6 Why does Wikipedia even work?

Now we from the recommendation problem to three chapters on influence in social networks. We start with forming consensus from conflicting opinions in this chapter before moving onto a collection of influence models in the next two.

Let us compare the four different consensus models we have covered in Chapters 3-6, as visualized in Figure 6.1:

- Google pagerank turns a graph of webpage connections into a single ranked order list according to their pagerank scores.
- Netflix recommendation turns a user-movie rating matrix into many ranked order lists, one list per user based on the predicted movie ratings for that user.
- Amazon rating aggregation turns a vector of rating scores into a single scalar for each product.
- Voting systems turn a set of ranked order lists into a single ranked order list, as we will see in this chapter.

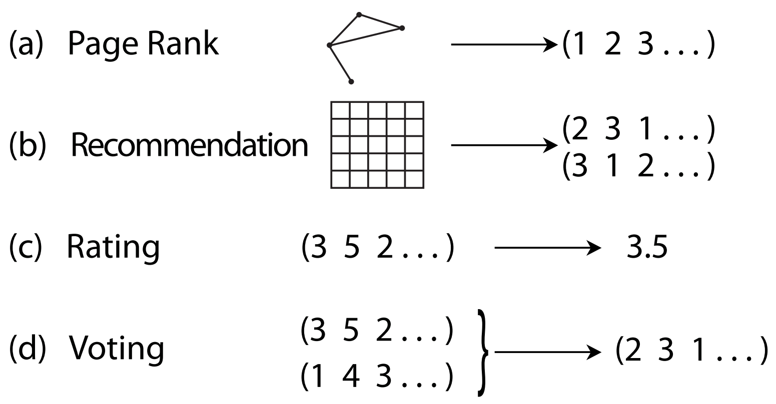

Figure 6.1 Comparison of the four types of consensus formation mechanisms with their inputs and outputs. The first three mechanisms have been covered in the last three chapters, and the last one will be part of this chapter.

6.1 A Short Answer

Crowdsourcing knowledge representation with unpaid and possibly anonymous contributors is very tricky. It faces many challenges for this idea to "work," including two major ones: how to create incentives for people to keep contributing, and how to handle disagreements among the contributors.

Launched in 2001, Wikipedia represented a convergence of three forces that had been gathering momentum: wikis for online collaboration among people, the free and open software movement, and the appearance of online encyclopedias. Within a decade, Wikipedia has generated 3.7 million articles in the US and 19 million articles worldwide. It has become one of the most popular sources of information online. For certain fields, like medicine, the quality of Wikipedia articles are consistently high. And for many fields, if you google a term, a Wikipedia entry will likely come up in the top few search results. It is quite amazing that Wikipedia actually "worked" as much as it did. As we will see in Chapter 11, when people interact with each other, there is often the risk of *tragedy of the commons*. How does Wikipedia turn that into effective collaboration?

Of course, there are also *limitations* to Wikipedia in its capacity as an encyclopedia:

- *Misinformation*: sometimes information on Wikipedia is plainly wrong, especially in articles with a small audience. But Wikipedia provides an effective self-correcting mechanism: it is open to edits from anyone.
- *Mistakes*: there are also honest mistakes, but again, anyone can edit an article, and the edit will stay there as long as no other contributor can present a stronger case otherwise.
- *Missing information*: no encyclopedia can be truly *complete* to everyone's liking, not even the largest encyclopedia in history.

There have been some high profile cases of limitation and abuse. Still, Wikipedia stands as a major success of online collaboration.

There had been other efforts aimed at creating free, online, open encyclopedia before, and the success of Wikipedia is often attributed to a "good faith collaboration" environment within the Wikipedia contributor community. If we count the number of pairwise links in a fully connected graph with n nodes, we have on the order of n^2 such links. But if we examine the number of opinion configurations, we have 2^n possibilities if each person has 2 choices. This n^2 vs. 2^n tension examplifies the positive and the negative sides of the **network effect**. Converging on one of these configurations is difficult, and Wikipedia mostly follows the principle of "rough consensus". The process of reaching a rough consensus can be understood from voting theory (even though it does not explicitly involve voting by providing a ranked order list) and from bargaining theory.

Wikipedia is free, open, dynamic, interactive, and extensively linked. There are natural pros and cons associated with such a model of encyclopedia that complements other forms of encyclopedia. Let us consider three distinct features of Wikipedia:

- It is free. How can people be motivated to contribute? Incentives do not have to be financial; the ability to influence others is a reward in its own right to most people. This requires the Wikipedia audience to be very large.
- Anyone can write or add to an article, including non-experts, anonymous writers, and people with conflicts of interest. The key is *check and balance.* Precisely because anyone can contribute, Wikipedia has a large body of writers who check others' writing frequently through a mechanism for debates and updates. Sometimes, however, a contributor or an IP address may be blocked if it is detected as a regular source of deliberate misinformation.
- Any subject may be contributed, including controversial ones. Sometimes, however, certain articles can be "protected" from too frequent edits to give time for the community of contributors to debate. How does Wikipedia avoid unbalanced treatment or trivial subjects? It turns out that there are Policies and Guidelines, and there are conflict resolution by editors.

The first and second features above provide Wikipedia with a strong, positive networking effect: a larger audience leads to more contributors, which in turn leads to more audience, provided that the quality of contributions are kept high. This brings us to the third feature above.

In general, how does Wikipedia enforce quality and resolve conflicting contributions? In addressing this question, we also bear in mind the obvious fact that Wikipedia is not a sovereign state with the power of a government. So issues such as voting, decisioning, and free speech do not have the same context.

To start with, there are three core Policies on Wikipedia to help ensure reliability and neutrality as much as possible:

- *Verifiability (V)*: each key point or data in an article must be externally verifiable, with link to the primary source for verification by readers.
- *No Original Research (NOR)*: this is to prevent people from using Wikipedia as a publication venue of their new results.
- *Neutral Point of View (NPOV)*: the basic rule is that a reader must not be able to tell the bias of the author in reading through a Wikipedia article. It is particularly important for controversial topics, but also the most difficult to use exactly in those cases, *e.g.*, contentious political, social, and religious topics. Unlike the above two policies, it is harder to enforce this one since "neutrality" is subjective.

Wikipedia also installs several mechanisms for debates and updates. One is the use of the history page and the talk page, available for public view through tags on top of each article's page. All previous versions and all the changes made are recorded too.

There is also a reputation system for contributors, similar to the reviewer rating system in Amazon. Each article can be rated on a 1-6 grade scale. For those who do not reveal their names, it is a reputation system of IP addresses of the devices from which contributions are sent. In addition, links across article pages are analyzed similar to Google pagerank's reputation analysis for ranking pages in each search.

But perhaps the ultimate mechanism still boils down to people negotiating. Depending on the stage of the article, expert and non-expert contributors may join the discussion. There is a hierarchy of Wikipedia communities, and debates among contributors who cannot come to an agreement will be put forward to a group of the most experienced editors. This committee of editors acts like a jury, listening to the various sides of the debate, and then tries to decide by "rough consensus."

How do we model the process of reaching a "rough consensus" through "goodwill collaboration?" Not easy. We will see in this chapter two underlying theories: bargaining theory and voting theory. But much translation is needed to connect either of these theories to the actual practice of Wikipedia.

- In a bargaining model, the the contributors need to reach a compromise, otherwise there would be no agreement. Each contributor's utility function, and the default position in the case of no agreement, need to reflect the goodwill typically observed in Wikipedia collaboration.
- In a voting model, each contributor has some partially ordered list of preferences, and a *threshold* on how far the group decision can be away from her own preferences before she vetoes the group decision and thereby preventing the consensus from being reached. (Sometimes a decision is actually carried out by explicit votes in the arbitration committee. And the committee members are also elected through a voting system.) Dealing with partial ordering, quantifying the distance between two ordered lists, and modeling the veto decision threshold are still much under-explored in the study of group interaction dynamics.

In contrast to coordination through well-defined pricing feedback signals that we will see in several later chapters, coordination though bargaining or voting is much harder to model. In the next section, we will present the basics of the rich theory of voting and social choice. Then in Advanced Material, we will briefly discuss the mathematical language for bargaining and cooperative games.

6.2 A Long Answer

Wikipedia consensus formation illustrates important issues in the general case of reaching consensus among a group of individuals that is binding for everyone, *i.e.*, how to sign a social contract with a binding force. This is different from presenting the rank ordered list for each individual to evaluate individually (like Netflix recommendation). It is different from coordinating individual actions through pricing (like web auction and Internet congestion control). Ranking preferences is also different from ranking objects (like webpages or products).

Voting is obviously essential for elections of those who will make and execute binding laws to be imposed on those casting the votes. It is also useful for many other contexts, from talent competitions to jury decisions. It is a key component of the **social choice theory**, and how individual preferences are collected and summarized.

Voting theory studies how to aggregate vectors, where the vectors are collections of individual preferences. We will see an axiomatic treatment of voting methods in this section. Later in Chapter 20, we will see another axiomatic treatment of scalarizing a vector of resource allocations. In both cases, turning a not-well-ordered input into an ordered output must satisfy certain intuitive properties, and it is often tricky to accomplish that.

6.2.1 Major types of voting

A **voting system** is a function that maps a set of voters' preferences, called a **profile**, to an **outcome**. There are N voters and M candidates. A profile in this chapter is a collection of ranked order lists, one list per voter that lists all the candidates. An outcome is a single ranked order list.

The requirement of complete ranked order lists as inputs can be too stringent. When there are many candidates, often a coarser granularity is used. For example, Wikipedia's arbitration committee members are elected by a binary voting system. Each voter divide the list of candidates into just two parts: those she votes "for" and those she votes "against". Then the percentage of "for" votes, out of all the votes received by each candidate, is calculated to rank order the candidates. The tradeoff between user-interface simplicity and voting result's consistency and completeness is something interesting to explore but does not have as many results as voting systems with complete ranked order input lists.

A voting system *aggregates* N lists into 1 list, like squeezing a 3 dimensional ball into a 2 dimensional "paste". Naturally, some information in the original set of N lists will be lost after this mapping, and that sometimes leads to results not conforming to our (flawed) intuition. We will focus on three commonly used voting systems to illustrate key points.

Perhaps the most popular voting system is **plurality voting**. We simply count the number of voters who put a candidate j in the first position in their lists. Call these numbers $\{V_j\}$, and the candidate j with the largest V_j wins and is put on the first position of the list in the outcome. Put the candidate with the second largest V_j in the second position, and so on. To simplify the discussion, we will assume no ties. There is also a variant called the **Kemeny rule**, where we count how many "least-liked" votes a candidate receives, and rank in reverse order of those numbers. There are other voting systems that try to determine the least objectionable candidate to help reach consensus. Plurality voting sounds reasonable and is often practiced, but there are many profiles that lead to counter-intuitive results.

A generalization of plurality voting is **positional voting**. Looking at a voter's list, we assign some numbers to each candidate based on its position in the list. The most famous, and the only "right" positional voting that avoids fundamental dilemmas in voting, is the **Borda count**. It is named after the French mathematician in the 18th century who initiated the scientific study of voting systems. By Borda count, the first position candidate in each list gets $M - 1$ points, the second position one gets $M - 2$ points, and so on, and the last position one gets 0 point, since being the last one in a list that must be complete carries no information about a voter's preference at all. Then, across all the lists, the candidate are ranked based on their total points across all the voters.

Yet another method is **Condorcet voting**, named after another French mathematician who founded the research field of voting study shortly after Borda. It is an aggregation of binary results from pairwise comparisons. All voting paradoxes must have at least three candidates. When $M = 2$, the result is always clear-cut, since each voter's preference can be characterized by one number, and aggregating scalars is unambiguous. So how about we look at each possible pair of candidates (A, B), and see how many voters think one is better than the other? This unambiguously decides the winner out of that pair: if more voters think A is better than B, denoted as A>B, then A is placed higher than B in the aggregated ranked order list. Now, if the pairwise decisions generate a consistent ranked order list, we are done. But it may not be able to, for example, when the three individual input lists for three candidates are: A>B>C, B>C>A, and C>A>B. We can see that pairwise comparison between A and B is A>B, similarly B>C and C>A. But that is *cyclic*, and thus logically inconsistent. There is no Condorcet voting output that is self consistent in this case.

6.2.2 A counter-intuitive example

Suppose the editors of Wikipedia need to vote on a contentious line in an article on "ice cream" about which flavor is the best: chocolate (C), vanilla (V), or strawberry (S); with $M = 3$ candidates and $N = 9$ voters. There are 6 positional possibilities for 3 candidates, and it turns out that half of these receive zero votes, while the other three possibilities of ranked-order receive the following votes:

- C V S: 4 votes
- S V C: 3 votes
- V S C: 2 votes

What should the aggregated ranked order list look like?

By plurality vote, the aggregation is clear: [C S V]. But something does not sound right. Here, those who favor strawberry over chocolate outnumbers those who favor chocolate over strawberry. So how could [C S V] be right?

Well, let us look at Condorcet voting then.

- C or V? Pairwise comparison shows V wins.
- S or V? Pairwise comparison shows V wins again.
- C or S? Pairwise comparison shows S wins.

The above 3 pairwise comparisons are all we need in this case, and aggregation is clear: V wins over both C and S, so it should come out first. Then S wins over C. So the outcome is clear: [V S C]. But wait, this is *exactly* the opposite of the plurality vote's outcome.

How about the Borda count? V gets 11 points, C 8 points, and S 8 points too. So V winns and C and S tie.

Three voting methods gave three different results. Weird. You may object: "But this profile input is synthesized artificially. Real world ones will not be like this."

Well, first of all, this synthesized input is indeed designed to highlight that our intuition of what constitutes an accurate aggregation of individual preferences is incomplete at best.

Second, there are many more paradoxes like this. In fact, we will go through a method that can generate as many paradoxes as we like. This is not an isolated incident.

Third, how do you define "real world cases"? Maybe through some intuitively correct statements that we will take as true to start the logical reasoning process. We call those **axioms**. But some seemingly innocent axioms are simply not compatible with each other. This is the fundamental, negative result of **Arrow's Impossibility Theorem** in social choice theory.

6.2.3 Arrow's impossibility results

Let us look at the following five statements that sound very reasonable about any voting system, *i.e.*, axioms that any voting system must satisfy. Two of them concern a basic property called **transitivity**, a logical self-consistency requirement: if there are three candidates (A, B, C), and in a list A is more preferred than B, and B is more preferred than C, then A is more preferred than C. Symbolically, we can write this as A>B>C ⇒ A>C. Had this not been true, we would have a cyclic, thus inconsistent preference: A>B>C>A.

Now the five axioms proposed by Arrow:

1. Each input list (in the profile) is complete and transitive.
2. The output list (in the outcome) is complete and transitive.
3. The output list cannot just be the same as one input list no matter what the other input lists are.
4. (Pareto Principle) If all input lists prefer candidate A over candidate B, the output list must do so too.
5. (IIA Principle) If between a given pair of candidates (A,B), each input list's preference does not change, then even if their preferences involving other

candidates change, the output list's preference based on A and B does not change.

The last statement above is called **Independence of Irrelevant Alternatives** (IIA), or pairwise independence. As we will see, these alternatives are actually not irrelevant after all.

You might think there should be a lot of voting systems that satisfy all the five axioms above. Actually, as soon as we have $M = 3$ candidates or more, there are *none*. If the surprise factor is a measure of a fundamental result's elegance, this impossibility theorem by Arrow in his Ph.D. thesis in 1950 is among the most elegant ones we will see.

How could that be? Something is wrong with the axioms. Some of them are not as innocent as they might seem to be at first glance. The first two axioms are about logical consistency, so we have to keep them. The third one is the underlying assumption of social choice, without which aggregation becomes trivial. So it must be either the Pareto axiom or the IIA axiom.

Usually in an axiomatic system, the axiom that takes the longest to describe is the first suspect for undesirable outcomes of the axiomatic system. And IIA looks suspicious. Actually, how A and B compare with each other in the *aggregate output list* should depend on other options, like candidate C's ranking in the *individual input lists*. To assume otherwise actually opens the possibility that transitive inputs can lead to cyclic output. When a voter compares A and B, there may be a C in between or not, and that in turn determines if the outcome is transitive or not. IIA prohibits the voting system from differentiating between those input lists that lead to only transitive outputs and those that may lead to cyclic outputs.

This will be clearly demonstrated in the next section, where we will see that a group of input rank lists, each transitive, can be just the same as a group of input rank lists where some of them are cyclic. Clearly, when an input is cyclic, the output may not be transitive. In the language of axiomatic construction, if axiom 5 can block axiom 2, no wonder this set of axioms is not self-consistent. The negative result is really a positive highlight on the importance of maintaining logical consistency and the flawed intuition in IIA.

In hindsight, Arrow's impossibility theorem states that when it comes to ranking three or more candidates, pairwise comparisons are inadequate. Then the next question naturally is: what additional information do we need? It turns out that scale, rather than just relative order, will lead to a "possibility result".

6.2.4 Possibility results

What is the *true intention* of the voters? Well, the answer is actually obvious: the entire profile *itself* is the true intention. Voting can be no more universally reflective of the "true intent" of voters than two points, say (1,4) and (3,2), on a 2D plane be compared to decide which is bigger.

Voting in political systems, despite the occurrence of counter-intuitive results stemming from our flawed intuition, is a universal right that provides the basic bridge between individuals and the aggregate, an effective means to provide check and balance against absolute power, and the foundation of consent from the governed to the government. No voting system is perfect, but it is better than the alternative of no voting. Moreover, some voting systems *can* achieve a possibility result.

For example, by replacing IIA with the **Intensity form of IIA** (IIIA), there are voting systems that can satisfy all the axioms. What is IIIA? When we write A>B, we now also have to write down the number of other candidates that sit in between A and B, this is the *intensity*. If none, intensity is zero. IIIA then states that, in the outcome list, the ranking of a pair of candidates depends only on the pairwise comparison *and* the intensity.

While the original 5 axioms by Arrow are not compatible, it turns out that the modified set of axioms, with IIA replaced by IIIA, is: the Borda count is a voting system that satisfies all five axioms now. This stems from a key feature of Borda count: the point spread between two adjacent positions in a rank list is the same no matter which two positions we are looking at. In essence, we need to *count* (the gaps between candidates) rather than just *order* (the candidates).

We have not, however, touched on the subject of manipulation, collusion, or strategic thinking, based on information or estimates about others' votes. For example, Borda count can be easily manipulated if people do not vote according to their true rank ordered list. In Chapter 7, we will look at information cascade as a particular example of influencing people's decision.

6.3 Examples

6.3.1 Sen's result

Another fundamental impossibility theorem was developed by Sen in the 1970s. This time, it turns out the following four axioms are incompatible:

1. Each input list is complete and transitive.
2. The output list is complete and transitive.
3. If all input lists prefer candidate A over candidate B, the output list must too.
4. There are at least two decisive voters.

The first three are similar to what are in Arrow's axioms, and the last one concerns a **decisive voter**: a voter who can decide (at least) one pair of candidates' relative ranking for the whole group of voters, *i.e.*, other voters' preferences do not matter for this pairwise comparison.

Just like Arrow's impossibility theorem, the problematic axiom, in this case axiom 4, precludes the insistency of transitivity of the output list. What axiom 4 implies is actually the following: one voter can impose strong negative externality on all other voters. This is illustrated next.

6.3.2 Constructing any counter examples you want

All examples of Sen's result can be constructed following a procedure illustrated in a small example.

Suppose there are $N = 5$ candidates and $M = 3$ voters. Voter 1 is the decisive voter for (A,B) pairwise comparison, voter 2 for (C,D) pair, and voter 3 for (E,A) pair. These will be marked in bold in tables that follow. We will show how decisive voters can preclude transitivity in the output list.

Let us start with a cyclic ranking for every voter: A>B>C>D>E>A, as shown in Table 6.1.

Table 6.1 Step 1 of constructing examples showing inconsistency of Sen's axioms. Each row represents the draft preferences of a voter. The columns contain a subset of the pairwise comparisons between 5 candidates. Pairwise preferences in bold indicate that they come from decisive voters.

	A,B	B,C	C,D	D,E	E,A
1	**A**>**B**	B>C	C>D	D>E	E>A
2	A>B	B>C	**C**>**D**	D>E	E>A
3	A>B	B>C	C>D	D>E	**E**>**A**

We do not want input rankings to be cyclic, so we need to flip the pairwise comparison order at least at one spot for each voter. But we are guaranteed to find, for each voter, two spots where flipping the order above will not change the outcome. Those are exactly the two pairwise comparisons where some other voter is the decisive voter. So flip the order in those two spots and you are guaranteed a transitive ranking by each voter. The resulting output, however, remains cyclic, as shown in Table 6.2.

This example not only demonstrates how easy it is to generate examples illustrating Sen's negative result, but also that each decisive voter is actually destroying the knowledge of the voting system on whether transitivity is still maintained. If a decisive voter ranks A>B, and another voter ranks not just B>A, but also B> k other candidates>A, then we say the decisive voter imposes a k *strong negative externality* to that voter. In cyclic ranking in Sen's system, each voter suffers strong negative externality from some decisive voter.

This example highlights again the importance of keeping track of the *position* of candidates in the *overall* ranking list by each voter, again motivating the Borda count. We simply cannot consolidate ranking lists by extracting some portion of each list in isolation. "Sensible voting" is still possible if we avoid that compression of voters' intentions.

Table 6.2 Step 2 of constructing examples showing the inconsistency of Sen's axioms. Pairwise preferences in quotation marks are those are flipped from the draft version in Table 6.1 to turn input rankings transitive without destroying the cyclic nature of the outcome ranking.

	A,B	B,C	C,D	D,E	E,A
1	**A**>**B**	B>C	"D>C"	D>E	"A>E"
2	"B>A"	B>C	**C**>**D**	D>E	"A>E"
3	"B>A"	B>C	"D>C"	D>E	**E**>**A**
Outcome	A>B	B>C	C>D	D>E	E>A

6.3.3 Connection to prisoner's dilemma

In fact, the prisoner's dilemma we saw back in Chapter 1 is a special case of Sen's negative result, and similar dilemmas can be readily constructed now. Recall that there are four possibilities: both prisoners 1 and 2 do not confess (A), 1 confesses and 2 does not (B), 1 does not confess and 2 does (C), and both confess (D).

Table 6.3 Prisoner's dilemma as a special case of Sen's impossibility result. Think of the two prisoners as two voters, and the four possible outcomes as four candidates. Four pairs of candidates are compared since they are the four individual actions afforded to the two prisoners. Those pairwise preferences that do not matter are marked in "-" since the other prisoner is the decisive voter for that pair. An additional pairwise comparison is obvious: both do not confess (candidate A) is better than both confess (candidate D).

	A,B	B,D	A,D	C,D	A,C
Prisoner 1	**B**>**A**	-	A>D	**D**>**C**	-
Prisoner 2	-	**D**>**B**	A>D	-	**C**>**A**
Outcome	B>A	D>B	A>D	D>C	C>A

Prisoner 1 is the decisive voter on (A,B) pair and (C,D) pair. Prisoner 2 is the decisive voter on (A,C) pair and (B,D) pair. Together with the obvious consensus A>D, we have an outcome that contains two cyclic rankings: D>B>A>D and A>D>C>A, as shown in Figure 6.2. This presence of cyclic output rankings in the outcome is another angle to understand the rise of the socially-suboptimal equilibrium D as the Nash equilibrium.

6.4 Advanced Material

The second conflict resolution mechanism, in a Wikipedia editorial decision as well as in many other contexts, is **bargaining**. Each bargaining party has a selfish

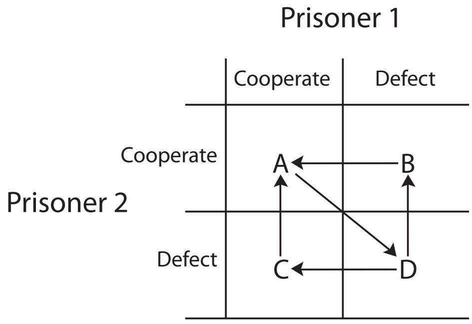

Figure 6.2 Prisoner's dilemma produces two cycles of ranking orders: D>B>A>D and A>D>C>A. This leads to a special case of Sen's impossibility result.

motivation and yet all the parties want to achieve some kind of agreement. If no agreement is achieved, then each party goes back to its own **disagreement point**. This interaction is studied in **cooperative game theory**.

We will first follow the more intuitive approach by Rubenstein in the 1980s before turning to Nash's axiomatic approach in his Ph.D. thesis in 1950, almost at the same time as Arrow's thesis. We will see an IIA style axiom too. But this time it is a possibility theorem that followed, in fact, a unique function modeling bargaining that satisfies all the axioms.

6.4.1 Bargaining: Interactive offers

Suppose there are two people, A and B, bargaining over how to divide a cake of size 1. This cake-cutting problem will be picked up again in Chapter 20 in our study of fairness, with extensions like "one cuts, the other chooses."

For now, consider the following procedure. At the end of each of the discrete timeslots with duration T, each person takes a turn to offer the other person how to share the cake. It is essentially a number $x_1 \in [0,1]$ for A and $x_2 = 1 - x_1$ for B. This iterative procedure starts with the initial offer at time 0 from A to B. If the offer is accepted, an agreement is reached. If it is rejected, the other person makes another offer at the next timeslot.

But wait a minute. This bargaining process can go on forever. Why would either person be motivated to accept an offer (other than one that gives her the whole cake)? There must be a price to pay for disagreeing. In Rubenstein's model, the price to pay is *time*. If an agreement is reached at the kth iteration, a person's payoff is

$$u_i = x_i e^{-r_i k T}, \quad i = 1, 2,$$

where r_i is a positive number capturing "the tolerance to wait and keep bargaining." So the payoff depends on both the deal itself (x_i) and when you seal the deal (k), with the second dependence sensitive to each person's bargaining power: the one with a larger r_i has more to lose by hanging on to keep bargaining for and rejecting offers.

We will not go into the details of the equilibrium properties of this procedure. But it is intuitively clear that if waiting for the next round of negotiation gives me the same payoff as accepting this round's offer, I might as well accept the offer. Indeed, it can be shown that the equilibrium offers (x_1^*, x_2^*) from each side satisfy the following two equations simultaneously:

$$
\begin{aligned}
1 - x_1^* &= x_2^* e^{-r_2 T} \\
1 - x_2^* &= x_1^* e^{-r_1 T}.
\end{aligned}
$$

There is a unique solution to the above pair of equations:

$$
\begin{aligned}
x_1^* &= \frac{1 - e^{-r_2 T}}{1 - e^{-(r_1 + r_2)T}} \\
x_2^* &= \frac{1 - e^{-r_1 T}}{1 - e^{-(r_1 + r_2)T}}.
\end{aligned}
$$

As the bargaining rounds get more efficient, *i.e.*, $T \to 0$, the exponential penalty of disagreement becomes linear: $\exp(-r_i T) \to 1 - r_i T$, and the solution simplifies to the following approximation for small T:

$$
\begin{aligned}
x_1^* &= \frac{r_2}{r_1 + r_2} \\
x_2^* &= \frac{r_1}{r_1 + r_2}.
\end{aligned}
$$

This proportional allocation makes sense: a bigger r_2 means a weaker hand of B, thus a bigger share to A at equilibrium.

6.4.2 Bargaining: Nash bargaining solution

Iterative bargaining is just one mechanism of bargaining. The model can be made agnostic to the mechanism chosen.

Let the payoff function U_i map from the space of allocation $[0, 1]$ to some real number. Assume these are "nice functions:" strictly increasing and concave. If no agreement is reached, a disagreement point will be in effect: (d_1, d_2). Assume disagreement is at least as attractive as accepting the worst agreement: $d_i \geq U_i(0)$, $\forall i$.

The set of possible agreement points is

$$
\mathcal{X} = \{(x_1, x_2) : x_1 \in [0, 1], x_2 = 1 - x_1\}.
$$

The set of possible utility pairs is

$$
\mathcal{U} = \{(u_1, u_2) : \text{there is some } (x_1, x_2) \in \mathcal{X} \text{ such that } U_1(x_1) = u_1, U_2(x_2) = u_2\}.
$$

Nash showed that the following four reasonable statements about a payoff point $\mathbf{u}^* = (u_1^*, u_2^*)$ can be taken as axioms that lead to a unique and useful solution:

1. (Symmetry) If two players, A and B, are identical ($d_1 = d_2$ and U_1 is the same as U_2), the payoffs received are the same too: $u_1^* = u_2^*$.
2. (Affine Invariance) If utility functions or disagreement points are scaled and shifted, the resulting $\mathbf{u}^*$ is scaled and shifted in the same way.
3. (Pareto Efficiency) There cannot be a strictly better payoff pair than $\mathbf{u}^*$.
4. (IIA) Suppose A and B agree on a point $\mathbf{x}^*$ that lead to $\mathbf{u}^*$ in $\mathcal{U}$. Then if in a new bargaining problem, the set of possible utility pairs is a strict subset of $\mathcal{U}$, and $\mathbf{u}^*$ is still in this subset, then the new bargaining's payoffs remain the same.

The first axiom on symmetry is most intuitive. The second one on affine invariance says changing your unit of payoff accounting should not change the bargaining result. The third one on Pareto effiency prevents clearly inferior allocation. The fourth one on IIA is again the most controversial one. But at least it does not preclude other axioms here, and indeed the above set of axioms are consistent.

Nash proved that there is one and only one solution that satisfies the above axioms, and it is the solution to the following maximization problem, which maximizes the product of the gains (over the disagreement point) by both A and B, over the set of payoffs that is feasible and no worse than the disagreement point itself:

$$\begin{array}{ll} \text{maximize} & (u_1 - d_1)(u_2 - d_2) \\ \text{subject to} & (u_1, u_2) \in \mathcal{U} \\ & u_1 \geq d_1 \\ & u_2 \geq d_2 \\ \text{variables} & u_1, u_2 \end{array} \tag{6.1}$$

This solution (u_1^*, u_2^*) is called the **Nash Bargaining Solution** (NBS).

Obviously there is a tradeoff between possible u_1 and possible u_2. If A gets payoff u_1, what is the payoff for B? Using the above definitions, it is:

$$u_2 = g(u_1) = U_2(1 - U_1^{-1}(u_1)).$$

We just defined a mapping, g, from player 1's payoff value to player 2's. This allows us to draw a g as a curve in the payoff plane. It is a similar visualization to the SIR feasibility region in Chapter 1.

If this function g is differentiable, we can differentiate the objective

$$(u_1 - d_1)(g(u_1) - d_2)$$

with respect to u_1, set it to zero, and obtain:

$$(u_1 - d_1)g'(u_1) + (g(u_1), d_2) = 0.$$

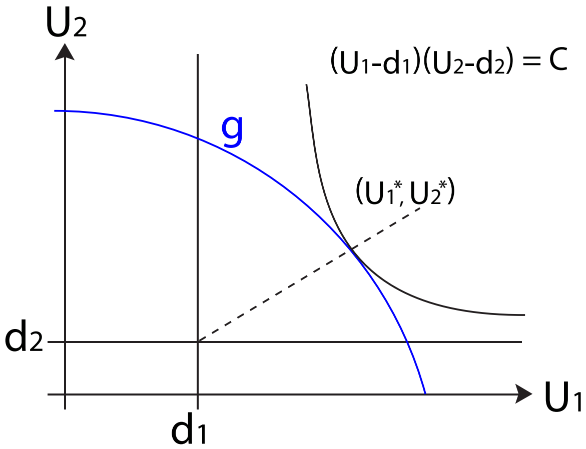

Figure 6.3 Nash bargaining solution illustrated on the two-player payoff plane. It is the intersection of the g curve, which captures the feasible tradeoff between the two players' payoffs, and the straight line normal to g and originating from the disagreement point (d_1, d_2).

Since $g(u_1) = u_2$, we have

$$g'(u_1) = -\frac{u_2 - d_2}{u_1 - d_1}.$$

This has a geometric interpretation, as illustrated in Figure 6.3. NBS (u_1^*, u_2^*) is the unique point on the graph of g where the line from the disagreement point (d_1, d_2) intersects g perpendicular to the slope of g. This clearly illustrates that the bigger the ratio d_1/d_2, *i.e.*, the bigger A's bargaining power relative to B's, the more favorable will u_1 be relative to u_2 at NBS.

For example, if both payoff functions are linear (which actually violates the strictly increasing property we assumed of U_i), then the cake cutting NBS has a simple solution: each person first takes the disagreement point allocation away, and then evenly splits the rest of the cake.

$$x_1^* = d_1 + 0.5(1 - d_1 - d_2)$$
$$x_2^* = d_2 + 0.5(1 - d_1 - d_2).$$

There is another way to model bargaining power: turn the objective function to

$$(u_1 - d_1)^\theta (u_2 - d_2)^{1-\theta},$$

where θ and $1-\theta$, for $\theta \in [0,1]$, are the normalized bargaining power exponents for A and B, respectively. There is an interesting connection between the iterative bargaining solution and this axiomatically developed NBS. In the limit of $T \to 0$, the iterative bargaining solution is the same as the NBS solution with asymmetric bargaining power ($\theta = r_2/r_1 + r_2, 1 - \theta = r_1/r_1 + r_2$) and disagreement point $(0, 0)$.

Further Reading

There is very little mature work on mathematically modeling Wikipedia, but a rich research literature on both voting theory and bargaining theory.

1. A comprehensive survey of the features in Wikipedia, including the policies, guidelines, and editorial procedure, can be found in the following book.

[AMY08] P. Ayers, C. Matthews, and B. Yates, *How Wikipedia Works*, No Starch Press, 2008.

2. Arrow's impossibility theorem was part of Arrow's Ph.D. dissertation, and originally published in 1950 and then in his book in 1951:

[Arr51] K. Arrow, *Social Choice and Individual Values*, Yale University Press, 1951.

3. One of the foremost researchers in voting theory today is Saari, who published several books interpreting and overcoming the negative results of Arrow and of Sen. Our treatment of IIIA, construction of examples of Sen's results, and connection to Prisoner's dilemma, all follow Saari's books. The following one is an accessible and rigorous survey:

[Saa06] D. Saari, *Dethroning Dictators, Demystifying Voting Paradoxes*, 2006.

4. Nash bargaining solution was axiomatically constructed as part of Nash's Ph.D. dissertation, and originally published in the following paper in 1950:

[Nas50] J. Nash, "The bargaining problem," *Econometrica*, vo. 18, pp. 155-162, 1950.

5. Among the many books devoted to bargaining theory since then is a concise and rigorous survey in the following book:

[Mut99] A. Muthoo, *Bargaining Theory with Applications*, Cambridge University Press, 1999.

Problems

6.1 *Differences between Borda count, Condorcet voting and plurality voting* ⋆

Consider an election which consists of 31 voters and 3 candidates A, B, C with the profile summarized as follows:

list	voters
$C > A > B$	9
$A > B > C$	8
$B > C > A$	7
$B > A > C$	5
$C > B > A$	2
$A > C > B$	0

What is the voting result by (a) Plurality voting (b) Condorcet voting (c) Borda count?

6.2 *List's list* $\star$

A three-member faculty committee must determine whether a student should be advanced to Ph.D candidacy or not by the student's performance on both the oral and written exams. The following table summarizes the evaluation result of each faculty member:

Professor	Written	Oral
A	Pass	Pass
B	Fail	Pass
C	Pass	Fail

(a) Suppose the student's advancement is determined by a majority vote of all the faculty members, and a professor will agree on the advancement if and only if the student passes both the oral and written exams. Will the committee agree on advancement?

(b) Suppose the student's advancement is determined by whether she passes both the oral and written exams. Whether the student passes an exam or not is determined by a majority vote of the faculty members. Will the committee agree on advancement?

6.3 *Anscombe's paradox* $\star\star$

Suppose there are three issues where a "yes" or "no" vote indicates a voter's support or disapproval. There are two coalitions of voters, the majority coalitions $\mathcal{A} = \{A_1, A_2, A_3\}$ and the minority coalitions $\mathcal{B} = \{B_1, B_2\}$. The profile is summarized as follows:

Voter	Issue 1	Issue 2	Issue 3
A_1	Yes	Yes	No
A_2	No	Yes	Yes
A_3	Yes	No	Yes
B_1	No	No	No
B_2	No	No	No

(a) What is the majority voting result of each issue?

(b) For each member in the majority coalition A, how many issues out of three does she agree with the voting result?

(c) Repeat (b) for the minority coalition B.

(d) Suppose the leader in A enforces "party discipline" on all members, namely, members in coalition A first vote internally to achieve agreement. Then on the final vote where B is present, all members in A will vote based on their internal agreement. What happens then to the final voting result?

6.4 *Nash Bargaining Solution* ⋆⋆

Alice (Bargainer A) has an alarm clock (good A_1) and an apple (good A_2); Bob (Bargainer B) has a bat (good B_1), a ball (good B_2), and a box (good B_3). Their utilities for these goods are summarized as follows:

Owner	Goods	Utility to Alice	Utility to Bob
Alice	Alarm Clock (A_1)	2	4
Alice	Apple (A_2)	2	2
Bob	Bat (B_1)	6	3
Bob	Ball (B_2)	2	1
Bob	Box (B_3)	4	2

What is the Nash bargaining result between Alice and Bob?

6.5 *Wikipedia (Open-ended question)*

Take a look at the History pages and Discussion pages of two Wikipedia articles: "Abortion", and "Pythagorean Theorem". Summarize 3 key (qualitative) differences you can see between them.