_2)$ where:

- $S_1 = \{T, B\}$, $S_2 = \{L, R\}$
- $u_1(T, L) = 1$, $u_1(B, R) = 2$, $u_1(T, R) = u_1(B, L) = 0$
- $u_2(T, L) = 1$, $u_2(B, R) = 2$, $u_2(T, R) = u_2(B, L) = 0$.

Example 93. A Cournot duopoloy is a game $\mathcal{G} = (S_1, S_2, u_1, u_2)$ with

- $S_1 = S_2 = [0, \infty)$,
- $u_1(s_1, s_2) = p(s_1 + s_2) \cdot s_1 - C(s_1)$
- $u_2(s_1, s_2) = p(s_1 + s_2) \cdot s_2 - C(s_2)$

where $p(\cdot)$ and $C(\cdot)$ are inverse demand function and cost function, respectively.

Definition 94. The **individual best response correspondences** $BR_1 : S_2 \rightrightarrows S_1$, $BR_2 : S_1 \rightrightarrows S_2$ are defined as:

$$\begin{cases} BR_1(s_2) := \arg\max_{\hat{s}_1 \in S_1} u_1(\hat{s}_1, s_2) \\ BR_2(s_1) := \arg\max_{\hat{s}_2 \in S_2} u_2(s_1, \hat{s}_2) \end{cases}$$

The **best response correspondence** $BR : (S_1 \times S_2) \rightrightarrows (S_1 \times S_2)$ involves putting BR_1 and BR_2 into a vector:

$$BR(s_1, s_2) := (BR_1(s_2), BR_2(s_1))$$

Example 95. Consider the game of rock-paper-scissors:

	R	P	S
R	0,0	-1,1	1,-1
P	1,-1	0,0	-1,1
S	-1,1	1,-1	0,0

The best response correspondence BR in this game is actually a function. Repeated applications of BR yields

$$\begin{pmatrix} \mathbf{R} \\ \mathrm{P} \end{pmatrix} \overset{BR}{\to} \begin{pmatrix} \mathrm{S} \\ \mathbf{P} \end{pmatrix} \overset{BR}{\to} \begin{pmatrix} \mathbf{S} \\ \mathrm{R} \end{pmatrix} \overset{BR}{\to} \begin{pmatrix} \mathrm{P} \\ \mathbf{R} \end{pmatrix} \overset{BR}{\to} \ldots$$

Definition 96. A **pure-strategy Nash equilibrium** (NE) is a strategy profile $(s_1^*, s_2^*) \in S_1 \times S_2$ such that $u_1(s_1^*, s_2^*) \geq u_1(s_1, s_2^*)$ for all $s_1 \in S_1$ and $u_2(s_1^*, s_2^*) \geq u_2(s_1^*, s_2)$ for all $s_2 \in S_2$.

Proposition 97. *A strategy profile $(s_1^*, s_2^*) \in S_1 \times S_2$ is a pure strategy NE if and only if $(s_1^*, s_2^*) \in BR(s_1^*, s_2^*)$.*

Example 98. Pure-strategy NE may not exist. The game of matching pennies, given by the payoff matrix below, does not have any pure-strategy NE.

	H	T
H	1,-1	-1,1
T	-1,1	1,-1

Example 99. Pure-strategy NE may not be unique. The game of coordination, given by the payoff matrix below, has two pure-strategy NEs.

	H	T
H	1,1	0,0
T	0,0	1,1

Definition 100. In a game $\mathcal{G} = (S_1, S_2, u_1, u_2)$ where S_1, S_2 are finite, a **mixed strategy** σ_i for player i is a probability distribution on S_i. We write $\sigma_i(s_i)$ for the probability that the distribution σ_i assigns to $s_i \in S_i$. We also write $\Delta(S_i)$ for the set of all mixed strategies of player i.

Definition 101. The **extended payoff functions** $\tilde{u}_1 : \Delta(S_1) \times \Delta(S_2) \to \mathbb{R}$ and $\tilde{u}_2 : \Delta(S_1) \times \Delta(S_2) \to \mathbb{R}$ are defined by:

$$\tilde{u}_1(\sigma_1, \sigma_2) = \sum_{s_1 \in S_1} \sum_{s_2 \in S_2} u_1(s_1, s_2) \cdot \sigma_1(s_1) \cdot \sigma_2(s_2)$$

$$\tilde{u}_2(\sigma_1, \sigma_2) = \sum_{s_1 \in S_1} \sum_{s_2 \in S_2} u_2(s_1, s_2) \cdot \sigma_1(s_1) \cdot \sigma_2(s_2)$$

Definition 102. A **mixed-strategy Nash equilibrium** is a mixed-strategy profile $(\sigma_1^*, \sigma_2^*) \in \Delta(S_1) \times \Delta(S_2)$ such that $\tilde{u}_1(\sigma_1^*, \sigma_2^*) \geq \tilde{u}_1(\sigma_1, \sigma_2^*)$ for all $\sigma_1 \in \Delta(S_1)$ and $\tilde{u}_2(\sigma_1^*, \sigma_2^*) \geq \tilde{u}_2(\sigma_1^*, \sigma_2)$ for all $\sigma_2 \in \Delta(S_2)$.

Theorem 103. *(Existence of mixed Nash equilibrium). Let $\mathcal{G} = (S_1, S_2, u_1, u_2)$ be a two-player normal-form game where S_1, S_2 are finite. Then $\mathcal{G}$ has at least one mixed Nash equilibrium.*

Proposition 104. *Suppose (σ_1^*, σ_2^*) is a mixed Nash equilibrium. Then for any $s_1 \in S_1$ such that $\sigma_1^*(s_1) > 0$, we have $\tilde{u}_1(s_1, \sigma_2^*) = \tilde{u}_1(\sigma_1^*, \sigma_2^*)$. Similarly, for any $s_2 \in S_2$ such that $\sigma_2^*(s_2) > 0$, we have $\tilde{u}_2(\sigma_1^*, s_2) = \tilde{u}_2(\sigma_1^*, \sigma_2^*)$.*

Corollary 105. *In a mixed Nash equilibrium, each player is indifferent between any two strategies that she plays with strictly positive probability.*

Definition 106. A pure strategy $s_1 \in S_1$ is **strictly dominant** if for any other $s_1' \in S_1$, $s_1' \neq s_1$, we have $u_1(s_1, s_2) > u_1(s_1', s_2)$ for every $s_2 \in S_2$. The analogous definition holds for player 2.

Proposition 107. *If player i has a strictly dominant strategy, then*

- in every pure NE i plays this pure strategy
- in every mixed NE i puts probability 1 on this pure strategy

Definition 108. A pure strategy $s_1 \in S_1$ is **strictly dominated** if there exists $s_1' \in S_1$ such that $u_1(s_1', s_2) > u_1(s_1, s_2)$ for every $s_2 \in S_2$. The analogous definition holds for player 2.

Proposition 109. *If player i has a strictly dominated strategy, then*

- in every pure NE, i does not play this pure strategy
- in every mixed NE i puts probability 0 on this pure strategy

9 Extensive form, signaling

Definition 110. An **extensive-form game** is represented in a game tree,

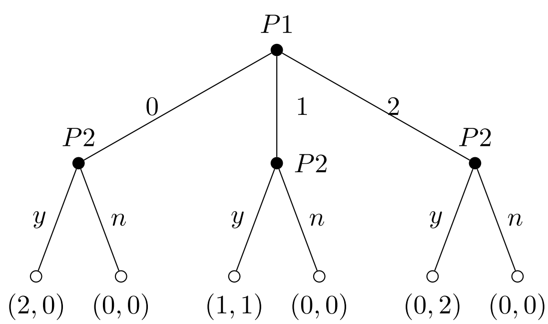


Figure 19: A game tree representation of the ultimatum game.

The game starts at the **root** of the tree. Every non-terminal node is labeled with the player who is to move. The **actions** available to the mover correspond to the different **edges** below the node. Each edge is labeled with the name of the action. Every terminal node is labeled with a vector of **payoffs**, one for each player. Once a terminal node is reached, the game ends and the associated payoffs are realized.

Definition 111. A **complete contingent strategy** (or "**strategy**" for short) for player i specifies which pure strategy i will pick at **every** node in the game tree where it is her turn to play.

Definition 112. Every subtree of the game tree defines a **subgame**. For example, the game in Figure 19 has 4 subgames (3 proper subgames and 1 improper subgame)

Definition 113. A **subgame-perfect equilibrium** (SPE) is a strategy profile such that no player has a strictly profitable unilateral deviation in **any** subgame.

Definition 114. A **Nash equilibrium** (NE) in an extensive-form game is a strategy profile such that no player has a strictly profitable unilateral deviation in the improper subgame (i.e. the whole extensive-form game itself).

Example 115. In Figure 19, the strategy profile where P1 plays 2, P2 plays n when P1 plays 0 or 1, plays y when P1 plays 2 is an NE, but not an SPE. The strategy profile where P1 plays 1, P2 plays n when P1 plays 0, plays y when P1 plays 1 or 2 is an SPE (and hence also an NE).

Fact 116. *Backwards induction can be used to find SPE payoff in an extensive-form game by successively replacing subgames with terminal nodes corresponding to the SPE payoffs in the deleted subgames.*

Definition 117. A **signaling game** consists of:

- a set of two **types** $\Theta = \{\theta_L, \theta_H\}$ for P1
- A **prior distribution** π on Θ
- S_1, set of actions for P1
- S_2, set of actions for P2
- Utility functions $u_i : S_1 \times S_2 \times \Theta \to \mathbb{R}$ for $i = 1, 2$

Remark 118. The type of P1 reflects some hidden characteristic that only P1 knows about, which nonetheless may affect the payoffs of both players in the game. For instance, in the market for lemons, the two types of sellers (P1) refer to whether the seller owns a good car (θ_H) or a bad car (θ_L).

Definition 119. A **perfect-Bayesian equilibrium** (**PBE**) consists of:

- A strategy $\sigma_1 : \Theta \to S_1$ for P1
- A strategy $\sigma_2 : S_1 \to S_2$ for P2
- No type of P1 has a profitable unilateral deviation

- A family of **posterior beliefs** $\{\pi(\cdot|s_1) : s_1 \in S_1\}$
- $\pi(\cdot|s_1)$ is generated by Bayes' rule whenever possible
- P2's action maximizes his expected utility after any s_1 by P1, according to his posterior belief $\pi(\cdot|s_1)$

Definition 120. A PBE is called **separating** if two types of P1 play different pure actions.

Definition 121. A PBE is called **pooling** if two types of P1 play the same pure action.

Definition 122. **Moral hazard** is when an informed player has private information about the action he takes. The action of the informed player affects the expected payoff of the uninformed player, who nevertheless only observes an outcome whose distribution depends on the action, but not the action itself.

Example 123. In the context of car insurance, the driver is the informed player while the insurance company is the uninformed player. The driver's effort (in driving safely) determines his chance of having an accident. This effort affects the expected payoff of the insurance company. However, the company only observes whether the driver has an accident, not the effort level itself.