1 Chapter 2

1. In the first example we saw an instance of

$$impossible \Rightarrow undesirable$$

whereas the second was an instance of

$$possible \Rightarrow undesirable$$

The third example is one in which

$$desirable \Rightarrow possible$$

and this raises the question, how about an example in which

$$desirable \Rightarrow impossible$$

?

You may be reflecting such a belief if you think, for instance, that any potential spouse you may like is bound to be married. Indeed, you may rely on statistics, and a reasonable theory that says that when you like someone, so do others, and therefore they are less likely to be available. But if we all believed that anything desirable is automatically impossible, these desirable spouses will end up being single. By analogy, it is true that you don't often see 100 euro bills on the sidewalk. But the reason is that they are indeed picked up. Someone who sees the bill believes that it might be real and is willing to try to pick it up. Thus, you are justified to believe that something that is really worthwhile may not be easy to find; but you will be wrong to assume that anything worthwhile is automatically unreachable.

2. Symmetry requires that we also look for examples in which

$$\begin{aligned} possible &\Rightarrow desirable \\ impossible &\Rightarrow desirable \\ undesirable &\Rightarrow possible \\ undesirable &\Rightarrow impossible \end{aligned}$$

Can you find such examples?

possible $\Rightarrow$ *desirable* : Habits may provide an example in which you do not try to optimize, and assume that something is what you want only because you know you can have it.

impossible $\Rightarrow$ *desirable* : By the same token, it will be also irrational to want things just because you don't have them. Whereas the previous example leads to too little experimentation, and may make you settle for suboptimal solutions, this example might lead to too much experimentation, and not let you settle on an optimal solution even if you found it.

undesirable $\Rightarrow$ *possible* : The pessimistic assumption that you might be doing something just because you hope not to is reminiscent of "If something can go wrong, it will".

undesirable $\Rightarrow$ *impossible* : This is the optimistic version of the above, a bit similar to "This won't happen to me" (referring to negative events such as accidents etc.).

> 3. George says, "I wish to live in a peaceful world. Therefore, I favor policies that promote world peace."
>
> a. Explain why this statement violates the separation of feasibility and desirability.
>
> b. Assume that George thinks that, if a peaceful world is impossible, he is not interested in living anymore, and, furthermore, he doesn't care about anything else that might happen in this world, to himself or to others. Explain why, under these assumptions, George's statement is compatible with rationality.

This problem is supposed to point out that often people think about what they want, and then reason backwards to see what's needed for that. This may be incompatible with rationality if they forget to ask themselves what is feasible. In George's case, if, as in (b), he doesn't care about anything but his one goal, then it makes sense to ignore what is precisely feasible: if peace is not feasible, he doesn't care about anything anyway. But for many people feasibility will be important. Even when people say that they want peace "at all cost", they do not literally mean it. And then what we expect of rational decision makers is not just to state what is desirable, but also see what is feasible.

> 4. In the above, the symbol $\Rightarrow$ referred to causal implication. First we have the antecedent (on the left hand side), and then, *as a result*, the consequent (on the right hand side) follows. Another notion of implication is *material implication*, often denoted by $\rightarrow$:
>
> $$\begin{aligned} p \;\rightarrow\; q \quad & iff \quad \neg p \vee q \\ & iff \quad \neg(p \wedge \neg q) \end{aligned} \tag{1}$$
>
> then some of the eight implications in (1) and (2) above are redundant. Which are they?[1]

With material implication, $p \rightarrow q$ is equivalent to $\neg q \rightarrow \neg p$ (see Problem 5 below). Hence,

$$possible \rightarrow desirable$$

is equivalent to

$$undesirable \rightarrow impossible$$

and

$$possible \rightarrow undesirable$$

is equivalent to

$$desirable \rightarrow impossible$$

etc. Hence, half of the implications are redundant.

---

[1] Notation: $\neg p$ is the negation of p; i.e., not-p. $\vee$ means "or" and $\wedge$ means "and".

5. Convince yourself that, for the material implication (1),

    a. $p \to q$ is equivalent to $\neg q \to \neg p$

but

    b. $p \to q$ is *not* equivalent to $q \to p$.

There are several ways to convince yourself of (a). First, using "proof by negation": assume that $p \to q$. We want to show that $\neg q \to \neg p$. Assume that indeed $\neg q$, i.e., that q is false. Ask yourself whether p can be true. If it were (contrary to what we want to show), then we could use $p \to q$ to conclude that q is true as well, in contradiction to our assumption $\neg q$. Hence by assuming $\neg q$ we obtained $\neg p$, which is the first part of what we wanted to prove.

The second part is the converse: that is we assume $\neg q \to \neg p$, we also have $p \to q$. We could go through a similar proof, or use the previous one, observing that $\neg\neg p \leftrightarrow p$ and $\neg\neg q \leftrightarrow q$.

Another way to see (a) is to observe that $p \to q$ is simply the statement that "we cannot observe p and not q simultaneously. That is, of the four possible combinations of truth values of p and of q, only three combinations are possible. The possible one are marked by + below, and the impossible one – by $-$:

	q is false	q is true
p is false	$+$	$+$
p is true	$-$	$+$

If we denote "p is true" by A and "q is false" by B, the statement $p \to q$ means that A and B cannot happen together. To say that two events, A and B are incompatible is like saying "If A, then not B", or "If B, then not A".

To see (b), take a simple example such as: because all humans are mortal, *mortal* → *human*. But, because dogs are also mortal, it is false that *mortal* → *human*. This looks trivial in such simple examples, and yet, people make such mistakes often both (i) in the heat of a debate; and (ii) when probabilities are involved. (And this we will see later on.)

2 Chapter 3

> 1. To what degree is the function u in Proposition **??** and Theorem **??** unique? That is, how much freedom to you have, as the outside modeler, in choosing the utility function u for a given relation $\succsim$?

The utility function is unique up to a monotone transformation. That is, if u represents $\succsim$, then so will any other function

$$v : X \to \mathbb{R}$$

such that there exists a (strictly) monotonically increasing

$$f : \mathbb{R} \to \mathbb{R}$$

for which

$$v(x) = f(u(x)) \tag{2}$$

for every $x \in X$. Conversely, if both u and v represent $\succsim$, it is easy to see that, for every $x, y \in X$

$$u(x) > u(y) \Leftrightarrow v(x) > v(y)$$

and this means that there exists a (strictly) monotonically increasing $f : \mathbb{R} \to \mathbb{R}$ such that (2) holds.

> 2. Assume that apart from preferences between pairs of alternatives $x \succsim y$ or $y \succsim x$, more data is available, such as (i) the probability that x is chosen out of $\{x, y\}$; or (ii) the time it takes the decision maker to make up her mind between x and y; or (iii) some neurological data that show us the strength of preference between x and y. Consider different representation of preferences, corresponding to (i)-(iii), which will also restrict the set of utilities one can ascribe to the decision maker.

Let us assume that in reality there exists some numerical measure of desirability, $u(x)$, which is not directly observable to us. Yet, we may find the following observable manifestations of this measure:

(i) The probability of choosing x out of $\{x, y\}$ may be increasing as a function of the utility difference, $u(x) - u(y)$. Our standard model implicitly assumes that this probability is

$$\Pr(x\_over\_y) = \begin{cases} 1 & u(x) > u(y) \\ 0.5 & u(x) = u(y) \\ 0 & u(x) < u(y) \end{cases} \tag{3}$$

but this discontinuous function is not very realistic. Instead, we may consider a function such as the cumulative distribution function of the Normal (Gaussian) distribution with parameters $(0, \sigma)$, namely

$$\Pr(x\_over\_y) = \int_{-\infty}^{u(x)-u(y)} e^{-\frac{t^2}{\sqrt{2\pi}\sigma}} dt$$

such that $\Pr(x\_over\_y)$ converges to (3) as $\sigma \to 0$. This function would make utility differences observable by the probability of choice.

(ii) The standard model ignores response time, or the time it takes the decision maker to reach a decision. We might consider a function such as

$$R(x,t) = c + de^{-\theta[u(x)-u(y)]^2}$$

such that $c > 0$ is the minimal response time, obtained when the choice is very clear (i.e., when the absolute difference between the utility levels is approaching infinity), and the maximal response time, $c + d$, is obtained when the two alternatives are equivalent in the eyes of the decision maker.

(iii) Finally, the standard model treats anything that goes in our brains as unobservable. But recent neurological studies identify zones of the brain that tend to be activated when the alternatives are close to equivalent but not otherwise. Thus, neurological data may be another source of information on strength of preferences.

Overall, the lesson is that the fact that the utility function is "only ordinal", and that we cannot discuss strength of preferences is a result of our highly idealized assumption that only choice is observable, and that choice is deterministic as in (3). It is our choice to focus on such a model. In reality, however, much more information is available, and this additional information may suffice to pinpoint a "cardinal" utility function, that is, one that is more or less unique – at least up to a linear transformation of the type

$$v(x) = \alpha u(x) + \beta$$

with $\alpha > 0$.

> 3. Assume that $X = \mathbb{R}^2$ and that, due to some axioms, you are convinced that your utility function should be of the form
>
> $$u(x_1, x_2) = v_1(x_1) + v_2(x_2).$$
>
> Discuss how this additional structure may help you estimating your own utility function, and contrast this case with the (end of the dialog) we started out with.

In this case, one can try to learn something about one's preferences in complex choices from one's preferences in simple ones. For example, suppose that after intensive introspection you realize that your preferences satisfy

$$(x_1, x_2) \sim (y_1, y_2)$$

and

$$(z_1, x_2) \sim (w_1, y_2)$$

The first equivalence means that

$$v_1(x_1) - v_1(y_1) = v_2(y_2) - v_2(x_2) \tag{4}$$

and the second –

$$v_1(z_1) - v_1(w_1) = v_2(y_2) - v_2(x_2) \tag{5}$$

Next suppose that we also find that

$$(x_1, s_2) \sim (y_1, r_2)$$

Which means that

$$v_1(x_1) - v_1(y_1) = v_2(r_2) - v_2(s_2) \tag{6}$$

It then follows that we should also have

$$(z_1, s_2) \sim (w_1, r_2)$$

because we already know that (combining (4) and (5))

$$v_1(z_1) - v_1(w_1) = v_2(y_2) - v_2(x_2) = v_1(x_1) - v_1(y_1)$$

and, because of (6), also

$$v_1(z_1) - v_1(w_1) = v_1(x_1) - v_1(y_1) = v_2(r_2) - v_2(s_2).$$

In other words, additional structure on the utility function will make the elicitation of utility a non-circular exercise.

4. Prove that, if $\succsim$ is transitive, then so are $\succ$ and $\sim$.

In this type of proofs the main thing is to keep track of what is given and what should be proved. Most mistakes in this exercises arise from getting confused about these. Also, much of the proof is a translation of the symbols using their definitions. For these reasons it is best to write things down very carefully and precisely, even though it might look silly or boring.

Let us begin with $\sim$. Assume that (for some $x, y, z \in X$) $x \sim y$ and that $y \sim z$. We need to show that $x \sim z$. Let us first translate both premise and desired conclusion to the language of the relation $\succsim$ about which we know something (i.e., that it is transitive).

By definition of $\sim$, $x \sim y$ means that

$$x \succsim y \qquad \text{and} \qquad y \succsim x \tag{7}$$

whereas $y \sim z$ is a shorthand for

$$y \succsim z \qquad \text{and} \qquad z \succsim y. \tag{8}$$

What we need to prove is that $y \sim z$, namely that

$$x \succsim z \qquad \text{and} \qquad z \succsim x. \tag{9}$$

The first parts of (7) and (8) are, respectively, $x \succsim y$ and $y \succsim z$, and given the transitivity of $\succsim$, they yield $x \succsim z$. This is the first part of (9).

Similarly, the second parts of (7) and (8) are, respectively, $y \succsim x$ and $z \succsim y$, which, given transitivity (of $\succsim$) imply $z \succsim x$. This is the second part of (9).

Since this is true for any x, y, z with $x \sim y$ and $y \sim z$, transitivity of $\sim$ has been established.

Next turn to transitivity of $\succ$. Assume that (for some $x, y, z \in X$) $x \succ y$ and $y \succ z$. We need to show that $x \succ z$. Again, let us translate both the premises and the desired conclusion to the language of $\succsim$.

By definition of $\succ$, $x \succ y$ means that

$$x \succsim y \qquad \text{and} \qquad \neg\,(y \succsim x) \tag{10}$$

while $y \succ z$ is the statement

$$y \succsim z \qquad \text{and} \qquad \neg\,(z \succsim y)\,. \tag{11}$$

We need to show that $x \succ z$ which means

$$x \succsim z \qquad \text{and} \qquad \neg\,(z \succsim x)\,. \tag{12}$$

The first part of (12) follows from the first parts of (10) and of (11) by transitivity of $\succsim$. The second part of (12) will be proved by negation: suppose that, contrary to our claim, $z \succsim x$ does hold. Combining this with $x \succsim y$ (the first part of (10)) we get, by transitivity of $\succsim$, that $z \succsim y$. But this would contradict the second part of (11). Hence $z \succsim x$ cannot be true and the second part of (12) is also true. Again, this holds for every x, y, z, and this completes the proof.

5. Assume that $\succsim$ is complete. Prove that u represents $\succsim$ if and only if, for every $x, y \in X$,

$$x \succ y \qquad \Leftrightarrow \qquad u(x) > u(y).$$

The simplest way to see this is to observe that, for real numbers a, b,

$$\neg(a \geq b) \Leftrightarrow b > a$$

and, because $\succsim$ is complete, a similar fact holds for preferences: for every $z, w \in X$,

$$\neg(x \succsim y) \Leftrightarrow y \succ x.$$

Once this is established, we can use the contrapositive (exercise 5a from the previous chapter to conclude the proof. But before we do it, a word of warning: we know that

$$x \succsim y \qquad \Leftrightarrow \qquad u(x) \geq u(y) \qquad \forall x, y \in X \tag{13}$$

and we need to show that,

$$x \succ y \qquad \Leftrightarrow \qquad u(x) > u(y) \qquad \forall x, y \in X. \tag{14}$$

This is rather simple unless we get ourselves thoroughly confused with the x's in (13) and in (14). It is therefore a great idea to replace (14) the latter by

$$z \succ w \qquad \Leftrightarrow \qquad u(z) > u(w) \qquad \forall z, w \in X. \tag{15}$$

You may take a few second to verify that (14) and (15) mean the same thing. Since we range over all x, y in (14) and over all z, w in (15), these variables have no existence outside the respective expressions. Replacing "all x" by "all z" is similar to changing the index inside a summation. That is, just as

$$\sum_{i=1}^{n} a_i = \sum_{j=1}^{n} a_j$$

the statements (14) and (15) are identical.

If we agree on this, we can now observe that, for every x, y,

$$x \succsim y \qquad \Rightarrow \qquad u(x) \geq u(y)$$

is equivalent to

$$\neg\,(u(x) \geq u(y)) \qquad \Rightarrow \qquad \neg\,(x \succsim y)$$

or to

$$u(y) > u(x) \qquad \Rightarrow \qquad y \succ x.$$

Thus, for every $z(= y)$, and $w(= x)$,

$$u(z) > u(w) \qquad \Rightarrow \qquad z \succ w.$$

Similarly, for every x, y,

$$u(x) \geq u(y) \qquad \Rightarrow \qquad x \succsim y$$

is equivalent to

$$\neg\,(x \succsim y) \qquad \Rightarrow \qquad \neg\,(u(x) \geq u(y))$$

or to

$$y \succ x \qquad \Rightarrow \qquad u(y) > u(x)$$

and, again, for every $z(= y)$, and $w(= x)$,

$$z \succ w \qquad \Rightarrow \qquad u(z) > u(w).$$

6. Assume that $X = [0, 1]^2$ and that $\succsim$ is defined by

$$(x_1, x_2) \succsim (y_1, y_2)$$

if

$$\begin{array}{l} [x_1 > y_1] \\ \textit{or} \\ \left[(x_1 = y_1) \quad \textit{and} \quad (x_2 \geq y_2)\right]. \end{array}$$

> Prove that $\succsim$ is complete and transitive but not continuous. Prove that $\succsim$ cannot be represented by any utility u (continuous or not).

To see that $\succsim$ is complete, consider (x_1, x_2) and (y_1, y_2). If $x_1 > y_1$ then $(x_1, x_2) \succsim (y_1, y_2)$. Similarly, $y_1 > x_1$ implies $(y_1, y_2) \succsim (x_1, x_2)$. We are left with the case $x_1 = y_1$. But then $x_2 \geq y_2$ (and $(x_1, x_2) \succsim (y_1, y_2)$) or $y_2 \geq x_2$ (and then $(y_1, y_2) \succsim (x_1, x_2)$).

Next turn to transitivity. Assume that $(x_1, x_2) \succsim (y_1, y_2)$ and $(y_1, y_2) \succsim (z_1, z_2)$. If $x_1 > y_1$ or $y_1 > z_1$, then $x_1 > z_1$ and $(x_1, x_2) \succsim (z_1, z_2)$ follows. Otherwise, $x_1 = y_1 = z_1$. Then $(x_1, x_2) \succsim (y_1, y_2)$ implies $x_2 \geq y_2$ and $(y_1, y_2) \succsim (z_1, z_2)$ implies $y_2 \geq z_2$. Together we have $x_2 \geq z_2$ which implies (since we already know that $x_1 = z_1$) that $(x_1, x_2) \succsim (z_1, z_2)$.

To see that continuity does not hold, consider $y = (y_1, y_2)$ with

$$\begin{aligned} y_1 &= 0.5 \\ y_2 &= 1 \end{aligned}$$

and the set

$$B(y) = \{((x_1, x_2) \, | (y_1, y_2) \succ (x_1, x_2)\} .$$

You may verify that $(0.5, 0) \in B(y)$ but for every $\varepsilon > 0$, $(0.5 + \varepsilon, 0) \notin B(y)$ (because $(0.5 + \varepsilon, 0) \succ (0.5, 1) = y$). Hence $B(y)$ is not open, which is sufficient to show that continuity of $\succsim$ does not hold.

We know from Debreu's theorem that $\succsim$ cannot be represented by a continuous utility function. This can also be verified directly in the example above. Indeed, if there were a continuous u that represented $\succsim$, we would have

$$u\left((0.5 + \varepsilon, 0)\right) > u(y) > u\left((0.5, 0)\right) \tag{16}$$

for every $\varepsilon > 0$. But this is incompatible with continuity, because

$$(0.5 + \varepsilon, 0) \to (0.5, 0)$$

as $\varepsilon \to 0$, and continuity would have implied that

$$u\left((0.5 + \varepsilon, 0)\right) \to u\left((0.5, 0)\right)$$

whereas (16) means that the left hand side above is bounded below by a number $(u(y))$ strictly larger than $u\left((0.5,0)\right)$.

To see that no utility function can represent $\succsim$ requires a little more knowledge of set theory. We can try to give the intuition here: if u represented $\succsim$, then we would find that the function

$$w(z)=u\left((z,0)\right)$$

has a discontinuity from the right at $z=0.5$. That is, as we have just seen,

$$\lim_{\varepsilon\to 0}u\left((z+\varepsilon,0)\right)>u\left((z,0)\right).$$

This is not the end of the world, as we don't expect u to be continuous any more. But the above is true not only for $z=0.5$, but for *any* $z\in(0,1)$. And, what complicated matters more, is that $w(z)$ is a monotone function (the higher z, the better is $(z,0)$ and the higher should be $u\left((z,0)\right)=w\left(z\right)$).

The contradiction arises from the fact that a monotone function can have "jumps", but not *everywhere*. Roughly, this has to do with the fact that the set of "jumps" of a monotone function is countable, whereas the interval $(0,1)$ isn't.

This issue of countability may be hard to grasp. It is, if you see it for the first time. So let us conclude with two lessons. First, you should be aware that there are hierarchies of infinities in set theory – among infinite sets that are larger infinities and smaller infinities. The smallest infinity is that of the natural numbers, $\{1,2,3,...\}$, which is called countable infinity. The points in $(0,1)$ (or in any interval of positive length) are not countable. In a well-defined sense, there are more points in $(0,1)$ than there are natural numbers.

The second lesson is conceptual. The lexicographic example above might seem like a mathematical oddity. But lexicographic relations often appear in everyday speech. For instance, you can imagine a politician saying that we will give our public the best health care possible, but, subject to this level of health care, we will save on costs. Or that we will promote minority candidates, provided, "of course", that we do not compromise on quality. These are examples of lexicographic relations. These are also often examples of dishonesty. Typically, trade-off do exist. If you need to save money on health care, you might have to compromise on the quality of health care. If you want to promote a social agenda and help minorities, you might have to compromise on quality. Politicians will often try to disguise this compromise. The lexicographic example above, showing that we can easily describe a function that cannot be represented numerically, suggests that maybe our politicians do not really mean what they say. That is, it might be more honest to say, "we have to cut on health costs, and we'll try to do it without hurting the quality of serving too much". Or, "It's important to have affirmative action, and we're willing to pay some price for that". So even if you didn't get the last bit about uncountable sets, next time you hear someone describing preferences lexicographically, ask yourself whether they really mean what they say.

3 Chapter 4

> 1. You are organizing an inter-disciplinary conference, and you wish to have a good mix of psychologists, economists, and sociologists. There are many scientists of each type, but the cost of inviting them grows with distance – it is relatively inexpensive to invite those that are in your city, but it gets expensive to fly them from remote countries. State the problem as a constrained optimization problem. Is this a convex problem? What do the first order conditions tell you?

Suppose that you invite x_1 psychologists, x_2 economists, and x_3 sociologists, and let us make the unrealistic but convenient assumption that these are real numbers, i.e., that scientists are divisible. Let $u(x_1, x_2, x_3)$ be a measure of how good the conference is as a function of the number of scientists you invite from each group. And even more unrealistic assumption here is that all psychologists are interchangeable, as are all economists (among themselves) and all sociologists. This is implicit in the formulation asking "how many... do we invite?", ignoring their identity.

The story implicitly refers to only one constraint, namely, cost. However, it is clear that cost is not linear, because it grows with distance. Let, then $c_i(x_i)$ be the cost of inviting x_i scientists of type i (1 for psychologists, 2 for economists, and 3 for sociologists), and let B be your overall budget. The optimization problem is then

$$\begin{aligned} & Max_{x_1,x_2,x_3} u(x_1, x_2, x_3) \\ & \text{subject to} \\ c_1(x_1) + c_2(x_2) + c_3(x_3) & \leq B \\ x_i & \geq 0 \end{aligned}$$

This will be a convex problem provided that the cost functions are weakly convex and that the utility function is quasi-concave. Specifically, if the cost functions are (weakly) convex, then, for every $i = 1, 2, 3$, every $x_i, y_i \geq 0$, and every $\alpha \in [0, 1]$,

$$\alpha c_i(x_i) + (1-\alpha)c_i(y_i) \geq c_i\left(\alpha x_i + (1-\alpha)y_i\right)$$

and this means that is $x = (x_1, x_2, x_3)$ and $y = (y_1, y_2, y_3)$ are in the feasible set, then so is $\alpha x + (1-\alpha)y$: non-negativity of x_i and of y_i implies non-negativity of $\alpha x_i + (1-\alpha)y_i$, and

$$\begin{aligned}
\sum_{i=1}^{3} c_i\left(\alpha x_i + (1-\alpha)y_i\right) &\leq \\
\sum_{i=1}^{3} \left[\alpha c_i(x_i) + (1-\alpha)c_i(y_i)\right] &= \\
\alpha \sum_{i=1}^{3} c_i(x_i) + (1-\alpha)\sum_{i=1}^{3} c_i(y_i) &\leq \\
\alpha B + (1-\alpha)B &= B.
\end{aligned}$$

Quasi-concavity of u means precisely that the "better than" sets are convex, i.e., the set

$$\left\{x \in R^3 \left| u(x_1, x_2, x_3) \geq \gamma \right.\right\}$$

is convex for every γ. With a convex feasible set and convex "better than" sets, the problem is convex.

The first order conditions can be obtained from taking the derivatives of the Lagrangian,

$$L(x_1, x_2, x_3, \lambda) = u(x_1, x_2, x_3) + \lambda\left[B - c_1(x_1) + c_2(x_2) + c_3(x_3)\right]$$

which yield

$$\frac{\partial L}{\partial x_i} = u_i(x_1, x_2, x_3) - \lambda c_i'(x_i).$$

with $u_i(x_1, x_2, x_3) = \frac{\partial u}{\partial x_i} u(x_1, x_2, x_3)$. Equating all to zero, we get

$$\frac{u_i(x_1, x_2, x_3)}{c_i'(x_i)} = \lambda$$

that is, the ratio of the marginal utility to marginal cost should be the same across all decision variables x_i. Given that the problem is convex, if we found such a point, it is optimal. Note, however, that there may not exist such a point, and the optimal problem may well be at a "corner" solution, for example, if the sociologists turn out to be too expensive and the optimal solution is to invite none of them ($x_3 = 0$).

> 2. Provide an example of a consumer problem in which the optimal solution does not satisfy the first order conditions. (Hint: you may use two goods and use a simple budget set such as that defined by $x_1 + x_2 \leq 100$.)

Given the budget constraint $x_1 + x_2 \leq 100$, consider the utility function

$$u(x_1, x_2) = 2x_1 + x_2$$

Clearly, the optimal solution is at $(100, 0)$. You can also generate such an example if the utility function is strictly quasi-concave: all you need to guarantee is that the slope of the indifference curves will be steep enough so that there will be no tangency point between these curves and the budget line. Specifically, if throughout the range

$$\frac{u_1}{u_2} > 1$$

the optimal solution will be at $(100, 0)$ without the marginality condition holding.

> 3. Assume that you have to allocate a given amount of time among several friends. Unfortunately, since they live far away, you can't meet more than one friend at the same time. Let x_i the amount of time you spend with friend i. Formulate the problem as a constrained optimization problem. Is it convex?

The problem might look like

$$\begin{aligned} & Max_{x_1,...,x_n} u(x_1,..,x_n) \\ & \text{subject to} \\ x_1+...+x_n & \leq B \\ x_i & \geq 0 \end{aligned}$$

that is, like a standard consumer problem where the prices of all "goods" are 1.

If you like to see each friend as much as possible, u will be monotonically increasing. For the problem to be convex, you would like u to be quasi-concave. That is, if, for two feasible time-allocation vectors, $(x_1,..,x_n),(y_1,..,y_n)$, if each guarantees a utility value of c at least, so should

$$\lambda(x_1,..,x_n)+(1-\lambda)(y_1,..,y_n).$$

This is a reasonable condition if, at any level of the variables, you prefer to "mix", and to have some variety. But if may not hold if the values are too low. For instance, if you mainly derive pleasure from gossip, it seems that frequent changes among friends is a great thing. But if you wish to get into a deep conversation about your emotional life, you may find that one hour with one friends is better than six 10-minute sessions with different friends.

> 4. Show that, in the presence of discounts for quantity (that is, the price per unit goes down as you buy large quantities) the feasible set of the consumer is not convex.

Suppose that the prices of goods 1 and 2 are $p_1=p_2=1$, and that you have income of $I=200$. But if $x_1>100$, the price of good 1 drops to 1/2. Then the feasible (budget) set is bounded above by the segment connecting $(0,200)$ and $(100,100)$ (for $0\leq x_1\leq 100$) and by the segment connecting $(100,100)$ and $(300,0)$. Thus, the points $A=(0,200)$ and $B=(300,0)$ are in the feasible set, but the point $(100,\ 133\frac{1}{3})$, which is on the segment connecting them, isn't in the feasible set.

> 5. Show that the intersection of convex sets is convex.

Let there be two convex sets $A, B \subset \mathbb{R}^n$. Consider $C = A \cap B = \{x | x \in A$ and $x \in B\}$. To see that C is convex, consider $x, y \in C$ and $\lambda \in [0,1]$. We need to show that

$$\lambda x + (1-\lambda)y \in C.$$

By convexity of A (and since $x, y \in A$), we get $\lambda x + (1-\lambda)y \in A$. Similarly, $\lambda x + (1-\lambda)y \in B$. But this means that $\lambda x + (1-\lambda)y$ is both in A and in B, that is, in their intersection, C.

> 6. A half-space is defined by a (weak) linear inequality. That is, for a linear function $f : \mathbb{R}^n \to \mathbb{R}$ and a number $c \in \mathbb{R}$, it is the set of points
>
> $$H(f,c) \equiv \{x \in \mathbb{R}^n \, | f(x) \le c\}$$
>
> Show that the intersection of (any number of) half-spaces is convex.

First, we need to convince ourselves that a single half-space is convex. To see this, assume that

$$x, y \in H(f,c)$$

that is,

$$f(x), f(y) \le c.$$

Because f is linear, for any $\lambda \in [0,1]$ we have

$$f(\lambda x + (1-\lambda)y) = \lambda f(x) + (1-\lambda) f(y)$$

hence

$$f(\lambda x + (1-\lambda)y) \le c$$

and

$$\lambda x + (1-\lambda)y \in H(f,c)$$

that is, $H(f,c)$ is convex.

Next, we will show that any intersection of convex sets is convex. We will follow the same reasoning that applied in Problem 5 for two sets to any collection of sets. That is, assume that $\{A_\alpha\}_\alpha$ is some collection of convex sets, where α is an index that ranges over a certain set. (If α assumes only finitely many values, you can apply the conclusion of Problem 5 inductively. But the fact is true even if there are infinitely many α's.) Then

$$A^* = \cap_\alpha A_\alpha = \{x | x \in A_\alpha \quad \forall \alpha\}$$

is convex because, for any $x, y \in A^*$ and any $\lambda \in [0,1]$, we have

$$x, y \in A_\alpha \qquad \forall \alpha$$

and by convexity of A_α,

$$\lambda x + (1-\lambda) y \in A_\alpha \qquad \forall \alpha$$

and this means

$$\lambda x + (1-\lambda) y \in \cap_\alpha A_\alpha = A^*.$$

Hence, $x, y \in A^*$ implies that $\lambda x + (1-\lambda) y \in A^*$ for any $\lambda \in [0,1]$, and this is the definition of a convex set.

4 Chapter 5

> 1. A concave utility function can explain why people buy insurance with a negative expected value. And a convex utility function can explain why they buy lottery tickets, whose expected value is also negative. But how would you explain the fact that some people do both simultaneously?

One explanation is that the utility function looks like an inverse "S": concave up to a certain point, and convex thereafter. Imagine that w is the "inflection point" – the wealth level above which u is convex and below which it is concave. Then, if the decision maker is at w, considering a major loss (as in the case of insurance) she behaves like a risk-averse person, but considering a major gain (as in a lottery), she behaves in a risk-loving manner.

The problem with this explanation is that it seems unlikely that all the people who both (i) insure their property and (ii) buy lotteries are at the inflection point of their indifference curve. Another explanation is that this inflection point "moves" around with the current wealth level: that the utility function depends on the wealth the individual already has.

This is very similar to the idea of a "reference point" of Kahneman and Tversky. They argued that people respond to changes in the wealth level, rather than to the absolute levels of wealth. Moreover, they suggested that people react differently to gains as compared to losses. However, they found in their experiments that people are risk averse when it comes to gains and risk loving when it comes to losses. This appears to be in contradiction to our S-shaped utility above. At the same time, the sums of gains and losses involved in lotteries and insurance problems are much larger than the sums used in experiments.

Another explanation of the gambling behavior is that gambling is not captured by expected utility maximization at all. Rather, gambling has an entertainment value (people enjoy the game) or a fantasy value (people enjoy to fantasize about what they will do with the money they win). And these cannot be captured by expectation of a utility function which is defined over outcomes alone.

> 2. Assume that a decision maker's preference $\succsim$ is representable by median utility maximization. That is, for a function $u : X \to \mathbb{R}$, and a lottery $P \in L$, define
>
> $$med_P u = \inf \left\{ \alpha \,\middle|\, \sum_{u(x)\geq\alpha} P(x) \geq \frac{1}{2} \right\}$$
>
> and
>
> $$P \succsim Q \qquad \Leftrightarrow \qquad med_P u \geq med_Q u$$
>
> for all $P, Q \in L$.
>
> Show that $\succsim$ is a weak order, but that it violates continuity and independence.

To see that $\succsim$ is a weak order it sufficed to note that it is defined by maximization of a real-valued function. Since every lottery P is mapped to

$$V(P) = \inf \left\{ \alpha \,\middle|\, \sum_{u(x)\geq\alpha} P(x) \geq \frac{1}{2} \right\}$$

and the decision maker maximizes $V(P)$, the relation is complete and transitive (as is the relation $\geq$ on the real numbers).

To see that $\succsim$ is not continuous, assume for simplicity that $X = \mathbb{R}$ and $u(x) = x$, and consider the lotteries P, Q, R defined as follows:

$$P = \begin{cases} 10 & 0.5 \\ 0 & 0.5 \end{cases};$$

Q guarantees the outcome 5 with probability 1; and R guarantees the outcome 0 with probability 1. Then

$$V(P) = 10 \quad V(Q) = 5 \quad V(R) = 0$$

hence

$$P \succ Q \succ R.$$

However, for any $\alpha \in (0,1)$,

$$V\left(\alpha P + (1-\alpha)R\right) = 0 < V(Q)$$

and

$$Q \succ \alpha P + (1-\alpha)R$$

which contradicts the continuity axiom.

As for independence, consider the same P, Q, R above, and observe that

$$P \succ Q$$

but if we "mix" them with R and $\alpha = .7$, we get

$$\begin{aligned} V\left(.7P + .3R\right) &= 0 \\ V(.7Q + .3R) &= 5 \end{aligned}$$

thus

$$.7Q + .3R \succ .7P + .3R$$

in violation of the independence axiom, which would have implied

$$\alpha P + (1-\alpha)R \succ \alpha Q + (1-\alpha)R.$$

> 3. If $\succsim$ is representable by median u maximization as above, how unique is u? That is, what is the class of functions v such that median v maximization also represents $\succsim$?

In this case, u is unique up to (any) monotone transformation. The median ranking depends only on the ordering of the various outcomes, and thus any transformation that preserves this ordering can also serve as "the" utility function.

4. Suppose that the utility function from money, u, is twice differentiable and satisfies $u' > \delta > 0$ and $u'' < 0$. Let X be a bounded random variable with $EX > 0$.

a. Show that, for every wealth level W, there exists $\varepsilon > 0$ such that

$$E\left[u(W + \varepsilon X)\right] > u(W)$$

b. Show that there exists a wealth level W such that, for all $w \geq W$,

$$E\left[u(w + X)\right] > u(w)$$

Let us make the simplifying assumption that X assumes only two values. (The statements are true in general, but calculating expectations involves measure-theoretic terms that are probably not familiar to most students.) Specifically, assume that

$$X = \begin{cases} a & p \\ b & 1-p \end{cases}$$

with $a > 0 > b$ and $EX = pa + (1-p)b > 0$. Denote this expectation by $c = EX$.

We know that, if the utility function were linear ($u'' = 0$), the decision maker would prefer to add αX to her current wealth level w, for any $\alpha > 0$ and any w. This is so because for a linear u,

$$E\left[u(w + \alpha X)\right] = u(w) + \alpha EX > u(w).$$

Being risk averse, the above may not hold in general. However, we should expect it to be true if u can be approximated by a linear function, that is, if the decision maker is roughly risk neutral.

In clauses (a) and (b) we have different reasons for which we can think of the decision maker as "roughly risk neutral", that is, to approximate her utility function by a linear one. In the first case, the approximation is "local" – using the tangent to the utility function's graph as the linear approximation. In the second case, the main point is that the utility function has a decreasing but positive derivative, and it therefore has to converge to a constant derivative, i.e., to a linear function. Some more details:

a. Here we want to approximate $u(x)$ by

$$v(x) = u(W) + (x - W)u'(W)$$

that is, by the tangent to the curve of u at W.

To simplify notation a little bit, we may change the variable so that $W = 0$. (Formally, introduce a new variable $y = x - W$.) Also, since u is given up to a positive linear transformation, no loss of generality is involved in assuming that $u(0) = 0$ and $u'(0) = 1$. Under these assumptions, we also have

$$v(x) = x.$$

Thus, the expected v-value of αX is simply $\alpha c > 0$ (for any $\alpha > 0$).

Differentiability of u means that

$$\left|\frac{u(x) - x}{x}\right| \rightarrow_{x \rightarrow 0} 0.$$

Let us now consider the expected utility of $w + \varepsilon X = \varepsilon X$. We have

$$E[u(\varepsilon X)] = pu(\varepsilon a) + (1 - p)u(\varepsilon b)$$

and we wish to approximate it by the expected utility of $v(x) = x$, which is

$$E[\varepsilon X] = p\varepsilon a + (1 - p)\varepsilon b = \varepsilon c > 0.$$

Explicitly,

$$\begin{aligned}
E[u(\varepsilon X)] &= pu(\varepsilon a) + (1 - p)u(\varepsilon b) \\
&= p\varepsilon a + p[u(\varepsilon a) - \varepsilon a] + (1 - p)\varepsilon b + (1 - p)[u(\varepsilon b) - \varepsilon b] \\
&= \varepsilon c + p\varepsilon a \left[\frac{u(\varepsilon a) - \varepsilon a}{\varepsilon a}\right] + (1 - p)\varepsilon b \left[\frac{u(\varepsilon b) - \varepsilon b}{\varepsilon b}\right] \\
&= \varepsilon \left(c + pa \left[\frac{u(\varepsilon a) - \varepsilon a}{\varepsilon a}\right] + (1 - p)b \left[\frac{u(\varepsilon b) - \varepsilon b}{\varepsilon b}\right]\right)
\end{aligned}$$

or

$$\frac{E[u(\varepsilon X)]}{\varepsilon} = c + pa\left[\frac{u(\varepsilon a) - \varepsilon a}{\varepsilon a}\right] + (1-p)b\left[\frac{u(\varepsilon b) - \varepsilon b}{\varepsilon b}\right]$$

Since the two expressions in square brackets converge to zero as $\varepsilon \to 0$, the expression above converges to $c > 0$. This proves our claim.

The meaning of this result is that if we take a decision maker who has a constant (risk-free) asset W, and offer her to invest in an asset X with positive expected value, she would like to invest at least some amount $\varepsilon > 0$ in the asset X, even if she is risk averse.

This conclusion may not be entirely realistic because the expected utility gain, for a very small ε, may not exceed the transaction cost (say, of buying an asset), and it may be also just too small for the decision maker to notice.

b. Since the derivative of the utility function, u', is a positive and decreasing (because u is increasing and concave), we know that it converges to a limit:

$$u'(w) \searrow_{w\to\infty} d \geq \delta > 0.$$

(the notation $\searrow$ means "converges from above").

Consider the expected utility of getting the asset X with initial assets fixed at w:

$$E[u(w+X)] = pu(w+a) + (1-p)u(w+b).$$

We wish to show that for a large enough w it is higher than $Eu(w) = u(w)$. That is, we wish to show that the following expression is positive (for w large enough):

$$E[u(w+X)] - u(w).$$

Observe that

$$\begin{aligned} & E[u(w+X)] - u(w) \\ = \quad & pu(w+a) + (1-p)u(w+b) - [pu(w) + (1-p)u(w)] \\ = \quad & p[u(w+a) - u(w)] + (1-p)[u(w+b) - u(w)] \end{aligned}$$

We know that a difference of a the values of a differentiable function between two points is equal to the distance between the points times the derivative at some point between them. That is, if you consider w and $w+a$, there exists $w' \in [w, w+a]$ such that

$$u(w+a) - u(w) = au'(w')$$

and there also exists $w'' \in [w+b, w]$ such that

$$u(w+b) - u(w) = bu'(w'')$$

Using these, we can write

$$\begin{aligned}
& E[u(w+X)] - u(w) \\
= \; & pau'(w') + (1-p)bu'(w'') \\
= \; & pau'(w') + (1-p)bu'(w') + (1-p)b[u'(w'') - u'(w')] \\
= \; & cu'(w') + (1-p)b[u'(w'') - u'(w')] \\
= \; & u'(w')\left[c + (1-p)b\frac{u'(w'') - u'(w')}{u'(w')}\right]
\end{aligned}$$

As $w \to \infty$, $w', w'' \in [w+b, w+a]$ also converge to infinity, and $u'(w'), u'(w'') \to d$. This implies that

$$u'(w') - u'(w'') \to 0$$

and, because the denominator $u'(w') \geq d > 0$, the expression in square brackets above converges to $c > 0$. Hence, the entire expression converges to $dc > 0$, and for all w from that point on, $E[u(w+X)]$ will be strictly higher than $u(w)$.

The meaning of this result is that, when one becomes very rich, one tends to be risk neutral. This is not quite accurate, because, as Kahneman and Tversky pointed out, people react to changes to their "reference point", and not to absolute levels of overall wealth.

5. Show that, if $\succsim$ satisfies the vNM axioms, then, whenever $P \succ Q$,

$$\begin{aligned}
\alpha P + (1-\alpha)Q \;&\succsim\; \beta P + (1-\beta)Q \\
&\text{iff} \\
\alpha \;&\geq\; \beta
\end{aligned}$$

Assume that $P \succ Q$. Consider $\alpha \in (0,1)$. Use the independence axiom with P, Q, and $R = Q$ to obtain

$$\alpha P + (1-\alpha)Q \succ Q$$

and the same axioms with P, Q, and $R = P$ to obtain

$$P \succ \alpha P + (1-\alpha)Q.$$

Thus, whenever $P \succ Q$,

$$P \succ \alpha P + (1-\alpha)Q \succ Q.$$

Next, consider $\alpha, \beta \in (0,1)$. If $\alpha = \beta$ then the equivalence $\alpha P + (1-\alpha)Q \sim \beta P + (1-\beta)Q$ is trivial (because it is precisely the same lottery on both sides). Assume, then, without loss of generality, that $\alpha > \beta$. The point to note is that $\beta P + (1-\beta)Q$ can be described as a combination of $\alpha P + (1-\alpha)Q$ and Q. Specifically, denote

$$\begin{aligned} P' &= \alpha P + (1-\alpha)Q \\ Q' &= Q \\ \gamma &= \frac{\beta}{\alpha} \in (0,1) \end{aligned}$$

Then we have $P' \succ Q'$, and by the first part of the proof,

$$P' \succ \gamma P' + (1-\gamma)Q'$$

but

$$\begin{aligned} & \gamma P' + (1-\gamma)Q' \\ = \; & \gamma[\alpha P + (1-\alpha)Q] + (1-\gamma)Q \\ = \; & \frac{\beta}{\alpha}[\alpha P + (1-\alpha)Q] + (1-\frac{\beta}{\alpha})Q \\ = \; & \beta P + (1-\beta)Q \end{aligned}$$

and the conclusion $\alpha P + (1-\alpha)Q = P' \succ \beta P + (1-\beta)Q$ follows.

6. a. Show that, if $x, y, z \in X$ satisfy $x \succ y \succ z$ (where we abuse notation and identify each $x \in X$ with the lottery $P_x \in L$ such that $P_x(x) = 1$), there exists a unique $\alpha = \alpha(x, y, z)$ such that

$$\alpha x + (1 - \alpha)z \sim y$$

To see this, consider the sets

$$A = \{\alpha \in [0, 1]\, |\alpha x + (1 - \alpha)z \succ y\}$$

and

$$B = \{\alpha \in [0, 1]\, |y \succ \alpha x + (1 - \alpha)z\}$$

We know that $0 \in B$ and $1 \in A$, and from Problem (5) we also know that both A and B are contiguous intervals. Obviously, they are disjoint. The question is – can they cover the entire segment $[0, 1]$, or does there have to be something in between?

The answer is given by the continuity axiom. It says that both A and B are open: if α is in A, then, for a small enough ε, $\alpha - \varepsilon$ is also in A. Similarly, if $\alpha \in B$, then, for a small enough ε, $\alpha + \varepsilon \in B$. Together, this implies that A and B cannot cover all of $[0, 1]$, and a point

$$\alpha \in [0, 1] \backslash (A \cup B)$$

has to satisfy

$$\alpha x + (1 - \alpha)z \sim y.$$

To see that this α is unique, it suffices to use Problem (5): the strict preference between x and z implies that no distinct α, β can yield an equivalence

$$\alpha x + (1 - \alpha)z \sim y \sim \beta x + (1 - \beta)z.$$

b. Assume that, for some $x, z \in X$, we have $x \succsim y \succsim z$ for all $y \in X$. Define

$$\begin{aligned} u(y) &= 1 \quad if \quad y \sim x \\ u(y) &= 0 \quad if \quad y \sim z \\ u(y) &= \alpha(x,y,z) \quad if \quad x \succ y \succ z \end{aligned}$$

Explain why maximization of the expectation of u represents $\succsim$. ("Explain" means "sketch the proof", but if you write the complete proof formally that's even better.)

Consider a lottery

$$P = (p_1, x_1; p_2, x_2; ...; p_n, x_n)$$

and since

$$x_1 \sim \alpha(x, x_1, z)x + (1 - \alpha(x, x_1, x))z$$

we can "replace" x_1 in the lottery by $\alpha(x, x_1, z)x + (1 - \alpha(x, x_1, z))z$. More precisely, the independence axiom says that when we "mix" x_1 (with probability $\alpha = p_1$) with

$$\left(\frac{p_2}{1-p_1}, x_2; ...; \frac{p_n}{1-p_1}, x_n \right)$$

we might as well "mix" $\alpha(x, x_1, z)x + (1 - \alpha(x, x_1, z))z$ with the same lottery (and same $\alpha = p_1$). This gives us a lottery that is equivalent to P but does not use x_1. (It uses x, z, though.)

Continuing this way n times, we replace all the other outcomes by x, z. If we then calculate the probability of x in this lottery, we find that it is precisely

$$\sum_{i=1}^{n} p_i \alpha(x, x_i, z) = \sum_{i=1}^{n} p_i u(x_i)$$

that is, the expected utility of the lottery according to the utility function defined above.

5 Chapter 6

1. Explain what's wrong with the claim "Most good chess players are Russian; therefore a Russian is likely to be a good chess player".

As explained in class and in the text, this is a classical case of "ignoring base probabilities", i.e. of confusing the probability of A given B with that of B given A. It's possible that $P(A|B)$ is high while $P(B|A)$ is low.

2. When you sail along the shores of the Mediterranean, it seems that much more of the shoreline has hills and cliffs than one would have imagined. One theory is that God created the Earth with the tourism industry in mind. The other is that this is an instance of biased sampling. Explain why.

   [Hint: Assume that the Earth is unidimensional, and that its surface varies in slope. To be concrete, assume that the surface is made of the segments connecting $((0,0),(90,10))$ and $((90,10),(100,100))$ (where the first coordinate denotes distance and the second – height). Assume that the height of the water is randomly determined according to a uniform distribution over $[0,100]$. Compare the probability of the shore being at a point of a steep slope to the probability you get if you sample a point at random (uniformly) on the distance axis.]

The hint here basically gives the answer. Once you draw the curve, if you select a point at random (with a uniform distribution) on the x axis, the steep slope has probability of 10% of being chosen. If you select a random point on the y axis (again, using a uniform distribution), you get a probability of 90% for the steep slope. Thus, if you look around the Earth from a plane, it seems that mountains are very rare. But if you pour water (presumably a random quantity that generates a uniform distribution over the height of the water surface), you're much more likely to have the water height be at a steep slope.

Similarly, if I spend 11 hours and 50 minutes at home, run for ten minutes to get to my office, spend another 11:50 hours there and runs back, people who see me at a random point on the street might think that I'm running a very hectic lifestyle. But if you randomly sample me over time, you're most likely to conclude that I don't move at all.

> 3. Comment on the claim, "Some of the greatest achievements in economics are due to people who studies mathematics. Therefore, all economists should better study mathematics first".

Again, this is the same issue: it's possible that the probability of mathematical background given achievements in economics is high, but this doesn't mean that the probability of economic achievements given a mathematical background is also high.

> 4. Consider Problem 5 below (even if you do not solve it), and explain how many prejudices in the social domain may result from ignoring base probabilities.

Here I would like you to think of a social prejudice. I rather not promote any. I only suggest that you take one such prejudice, associating, say, an ethnic group with a certain characteristic, and ask yourself perhaps the prejudice is partly driven by ignoring base probabilities.

> 5. Trying to understand why people confuse $P(A|B)$ with $P(B|A)$, it is useful to see that qualitatively, if A makes B more likely, it will also be true that B will make A more likely:

a. Show that, for any two events A, B,

$$\begin{array}{rrcl} & P(A|B) & > & P(A|B^c) \\ iff & P(A|B) & > & P(A) > P(A|B^c) \end{array}$$

iff

$$\begin{array}{rrcl} & P(B|A) & > & P(B|A^c) \\ iff & P(B|A) & > & P(B) > P(B|A^c) \end{array}$$

where A^c is the complement of A. (Assume that all probabilities involved are positive, so that all the conditional probabilities are well-defined.)

b. If the proportion of Russians among the good chess players is higher than their proportion overall in the population, what can be said?

a. To start, consider the first equivalence,

$$\begin{array}{rrcl} & P(A|B) & > & P(A|B^c) \\ iff & P(A|B) & > & P(A) > P(A|B^c). \end{array}$$

The second line clearly implies the first. So let us prove the converse: assume the first line and then derive the second.

Bayes's formula tells us that

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c).$$

Denoting $\beta = P(B)$ we have $P(B^c) = 1 - \beta$ and then

$$P(A) = \beta P(A|B) + (1 - \beta)P(A|B^c)$$

with $\beta \in [0, 1]$. That is, the unconditional probability $P(A)$ is a weighted average (with weights $P(B), P(B^c)$) of the two conditional probabilities $P(A|B)$ and $P(A|B^c)$. The weighted average is necessarily between the two extreme points.

Moreover, if the two are distinct, say, $P(A|B) > P(A|B^c)$, and if these are well defined (that is, $P(B), P(B^c) > 0$) then $0 < \beta < 1$ and $P(A)$ is strictly larger than $P(A|B^c)$ and strictly smaller than $P(A|B)$.

Next we wish to show that if $P(A|B) > P(A|B^c)$ then we can reverse the roles of A and B and get also $P(B|A) > P(B|A^c)$. (Clearly, the last equivalence is the same as the first, with the roles of A and B swapped.)

Assume that the probabilities of intersections of A and B and their complements are given by

$$\begin{array}{ccc} & B & B^c \\ A & p & q \\ A^c & r & s \end{array}$$

so that

$$\begin{aligned} P(A \cap B) &= p; \quad P(A \cap B^c) = q \\ P(A^c \cap B) &= r; \quad P(A^c \cap B^c) = s \end{aligned}$$

with $p + q + r + s = 1$. For simplicity, assume that $p, q, r, s > 0$.

Then

$$\begin{aligned} P(A) &= p + q; \quad P(A^c) = r + s \\ P(B) &= p + r; \quad P(B^c) = q + s \end{aligned}$$

and

$$\begin{aligned} P(A|B) &= \frac{p}{p+r}; \quad P(A|B^c) = \frac{q}{q+s} \\ P(B|A) &= \frac{p}{p+q}; \quad P(B|A^c) = \frac{r}{r+s} \end{aligned}$$

The condition

$$P(A|B) > P(A|B^c)$$

is

$$\frac{p}{p+r} > \frac{q}{q+s}$$

and it is equivalent to

$$p(q+s) > q(p+r)$$

or

$$ps > qr$$

which is equivalent to

$$\begin{aligned} ps + pr &> qr + pr \\ p(r+s) &> r(p+q) \\ \frac{p}{p+q} &> \frac{r}{r+s} \end{aligned}$$

that is, to

$$P(B|A) > P(B|A^c).$$

b. In words, we have found that if A makes B more likely (that is, more likely than B used to be before we knew A, or, equivalently, A makes B more likely than does A^c), the converse is also true: B makes A more likely (than A was before we knew B, or, equivalently, B makes A more likely than does B^c).

In our case, if the proportion of Russians among the good chess players is larger than in the population at large, we can say that

(i) the proportion of Russians among chess players is higher than among the non-chess players;

(ii) the proportion of good chess players among Russians is higher than among non-Russians;

(iii) the proportion of good chess players among Russians is higher than in the population at large.

Importantly, we cannot say anything quantitative that would bring us from $P(A|B)$ to $P(B|A)$ without knowing the ratio $P(B)/P(A)$.

> 6. Consider a regression line relating the height of children to that of their parents. We know that its slope should be in $(0,1)$. Now consider the following generation, and observe that the slope should be again in $(0,1)$. Does it mean that, due to the "regression to the mean", all the population will converge to a single height?

The answer is negative. The regression to the mean is observed when we try to predict a single case, not the average of the population. Indeed, particularly tall parents will have children that are, on average, shorter than they are (but taller than the average in the population), and particularly short parents will have, on average, taller children. These would be the extremes "feeding" the mean. At the same time, we will have the opposite phenomena: parents with average height will have children in both extremes. (In particular, a parent with "tall" genes who happened to have been poorly nourished might be of average height, yet will pass on the "tall" genes.)

Moreover, if you regress the height of the parents on the height of the children you are also likely to find a nice positive correlation – again with the regression to the mean. (Recall that correlation does not imply causation: the parents' height is a cause of the height of the children, and not vice versa, but the correlation goes both ways.) If you were to agree that the children's generation would have a lower variance than the parents' generation, you should also endorse the opposite conclusion...

6 Chapter 7

> 1. In order to determine a unique utility function for each individual, to be used in the summation of utilities across individuals, it was suggested to measure individual's vNM utility functions (for choice under risk), and to set two arbitrary outcomes to given values (shared across individuals). Discuss this proposal.

The proposal is not without merit. Fixing two outcomes that are considered to be of more or less universally agreed upon values makes sense. Of course, nothing is ever "objective". An individual who wishes to commit suicide might prefer death to life, so that we can't even agree on what seems like an obvious ranking. Yet, we can hope that this is exceptional. Moreover, we can take a paternalistic point of view and decide to ignore such preferences even if they do exist, ascribing to the person a preference for life over death, or for more money over less money, independently of what they actually prefer.

There are, however, two other difficulties with this proposal. First, it is not clear that the utility function used for describing behavior under risk is the "right" one for social choice. Assume that one individual is risk averse and the other one is risk neutral. We have to share 1 euro between them. Let us assume that we normalized their utility functions so that they both have $u_i(0) = 0$ and $u_i(1) = 1$. If u_1 is concave and u_2 is linear, the maximization of

$$u_1(x) + u_2(1 - x)$$

will be obtained where $u_1'(x) = 1$. If, for example,

$$u_1(x) = \sqrt{x}$$

we end up giving

$$x = 0.25 > 0.5$$

to individual 1. That is, being risk averse, this individual gets less of the social resources, and it's not obvious that we would like to endorse this.

Finally, once a procedure such as the above is put into place, we should expect individuals to be strategic about it. If one knows that the responses one gives to vNM questionnaires eventually determine social policy, one may choose to provide untruthful reports (say, pretend to be less risk averse than one really is) in order to get a larger share of the pie.

> 2. The "Eurovision" song contest uses a scoring rule, according to which each country ranks the other countries' songs and gives them scores according to this ranking. It has been claimed that the scores given favor standard songs over more innovative ones. Does this claim make sense? Is it more convincing when the score scale is convex or concave?

If the score scale is "convex", say,

$$1, 2, 4, 8, 16, ...$$

it's worthwhile to be half the time at the higher end of the scale and the other half at the lower end, as compared to being around the middle all the time. If you have a choice between a "risky" song, which might be loved by some and abhorred by others, or a less risky one, which is likely not to arouse strong emotions in anyone, you would prefer the riskier song.

By contrast, a concave scale such as

$$5, 9, 12, 14, 15$$

generates the opposite incentives for similar reasons.

> 3. It turns out that, for two particular individuals, Pareto domination defines a complete relation. (That is, for every two distinct alternatives, one Pareto dominates the other.) Assume also that
>
> $$u(X) = \{(u_1(x), u_2(x)) \,|x \in X\}$$

is convex. What can you say about the utility functions of these individuals?

First we argue that, if the set

$$u(X) = \{(u_1(x), u_2(x)) \, | x \in X\}$$

is convex, it has to be a straight line segment (in $\mathbb{R}^2$, where the first coordinate is $u_1(x)$ and the second is $u_2(x)$). To see this, assume that $u(X)$ is not contained in a segment. Connect two points in $u(X)$. Since the latter is not included in the line defined by these two points, you have points off the line. By convexity, you have an entire non-trivial triangle (with positive area) in $u(X)$. But in such a triangle you can find two points that are not ranked by Pareto domination. Hence $u(X)$ is contained in a segment.

If the segment has a negative slope, we will find on it points that are not ranked by Pareto domination again. Hence we conclude that this segment can be parallel to the x axis, or parallel to the y axis, or else it has a finite but positive slope. In all of these cases we obtain the conclusion that the utility function of one individual is a linear function of the other (with the possibility of zero coefficient if the segment is parallel to one of the axes, making one individual indifferent among all alternatives).

4. Assume that individual i has a utility function u_i. For $\alpha = (\alpha_1, ..., \alpha_n)$ with $\alpha_i > 0$, let

$$u_\alpha = \sum_{i=1}^{n} \alpha_i u_i.$$

Show that if x maximizes u_α for some α, it is Pareto efficient.

Assume that x maximizes u_α for some $\alpha > 0$, but suppose, by negation, that x is not Pareto efficient. Then there exists y such that $u_i(y) \geq u_i(x)$ for all i, with a strict inequality for at least one i, say $i = i_0$. Since we assumes that all the coefficients are strictly positive, we know that $\alpha_{i_0} > 0$. This means that

$$u_\alpha(y) > u_\alpha(x)$$

contrary to the assumption that x is a maximizer of u_α.

> 5. Is it true that every Pareto efficient alternative maximizes u_α for some α? (Hint: for $n = 2$, consider the feasible sets
>
> $$X_1 = \{(x_1, x_2) \,|\, \sqrt{x_1} + \sqrt{x_2} \le 1; \quad x_1, x_2 \ge 0\}$$
>
> and
>
> $$X_2 = \left\{(x_1, x_2) \,\middle|\, x_1^2 + x_2^2 \le 1; \quad x_1, x_2 \ge 0\right\}$$
>
> where $u_i = x_i$.)

The answer is negative, as suggested by the second example in the hint. If the feasible set X is convex, the answer would have been almost positive. To be precise, it would have been positive if we allowed some α_i to be zero. Indeed, for the convex feasible set X_2 all points where $x_1, x_2 > 0$ are optimal for a utility function

$$\alpha_1 u_1 + \alpha_2 u_2$$

with $\alpha_1, \alpha_2 > 0$, but for the extreme points $(0, 1)$,$(1, 0)$ you need to set one of the α_i to 0.

However, with a non-convex set such as X_1, none of the Pareto efficient points can be described as a maximizer of a utilitarian function.

> 6. Show that, under approval voting, it makes sense for each individual to approve of her most preferred alternative(s), and not to approve of the least preferred one(s) (assuming that the voter is not indifferent among all alternatives).

Assume that a is my preferred alternative. Suppose that I am about to cast a ballot that approves of some set of alternatives B, which does not contain a. (B may be empty, though we agreed that, under approval voting, an empty set is equivalent to abstention.) Next consider switching from the ballot B to

$B \cup \{a\}$, i.e., adding the most preferred alternative to the set of approved-of alternatives.

Consider two possibilities: (i) the other voters casts ballots that, together with my vote B, makes a among the top alternatives; and (ii) the others' votes, together with my B, does not put a at the top. In case (i), a will certainly be at the top – in fact, it will become the unique top alternatives if it was not the unique one before. Thus, I only stand to gain from adding a: either it used to be the unique one at the top, and then my vote does not change that, or else it singles a as the unique one among several that used to be equally popular.

Next, consider case (ii). Here a was not among the winners of the vote. Adding it to my ballot B might not change anything, or add a to the set of winners. But in the latter case, it reduces the probability of alternatives I like less than a in favor of a, which is my most preferred alternative. Thus, in this case again I can only gain by switching from B to $B \cup \{a\}$.

7 Chapter 8

> 1. Assume that the prisoner's dilemma is player T times between two players. Show that playing D is not a dominant strategy, but that the only Nash equilibria still result in consecutive play of D.

First, let us convince ourselves that playing D at the first stage is no longer dominant. To see this, imagine that the other player's strategy is to respond to your first move as follows: if you play C in the first stage, they play C throughout the rest of the game, but if you play D in the first stage – they play D in the remaining $(T-1)$ stages. If T is large and you are not too impatient, it makes sense to forego the gain in the first period in order to get the higher payoffs guaranteed by the other player playing C after that period.

Second, we wish to convince ourselves that at all Nash equilibria the players play only D. We will do this by "backward induction". First, consider the last stage, $t = T$. In this stage there is no future and playing D is dominant. To be precise, if the players choose strategies that lead to a certain node, their only best response to each other's strategy is to play D (with probability 1) at this last node.

Let us now consider the penultimate stage, $t = T-1$. Can it be the case that a player plays C at this stage? The answer is negative: suppose you decided to play C. Why won't you switch to D, which gives you a strictly higher payoff at stage $t = T-1$? The only reason can be that this switch will be "punished" by the other player reply in the following stage ($t = T$). However, you already know that the other player will play D. Or, more precisely, for there to be a "punishment" threat, it should be the case that if you do stick to the presumed equilibrium strategy (and play C), you will be rewarded by the other player's playing C (or at least C with a positive probability) in the last stage. But we have concluded that any node that can be reached by the equilibrium strategies is one in which the players play D (with probability 1).

In the same manner we continue, and prove, by induction on $k \geq 1$, that at any node that is in stage $T-k, T-k+1, ..., T$ and that is reached (with positive probability) by the equilibrium play, the players play D. Applying the conclusion to $k = T-1$ completes the proof.

> 2. Consider the following story. In a certain village there are n married couples. It is the case that, should one married woman be unfaithful to her husband, all other men are told about it immediately, but not the husband. This fact is commonly known in the village. The law of the land is that, should a husband know that his wife has been unfaithful to him, he must shoot her to death on the same night. But he is not allowed to hurt her unless he knows that for sure.
>
> One day a visitor comes to the village, gets everyone to meet in the central square, and says "There are unfaithful wives in this village". He then leaves.
>
> That night, and the following one, nothing happens. On the third night, shots are heard.
>
> a. How many shots where heard on the third night?
>
> b. What information did the visitor add that the village inhabitants did not have before his visit?

a. There were three shots. The reasoning is as follows. Let k be the number of unfaithful wives. Assume first that $k = 1$, that is, that there is exactly one unfaithful wife in the village. In this case, all men apart from her husband know that "There are unfaithful wives in this village". But the husband doesn't know whether there are ($k = 1$) or there aren't ($k = 0$). Importantly, this husband knows that the other married women are faithful to their respective husbands, because he knows that, had one of them been unfaithful, he would have known about her. But he knows of none, and he knows that he knows of none. So he can conclude that the other women are faithful. Hearing the news that some women are not faithful ($k \geq 1$) proves to him that his wife isn't faithful to him, and he will kill her on the first night.

Next assume that there are exactly two unfaithful wives ($k = 2$). Call then A and B. The husband of each of them knows that there are some ($k \geq 1$) unfaithful wives, because he knows for sure that the other one (not his wife) is unfaithful. That is, A's husband knows that B is unfaithful, but doesn't know whether A is, and B's husband knows that A is unfaithful but doesn't know whether B is. Hence, the husbands of both A and B are not too excited when they hear that "There are unfaithful wives in this village". Each should say to himself, "Well, I don't know about my wife, but the fact that some are unfaithful is not news to me." However, A's husband should also reason as follows: "if my wife, A, were faithful to me, then B would be the only unfaithful wife in the village (that is $k = 1$). In this case, by the reasoning for the case $k = 1$, B's poor husband just learnt that B is unfaithful to him and he'll shoot her tonight." Excited about the prospect of a sensational killing, A's husband goes to sleep. In the morning, he is awakened by the birds chirping rather than by the sound of a shot. And then he has to say to himself, "Oh, my God, B's husband didn't shoot her tonight. This means that he's not sure that she's unfaithful. This means that he had already known that there are some unfaithful wives, i.e., that $k \geq 1$. But, not knowing about B herself, he can only know that my dear wife, A, is unfaithful to me." Equipped with this sad conclusion, A's husband waits until night falls, and then he shoots his wife. By the same reasoning, so does B's husband.

Similarly, you can prove by induction that, if there are exactly k unfaithful wives, then all their husbands will know for sure that their wives are unfaithful on the k-the night, and shoot them on that night. Thus, if the shots were heard on the third night, there were exactly three unfaithful wives in the village.

b. The information added by the visitor was not that there are unfaithful wives. Indeed, with $k \geq 2$ all husbands know that there are some unfaithful wives. The additional piece of information was that "There are unfaithful wives in this village" is common knowledge. That is, by the public declaration the visitor made a fact, that was already known, also commonly known.

To see this more clearly, observe that, if $k = 1$, everyone but one husbands knows the proposition $p =$"There are unfaithful wives in this village". In this case, this husband does learn something from the declaration. If, however, $k = 2$, everyone knows p, but it is not true that everyone knows that everyone knows p. As analyzed before, if the only two unfaithful wives are A and B, then A's husband knows p, and he also knows that all other husbands apart from B's husband know p, but he does not know that B's husband knows p. As far as he knows, his wife, A, may be faithful, and then B's husband wouldn't know whether p is true or not (i.e., would not know whether B is faithful or not). Similarly, if $k = 3$, everyone knows p, and everyone knows that everyone knows that p, but it is not true that (everyone knows that)^3 p. Thus, for any k, there is some hierarchy of knowledge that is missing – and this level is what the visitor adds by his public announcement.

> 3. Consider an extensive for game, and show how a player might falsify common knowledge of rationality (by deviating from the backward induction solution). Show an example in which it may be in the player's best interest to do so.

(It will be helpful to draw a game tree here!) Consider the game in which Player I can choose to go "down" (D) and end the game with payoffs $(11, 5)$, or to play "across" (A). If she plays A, it's Player II's turn. He can choose between playing "down" (d), and end the game with payoffs $(x, 9)$, or "across" (a), in which case it's Player I's move again. In the last stage, Player I has to choose between "down" (δ) – with payoffs $(9, 0)$ – and "across" (α), with payoffs $(10, 10)$.

The backward induction solution is as follows: if we reach the last node, Player I would play across (α) (because $10 > 9$). Given that, in the second node Player II should play across (a) (because he gets 10 by the backward induction solution if he continues, and only 9 if he stops the game). Given this, we can now conclude that at the first node Player I should play "down" (D) (because this guarantees here 11 and the backward induction analysis says that she will get only 10 is she plays across).

However, what should player II think if he finds himself playing? The backward induction solution says that he will not have to play at all. So there is something wrong in the assumptions underlying the backward induction solution. What is it? We don't know, but maybe Player I is not rational? Maybe she's crazy? In this case, can Player II trust her that she will indeed prefer 10 to 9 at the last stage? Maybe she won't, and then Player II will get only 0? So perhaps it is safer for Player II to play down (d), guaranteeing herself 9, rather than taking a lottery with outcomes 10 and 0, with the unknown probability that Player I is rational?

Indeed, if x is rather low, this would make a lot of sense. But what happens if $x = 15$? In this case, this payoff is the best outcome for Player I throughout the game. That is, in this case it is in Player I's interest to saw doubt about her own rationality in Player II's mind. If Player II is sure that Player I is rational, he will play across. But if Player I manages to scare him to think that she's crazy, she will be better off. But then again, maybe precisely because of this Player II will not be scared? Maybe he'll say to himself, "Oh, I know the game she's playing. She is trying to scare me to give her the best payoff for her. But I will not be tricked. I'll play across and I'm sure that when it's her choice in the final node, she'll be rational alright" – ?

It's not clear indeed how players revise their theory of the game (and of the other players' rationality) in such situations. And you can see such example in real life, including political situations in which one may be better off if others think that one is crazy, but pretending to be crazy may not be easy if the motives for doing so are too transparent.

4. Compute the mixed strategies equilibria in games 3,4,5 above.

Consider Game 3 first. Assume that Player I plays R with probability p and

L with probability $(1-p)$. Assume that Player II plays R with probability q and L with probability $(1-q)$. If Player I uses a truly mixed strategy, i.e., if $0<p<1$, it has to be the case that the expected utility she gets from both pure strategies is the same. To see this, observe that the expected utility is linear in p:

$$EU^1\left((p,1-p)\right) = pEU^1\left((1,0)\right) + (1-p)EU^1\left((0,1)\right).$$

If the expected utility from playing R, $EU^1\left((1,0)\right)$ were higher than the expected utility from playing L, $EU^1\left((0,1)\right)$, the only optimal response for Player I would have been $p=1$. Conversely, if $EU^1\left((1,0)\right) < EU^1\left((0,1)\right)$, the only optimal response is $p=0$. Hence the only way that $p \in (0,1)$ can be optimal is if

$$EU^1\left((1,0)\right) = EU^1\left((0,1)\right).$$

Notice that in this case Player I is completely indifferent between playing $(p,1-p)$ and playing $(1,0)$, $(0,1)$ or any other mixed strategy. This may sound a little weird, and, indeed, some people are not completely convinced by the concept of mixed strategies Nash equilibria in games that are not zero-sum (where there exists other justifications of the concept). But let's acknowledge these doubts and move on.

Given that Player II plays $(q,1-q)$, we can compute these expected utilities:

$$\begin{aligned} EU^1\left((1,0)\right) &= q*1+(1-q)*0 = q \\ EU^1\left((0,1)\right) &= q*0+(1-q)*1 = 1-q \end{aligned}$$

and the equation $EU^1\left((1,0)\right) = EU^1\left((0,1)\right)$ means that $q=1-q$ or $q=0.5$. The same calculation applies to Player II, given that Player I plays $(p,1-p)$ and it yields $p=0.5$.

For Game 4 the same type of calculations (with the same notation for p and q, though the names of the pure strategies are different) yield:

$$\begin{aligned} EU^1\left((1,0)\right) &= q*3+(1-q)*0 = 3q \\ EU^1\left((0,1)\right) &= q*0+(1-q)*1 = 1-q \end{aligned}$$

and

$$EU^1\left((1,0)\right) = EU^1\left((0,1)\right).$$

implies

$$\begin{aligned} 3q &= 1-q \\ q &= 0.25 \end{aligned}$$

and, similarly, we also get $p = 0.25$.

In Game 5 we have (again, with the same meaning of p and q)

$$\begin{aligned} EU^1\left((1,0)\right) &= q*2+(1-q)*0 = 2q \\ EU^1\left((0,1)\right) &= q*0+(1-q)*1 = 1-q \end{aligned}$$

and

$$\begin{aligned} 2q &= 1-q \\ q &= \frac{1}{3} \end{aligned}$$

but for Player II we get

$$\begin{aligned} EU^2\left((1,0)\right) &= p*1+(1-p)*0 = p \\ EU^2\left((0,1)\right) &= p*0+(1-p)*2 = 2(1-p) \end{aligned}$$

and

$$\begin{aligned} p &= 2(1-p) \\ p &= \frac{2}{3} \end{aligned}$$

That is, each player chooses the strategy that corresponds to their preferred equilibrium with probability $\frac{2}{3}$.

5. Show that a 2x2 game, in which all payoffs are different, cannot have precisely two Nash equilibria.

Let there be a game

	L	R
T	a, α	b, β
B	c, γ	d, δ

Since all payoffs are different, we may assume without loss of generality that $a > c$. Otherwise, $c > a$, and we can rename the strategies to make a the higher payoff.

Let us now consider b and d. If $b > d$, then strategy T strictly dominates strategy B for Player I. In this case, in each equilibrium player I will play T with probability 1. And then the only equilibrium will be obtained when Player II plays L (with probability 1), if $\alpha > \beta$, or R (with probability 1), if $\alpha < \beta$. That is, if $b > d$ the game has a unique equilibrium in pure strategies and no equilibria in mixed strategies. The number of equilibria is then 1.

Next consider the case in which $b < d$. Before we continue we might simplify the analysis a bit. To be honest, we take this chance to make a theoretical point more than to simplify the analysis.

Recall that the vNM utility functions are given up to multiplication by a positive constant and an addition of a constant. In fact, if we only consider this particular game, we can also add an arbitrary constant to the payoffs of Player I in each column, and an arbitrary constant to the payoffs of Player II in each row. (Convince yourself that such a "shift" of the utility function in a given column for Player I or in a given row for Player II does not change the best-response set: a strategy is a best response for a player after such a shift if and only if it used to be a best response before the shift.)

Hence, we can assume without loss of generality that $c = 0$ (by subtracting c from Player I's payoffs in column L) and that $b = 0$ (by subtracting b from Player I's payoffs in column R) and obtain the game

	L	R
T	a, α	$0, \beta$
B	$0, \gamma$	d, δ

with $a, d > 0$. (Technically speaking, now it is no longer true that "all payoffs are different", but what we care about are the payoffs that can be compared by a given player who considers to switch a strategy. Hence the fact that there are two zeroes above does not change the fact that there are no indifferences when players compare their payoffs given different choices of themselves but the same choice of the other.)

We now turn to consider Player II's payoffs. If $\alpha < \beta$ and $\gamma < \delta$, then R is a strictly dominant strategy for Player II and the unique equilibrium is (B,R). Similarly, if $\alpha > \beta$ and $\gamma > \delta$, then L is a dominant strategy and the unique equilibrium is (T,L). Thus we are left with the interesting case in which Player II does not have a dominant strategy either. This means that either

(i) $\alpha < \beta$ and $\gamma > \delta$

or

(ii) $\alpha > \beta$ and $\gamma < \delta$.

Note that these cases are not symmetric anymore. If you wish to switch the names of the columns now you will change some of the assumptions about Player I's payoffs.

We may still simplify notations a bit by assuming that at least one zero appears among player II's payoffs in each row. We can decide, for instance, that $\beta = 0$ and consider the cases in which α is positive or negative. Or we may choose to work only with non-negative payoffs, and then we'll set each time a different parameter to zero. Let's do this, one case at a time:

In Case (i) we may assume without loss of generality that $\alpha = \delta = 0$ and we get the game

$$\begin{array}{ccc} & \text{L} & \text{R} \\ \text{T} & a, 0 & 0, \beta \\ \text{B} & 0, \gamma & d, 0 \end{array}$$

with $a, d, \beta, \gamma > 0$. In this case there is no pure strategy Nash equilibrium. An equilibrium in mixed strategies $((p, 1-p), (q, 1-q))$ will have to satisfy

$$\begin{aligned} qa &= (1-q)d \\ (1-p)\gamma &= p\beta \end{aligned}$$

that is,

$$\left(\left(\frac{\gamma}{\beta+\gamma}, \frac{\beta}{\beta+\gamma} \right), \left(\frac{d}{a+d}, \frac{a}{a+d} \right) \right)$$

is the unique mixed strategy Nash equilibrium, and the unique Nash equilibrium overall.

In Case (ii) we may assume that $\beta = \gamma = 0$ and the game is as follows:

$$\begin{array}{ccc} & \text{L} & \text{R} \\ \text{T} & a, \alpha & 0, 0 \\ \text{B} & 0, 0 & d, \delta \end{array}$$

with $a, d, \alpha, \delta > 0$.

In this case, both (T,L) and (B,R) are pure strategies Nash equilibria. Are there any mixed ones? If $((p, 1-p), (q, 1-q))$ is a Nash equilibrium in mixed strategies, it will have to satisfy

$$\begin{aligned} qa &= (1-q)d \\ p\alpha &= (1-p)\delta \end{aligned}$$

Indeed,

$$\left(\left(\frac{\delta}{\alpha+\delta}, \frac{\alpha}{\alpha+\delta} \right), \left(\frac{d}{a+d}, \frac{a}{a+d} \right) \right)$$

is a mixed strategy Nash equilibrium. Overall, there are three equilibria in the game.

To conclude, if one of the players has a dominant strategy, the game will have a unique Nash equilibrium, and it will be pure. Otherwise, there might be a unique Nash equilibrium in mixed strategies (if the game is of the type of "matching pennies", or three equilibria of which two are pure (if the game is of the type of a "coordination game", or "battle of the sexes").

> 6. A computer sends a message to another computer, and it is commonly known that it never gets lost and that it takes 60 seconds to arrive. When it arrives, it is common knowledge (between the two computers) that it has, indeed, been sent and arrived. Next, a technological improvement was introduced, and the message can now take any length of time between 1 and 60 seconds. How long after the message was sent will it be commonly known that it has been sent?

Suppose that the message was sent at time 0 (measured in seconds) and arrived at time t, $1 \leq t \leq 60$. At time t, the receiver knows that the message has arrived. Does the sender know that it has arrived? Well, this depends. If $t = 60$, the sender will know that the message has arrived, because 60 seconds is the upper bound on the transmission time. But if t is smaller than 60, this will help the receiver to know it sooner, but not the sender. The sender will have to wait until the 60-th second to know that her message had indeed arrived.

When will the receiver know (for sure) that the sender knows (for sure) that the message has arrived? The receiver knows the previous paragraph, so that he knows that the sender is going to wait 60 second from the time she sent the message until she can surely say that the message has arrived. When was the message sent? The receiver can't know for sure. Getting the message at time t, he has to consider various possibilities: it might have been a quick transmission, sent at $t-1$ and arriving within one second, or a sluggish one, sent at $t-60$ and taking the maximal length of time, 60 seconds. When will the receiver know that the sender knows that the message has been sent? Well, the receiver will have to wait 60 second after transmission time, which is somewhere between $t-60$ and $t-1$. The maximum is obtained at $t-1$. That is, after having received the message, the receiver has to wait another 59 seconds to know for sure that the sender knows for sure that the message has arrived.

When will the sender know that the receiver knows that the sender knows that the message has arrived? Well, she knows the analysis in the previous paragraph. That is, she knows that the receiver has to wait 59 seconds from the time that the message has actually arrived until he (the receiver) knows that she knows that it has arrived. Sending the message at time 0, she has to consider the maximal t, that is $t = 60$, and add to it another 59 seconds, and then, only at $t = 119$, can she say that she knows that he knows that she knows that the message has arrived.

And when will the receiver know that the sender know that the receiver knows that the sender knows that the message has arrived? The receiver has to wait 119 second from the time the message has been sent, which means,

$119 + 59 = 178$ seconds from the time he received it. Taking this into account, the sender knows that she has to wait $178 + 60 = 238$ seconds from the time of transmission until she know that he knows that...

In short, the fact that the message has arrived will never be common knowledge!

8 Chapter 9

1. Discuss the reasons for which equilibria might not be efficient in the following cases:

a. A physician should prescribe tests for a patient

b. A lawyer assesses the probability of success of a legal battle

c. A teacher is hired to teach a child.

All of these cases are examples of principal-agent problems with incomplete information. A physician is an expert hired by the patient (directly or indirectly). The physician knows more than does the patient about the patient's condition, possible treatments, etc. Consider a test that the physician might prescribe, which is very expensive or unpleasant. If she bears no part of the cost, she might be overly cautious and prescribe the test, simply because she would feel more comfortable with the additional information. The patient may prefer to forego the test and avoid the cost or pain involved, but he does not have the information to make this decision. If, however, the physician does bear some of the cost, say, she has a budget for tests, then she has an incentive to save money even if the test is necessary. Again, the patient can't directly check whether the physician's recommendation is the same recommendation he would have arrived at given the information. Thus, in both accounting systems, equilibria may not be Pareto efficient.

Similar problems arise when the lawyer – an expert – has to advise the client whether to pursue a legal battle. If the only costs and benefits involved are monetary, it is possible to agree on a fee that is proportional to the client's outcome, and thus to align the interests of the informed agent (the lawyer) with the uninformed principal (the client). But since there are other costs (such as psychological cost of uncertainty), one may again find that equilibria are inefficient.

Finally, in education we find a double-agent problem. The parent hires the teacher to teach, but both the child and the teacher would prefer to tell each other jokes or to cut the lesson by a few minutes. The parent may condition the teacher's compensation on the child's performance on a certain test, but it's hard to disentangle the child's talent and the teacher's efforts, and to make the compensation proportional to the latter. Again, inefficiency is to be expected.

> 2. The dean has to decide whether to give a department overall budget for its activities, or to split it among several activities such as "conferences", "visitors", and so forth. Discuss pros and cons of the two options.

The argument for an overall budget is the classical argument for free markets. Rather than a central planner, who dictates the details of economic activities, the free market intuition suggests that we decentralize the decision making process. Thus, as a dean you might say, "Who am I to judge what's the best trade-off between inviting visitors over vs. going to conferences abroad? Let the department make these choices. I should trust them that they know best how useful are conferences and which ones to go to, which visitors to invite, etc."

However, this free-market intuition should be qualified. First, there is a problem of incomplete information as in any principal-agent problem: the principal may not know whether the faculty go to a conference in a charming Mediterranean island because it's the most important conference in the field, or because its location is nice. Since the faculty's payoff is not precisely aligned with the school's, it's also not clear whether the right trade-off has been struck between travelling and inviting visitors, and whether the choice of visitors was perfectly objective, etc.

On top of this, there may be problems of externalities involved. For example, inviting visitors may benefit other departments, and this externality may not be "internalized" by the department making the decision.

> 3. Consider the student-course assignment problem. Show that for every n it is possible to have examples in which n is the minimal

> number of students that can find a Pareto-improving re-allocation of courses.

Let there be n students and n courses, denoted $\{a_1, ..., a_n\}$. Consider the preferences given by the following table. Each column designates a student, and the courses in that column are listed from top (most preferred) to bottom (least preferred):

1	2	3	...	n
a_1	a_2	a_3	...	a_n
a_2	a_3	a_4		a_1
a_3	a_4	a_5		a_2
...	...			
a_n	a_1	a_2		a_{n-1}

That is, the preferences of individual i are obtained from those of individual $i-1$ by taking the best alternative in the eyes of $i-1$ and "moving" it to the bottom, without changing the ranking of the other pairs of alternatives.

Now assume that the allocation is such that each individual has her second best choice. That is, 1 has a_2, 2 has a_3, and so on (with a_1 in the hands of individual n). Clearly, there is a Pareto-improving trade by which each gets her most preferred alternative instead of her second-most-preferred. However, no proper subset of the individuals can obtain a Pareto improving trade. For example, assume that individual n is not among the traders. In this case, individual 1 cannot get a_1, which is the only alternative she is willing to trade for what she has, namely, a_2. This means that individual 1 will also not be part of the trade. This, in turn, means that individual 2 cannot be convinced to give up her current holding, a_3, and so on.