19. *Threshold Models with Feedbacks*

Integration of racial, ethnic, and other groups that mark significant lines of social inequality is a vital ideal for a democratic society.

—Elizabeth Anderson

In this chapter, we cover models of threshold-based behaviors. Threshold-based behavior occurs when people's actions change when an external variable exceeds or falls below a threshold. Threshold-based behavior occurs when a person buys a coat when the price falls below $100 or joins a social movement when its membership reaches 1,000 people. Threshold-based behavior is intuitive and easy to analyze, and it often produces counterintuitive results, such as when tolerant behavior produces segregation. Threshold-based models can often produce tipping points. For example, when a person's decision to join a social movement depends on the number of people already in the movement, as more people join the movement the total number of participants is more likely to exceed other people's thresholds to join, creating a tip.

The models in this chapter can be classified as *agent-based models*—computer programs that model each agent individually. Agent-based models allow for more granularity than systems dynamics models, which represent an entire population by a single stock variable. Agent-based models position agents in space and can include behavioral diversity. Those added degrees of freedom offer advantages, but we must keep in mind that too much detail undermines some of our reasons for modeling. We should not, for example, model each neuron in a person's head when building a model of how people choose whether to join a social movement. The optimal level of granularity will depend on the purpose of the model.

The chapter begins with Granovetter's model of riots, followed by a double riot version that models the growth process for start-ups. We then cover two segregation models. The first considers people moving between rooms at a party. The second considers segregation at the scale of cities. Next we introduce the ping-pong model, which includes negative feedbacks and can produce a cycle or an equilibrium. In the discussion at the end of the chapter, we again return to the Game of Life and see how the double-threshold rule creates a combination of positive and negative feedbacks. Positive feedbacks create correlated behaviors that negative feedbacks dampen. We also return to the topic of model granularity.

Granovetter’s Riot Model

In a threshold-based model, an individual takes one of two actions, depending on whether an aggregate variable exceeds a threshold. If the variable’s value exceeds the threshold, the individual takes one action. Otherwise, she takes the other action. Our first model captures riots and social movements. In it, each person makes a choice to either join the riot or hang back. The decision hinges on the number of people involved in the riot. The model does not take a normative position. The social movement or uprising could be a justifiable revolt against a despot or soccer hooligans destroying property. The model applies to both cases.

The *riot model* assigns each person a threshold. A person joins the riot when the number of joiners exceeds that threshold.[1] Initially, only those who have a threshold of zero join. For the purposes of this discussion, we assume a social movement rather than a riot, so in this case joining may involve standing in a central square or marching. Suppose that on the first day, 200 people with a threshold of zero start a movement. On day two, those 200 people continue to protest, and they are joined by those whose participation threshold lies below 200. If this consists of an additional 500 people, then on day three, people with thresholds above 700 join. This could be several thousand people.

Riot Model

Each of N individuals, indexed by i, has a **riot threshold**, $T(i) \in \{0, 1, \cdots N\}$. Initially, any individual with a riot threshold of zero, $T(i) = 0$, joins the riot. $R(t)$ equals the number of people rioting at time t. Individual i participates at time t if $T(i) < R(t - 1)$.

Analysis of the model reveals that the diversity of thresholds matters at least as much as the average threshold. We can see why by comparing three scenarios involving a population of 1,000 people. In the first scenario, everyone has a threshold of 10, so no social movement occurs. In the second scenario, 5 people have a threshold of zero, 10 people have a threshold of 1, and everyone else has a threshold of 20. In this scenario, 5 people join initially. The next day, 10 more join. Thereafter, no one else joins. In the third scenario, each person has a unique threshold ranging from 0 to 999. For convenience, we can number people from 0 to 999 according to their thresholds, where person i has a threshold of i. In the first period, person 0 joins. The second day, person 1 joins. On the third day, person 2 joins. Each day one more person joins, until all 1,000 people participate. The first scenario has the lowest average threshold, yet it produces no social movement because no one has a threshold of zero. In the second scenario, some people have thresholds of zero, but not a sufficient number to create a widespread movement. Only in the last scenario does the social movement take hold.

The model reveals the importance of the entire distribution of thresholds, not just the mean. It therefore shows the difficulty of predicting which social movements will be successful. The model can also guide action by informing revolutionaries who wish to start an uprising against a despot that in addition to having a core group of people to begin the movement, they also need to create a population of others who will join them. Variants of the riot model can be applied to standing ovations, to changes in political views (the acceptance of gay rights), to fashion changes

(wearing bow ties), and to market dynamics (joining in a stock market or real estate bubble). In each case, people's behavior may be approximated by a threshold-based rule, and that threshold varies across people. In each of these contexts, the likelihood of a large event—be it a mass movement or a fad of thicker-rimmed eyeglasses—may depend as much or more on the distribution of thresholds than on the mean value.

Market Creation and Double Riots

The riot model can be extended to cover internet start-ups that create new markets of buyers and sellers. To create a new market, a start-up must create a population of buyers and a population of sellers. A site that matches dog owners and dog sitters needs to sign up dog sitters as well as dog owners. Similar incentives exist for sites that offer package delivery, transportation, or housecleaning. Each must create two populations to succeed, and the populations must grow at approximately the same rate. Otherwise, either the sellers or the buyers will be unable to find a match and they will leave disappointed. In other words, the start-up must create a simultaneous *double riot*.

The successful start-up Airbnb provides a mini case study of a double riot. Airbnb matches people willing to rent a house, room, or apartment with people looking for a place to stay for a short period of time. Airbnb needed to build two populations: renters and people letting out their apartments. People looking for a place to rent would visit the site only if the site had a sufficient number of places available for rent. Therefore, Airbnb needed to sign up people willing to list apartments. The first two launches of Airbnb failed. Listing an apartment on the site required effort—downloading pictures and including other information. No one had an incentive to list until Airbnb had a large population of renters.

Thus, Airbnb needed enough listings to create a riot among renters—that is, to get renters to come to the site. They also needed enough renters to create a riot among those who wanted to list rooms and apartments. Whether Airbnb would take off would depend on the thresholds for the two groups. The bigger problem was getting people to list, as that required more effort. Airbnb overcame this problem by going door-to-door and helping people list their apartments. Once that happened, the renting riot began and the listing riot followed.[2] The business succeeded because the founders were able to bootstrap a sufficient number of initial renters so that a double riot ensued. They constructed the tail, and the tail wagged the dog.

Two Models of Segregation

Our next two models, both developed by Thomas Schelling, explore segregation. People segregate by race and ethnicity at multiple scales. We segregate by nation and by regions within nations. The United States is racially segregated by neighborhoods within cities, and even by tables within school cafeterias. These observations can be read as evidence of intolerance. That inference contradicts the increasing number of interracial and interethnic marriages. How can the same people choose not to live near or even eat lunch with others of different races, yet they choose to marry across racial groups?

We could account for these facts if the people in multiracial marriages belong to different social classes than those who sit at segregated tables. But that is not true. Interracial marriage occurs at all income levels, and segregated lunch tables can be found even at the most elite colleges and universities. Schelling's models can accommodate both sets of facts. They show how tolerant people can produce segregation.

The first of these models, *Schelling's party model,* can be thought of as a mashup of the random walk model and the riot model. The model describes a party that takes place in a house with two rooms. The hosts of the party have invited guests that visibly sort into two types. The types could be men and women, blacks and whites, Spaniards and Australians, or goths and jocks. The key assumption is that each person be able to distinguish everyone else's type.

Schelling's Party Model

Each of N individuals has an observable type A or B. Each person randomly chooses one of two rooms. At each moment a person moves to the other room with probability p. Person i has a **tolerance threshold** T_i and leaves her room if the percentage of people in the room of her type falls below that threshold.

To see how segregation arises despite tolerance, imagine a party with 20 Australians and 20 Brazilians. Each person is tolerant and will remain in a room so long as 25% of the people in the room have the same ethnicity. Suppose that initially one room contains 12 Australians and 9 Brazilians and the other contains 8 Australians and 11 Brazilians. No one feels compelled to move, but there will be random movements between rooms, and these will alter the percentages in each room. In figure 19.1, one room contains 11 Australians and only 4 Brazilians. This configuration hovers at a tipping point: if any of the Brazilians leaves, the percentage of Brazilians will fall below 25%, causing the 3 remaining Brazilians to leave as well. If that happens, Brazilians will never move into that room.

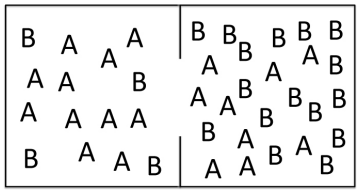

Figure 19.1: A Configuration Poised to Produce Segregation

Recall from the random walk model that a one-dimensional random walk will cross any threshold. The number of Australians in a room is a one-dimensional random walk. Therefore, if the party lasts long enough, segregation is inevitable. Even the best parties only last so long, so not all parties end in segregation. We also know that larger parties are less likely to segregate because the random walk has to cross a much higher threshold when the party has more people. At a party of 1,000 people in two equal

sized rooms, we would not expect the proportion of either type to fall below 25% of either type in a room. We would expect that to happen at a party of 12 people. We should therefore expect more segregation at small parties.

We should also expect more segregation when people have diverse tolerance thresholds. To see why, assume 10 Brazilians and 10 Australians and assign each person a tolerance threshold between 5% and 45% in such a way that each group has an average tolerance threshold of 25%, as shown in figure 19.2.

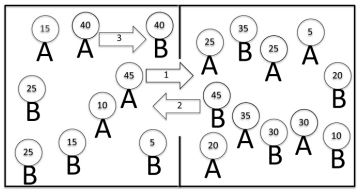

Figure 19.2: Relocations Caused by Heterogeneous Thresholds

The room on the left contains 5 Brazilians and 4 Australians, so the proportion of Australians is less than 45%, causing the least tolerant Australian to move to the room on the right (denoted by the arrow #1). When this Australian moves to the other room, she reduces the proportion of Brazilians in that room, causing the least tolerant Brazilian to move (denoted by arrow #2). These two relocations lower the percentage of Australians in the room on the left below 40%, causing the second-least-tolerant Australian to follow her into the room on the right (denoted by arrow #2). A cascade ensues. However, as shown in figure 19.3, the result need not be full segregation, as the most tolerant individuals are comfortable in either room. The model produces two effects of diverse thresholds: they make the tip to segregation more likely and make complete segregation less likely, as very tolerant people are content in either room.

This model can help explain variations in gender ratios across profession —why more women work as nurses and more men work in sales. Those differences could be due to preferences, but they could also arise if some people prefer to work with people of the same gender. This can be made more formal in a *revolving-door model*, which makes two empirically based assumptions: (i) women who exit a profession choose a new profession with more women, and (ii) women leave professions at a higher rate then men.[3] If women in the life sciences leave biomedical research at a faster rate than

men and take jobs in professions such as health care that employ more women, their actions increase gender segregation in both professions.

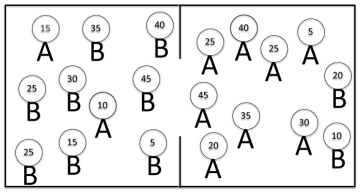

Figure 19.3: Segregation Resulting from Heterogeneous Thresholds

Schelling's Segregation Model

Schelling's segregation model places agents in distinct locations in geographic space, a checkerboard. Otherwise, it is identical to the party model. It assumes two types of people and makes the same behavioral assumption as Schelling's party model.

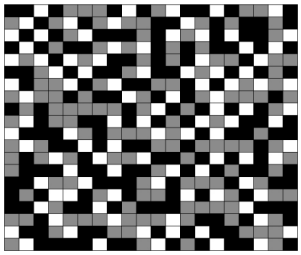

Figure 19.4: Initial Configuration in Schelling's Segregation Model

Schelling's Segregation Model

N individuals, each of whom has a type *A* or *B*, are randomly arranged on an *M*-by-*M* checkerboard with room for open spaces. Each person *i* has a **tolerance threshold**, T_i, and relocates to a random new location if the percentage of the people of her same type on the eight neighboring squares falls below her threshold.

If individuals have an average threshold near 50%, the model produces segregation, as shown in figure 19.5. Segregation arises because the individuals consider only their local neighborhoods, which have at most eight occupants. Almost any random initial configuration includes some people surrounded by others of the opposite type. If individuals move into regions with more individuals of their same type, they can cause relocations by people of the other type. As the relocations accumulate, segregation occurs. We need not walk through the logic again as to why threshold diversity exacerbates these effects.

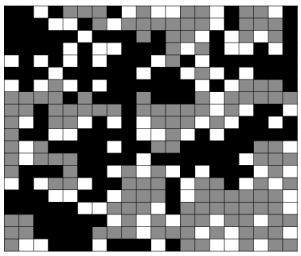

Figure 19.5: Configuration After Relocations in Schelling's Segregation Model

That tolerant people can produce segregated residential patterns serves as the foundational insight of Thomas Schelling's seminal book *Micromotives and Macrobehavior*: what occurs at the macro level need not align directly to the micro level motivations of individuals.

Threshold Models with Negative Feedbacks

Our next model, the *ping-pong model*, assumes threshold-based behavior that produces negative feedbacks. Recall that negative feedbacks can stabilize a system or, as we saw in the predator-prey model, produce cycles. The model assumes a finite number of entities, which could be people or mechanical, biological, or chemical devices. In each period, each entity takes either a positive (+1) or a negative (-1) action. In the first period, each agent chooses a random action. The initial state equals the sum of the actions. All subsequent states of the system equal the previous period's state plus the average of all actions and a random term. Each entity has a threshold, drawn from a distribution, and chooses the action that reduces the absolute value of the state if the absolute value exceeds its threshold. Put simply, if the state's value exceeds an entity's threshold (either negatively or positively), the entity does what it can to reduce the magnitude of the state.

Ping-Pong Model

Each entity in a population of size N randomly takes an initial positive (+1) or a negative (-1) action. The initial **state of the system**, S_0, is set equal to zero. All future states of the system, S_t, equal the average action plus a random variable:

$$S_t = \frac{\sum_{i=1}^{N} A_i(t)}{N} + S_{t-1} + \epsilon_t$$

Each entity i has a **response threshold** $T_i > 0$ drawn uniformly from the interval [0, RANGE]. An entity takes the same action as before if the magnitude of the state, $| S_t |$, is less than its threshold and takes an action to reduce the magnitude of the state otherwise.

If $| S_t | \leq T_i$, $A_i(t + 1) = A_i(t)$, otherwise $A_i(t + 1) = -\text{sign}S_i(t)$ where ε_t is randomly drawn from {-1, +1}.

Multiple applications of the ping-pong model should come to mind. Here are two. People allocate time and resources to multiple charitable causes. If cause receives too much attention or money, people may donate money to other charities to equalize their revenues. Or a country may have two UNESCO World Heritage sites that rely on volunteers. If one site has an abundance of volunteers, then some may reallocate their energies to the other site.

As foreshadowed by the name, the ping-pong model can produce cyclic behavior around the equilibrium. In one period too many people choose one action, and in the next too many choose the other. When all entities have a threshold of zero, all entities choose action one (+1) in one period and action minus one (-1) in the next.

To explore how threshold diversity contributes to whether people behave like ping-pong balls or find an equilibrium, we consider three cases, each involving 100 people. In the first case, we assume thresholds are uniformly

distributed between zero and 10. If the state in the first period equals -6, it will exceed the threshold of approximately 60 people. Approximately 30 of these 60 will have taken action one and will switch actions. The sum of the actions will now exceed 50, so the new state of the system (the average of the previous two periods) will exceed 20. This value exceeds all thresholds, producing the ping-pong effect shown in the top graph in figure 19.6.

If we increase the threshold diversity making them uniformly distributed between zero and 100, the ping-pong effect all but disappears. To see why, assume that the state equals -6 in the first period. On average, only six people's thresholds will be met. If three change actions, the state will move toward zero. This dampening of the deviation can be seen in the bottom graph in figure 19.6, which corresponds to thresholds between zero and 100. As we might expect, if we consider a moderate case, with thresholds uniformly distributed between zero and 60, we see a more moderate cycle, as visible in the middle graph. Thus, in systems with negative feedbacks, threshold diversity produces stability, but it has the opposite effect in the models with positive feedbacks.

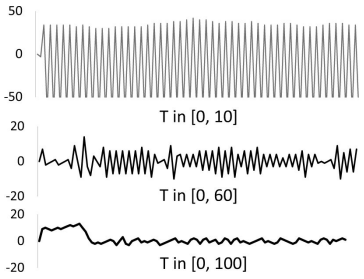

Figure 19.6: Time Series for Ping-Pong Model for Three Threshold Ranges

Summary: Model Granularity

The basic logic of feedbacks is straightforward: positive feedbacks reinforce actions, negative feedbacks dampen them. A system with only positive feedbacks will either blow up or collapse. A system with only negative feedbacks will either stabilize or cycle. A system with both positive feedbacks and negative feedbacks has the potential to produce complexity.

In systems dynamics models, feedbacks operate on stock variables (the number of hares) and rates (increasing the rate of purchases at the bakery). Agent-based models, such as the threshold models we covered here, produce feedbacks as a result of individual actions. These more granular models can include a diversity of thresholds, which exacerbates the effects of positive feedbacks and tempers negative feedbacks. Diversity in the tail of the distribution made riots much more likely. Diversity in tolerance thresholds increased segregation. Diversity of responses in systems with negative feedbacks stopped the wild fluctuations that existed with homogeneous feedbacks. In models of economic competition among firms, heterogeneous production costs can play a similar function. As prices rise or fall, firms vary in their responses owing to their differences in costs.[4]

This difference between the two types of models raises the issue of model granularity. Does one create a variable (or box) labeled “hares” and describe how the number of hares increases and decreases as a function of other variables, or does one model each hare individually? Disaggregating variables increases descriptive accuracy, but models are not graded on that criterion. Remember Box’s dictum that all models are wrong, and Borges’s map that was the same scale as the real world. Many modelers, including Einstein, have taken the position that we should seek the correct level of granularity. If writing a model to explain the force that is exerted by a human arm, for example, we need not include DNA.

For studying social systems, no ideal level may exist. We may need to explore them at multiple levels of granularity. By constructing multiple models, each at a distinct granularity, we can engage in a dialogue across those models. If we are attempting to make sense of trade patterns between

Sweden and Finland, we could start with the two countries as variables and identify broad macro patterns. We might then disaggregate each country's imports and exports into industries, and then into firms within those industries. Industry-level data would enable us to better explain past patterns and make more precise predictions of future patterns. The deeper dive into firms within those industries, including information such as their cost structures and growth trajectories, could produce even better results, but we would need a lot of information to construct a useful model with that many moving parts. We might even model the leadership within those firms. Most likely, this last level of unpacking would yield few benefits, but it could be that some leaders are known to pursue expansionist strategies.

To summarize, finer granularity need not be better. Models can have too much detail. Even when we can understand more granular models, we benefit from constructing coarser models as a point of comparison. By comparing the differences in the models' predictions, explanations, and policy prescriptions, we see how assumptions affect results. We see the conditionality of our assumptions. Thus, our many models should not only differ in the variables they include, they should also differ in the extent to which they disaggregate the same variables.

Algorithmic Riots

The riot model and the ping-pong model can provide insights into stock market crashes and price rebounds. We consider two illustrative cases here. The first occurred on Monday, October 19, 1987, when the Dow Jones Industrial Average fell by over 22%. The next day, this collapse reverberated around the globe. The cause of the crash remains a topic of analysis. At the time, the United States economy was in its fourth year of expansion. In the first eight months of the year, the Dow Jones Industrial Average had risen over 40%. Despite or perhaps because of this run-up in values, many people believed stock prices were overvalued. On the Sunday prior to the crash, the United States secretary of the treasury, James Baker, threatened to weaken the dollar if Germany did not lower tariffs, a comment that at the time did not seem of great import. The next day, the market crashed. Fourteen months later, the market returned to its previous value.

To apply the models, we assume a single financial asset representing the entire market. We assume that each person holding this asset has a *crash threshold*. If the price of the asset falls more than the crash threshold in a given day, the investor sells the asset, taking her money out of the market. This rule captures the behavior of trend or *noise traders* and creates a version of the riot model. If some percentage of investors woke up on October 19 and decided to sell a substantial amount of assets, they would have caused a drop in the market. If that drop exceeded other investors' crash thresholds, those others would sell assets as well, causing a downward spiral. The result is a classic positive feedback loop. Selling begets lower prices, which causes more selling.

We now add insights from the ping-pong model to the analysis. If prices falls too low, some people will apply a second rule, a *bargain threshold rule*. According to this rule, a person buys if the price falls below this value. Here, our investor acts based on value, not trends.

When prices fall dramatically, the bargain thresholds create a negative feedback. Buyers rush in to buy at a bargain price, halting the price fall.

Actual markets are more complex than this simple account of sellers with thresholds and buyers waiting in the wings. The stock market contains many types of traders, including large institutions, pension funds, foreign governments, speculators, portfolio insurers, and speculators, as well as small investors.[5] As a result of this diversity, someone is almost always willing to buy as prices fall, producing the negative feedback necessary to stabilize the market.

Early analyses of the crash emphasized the prevalence of *(computer) program trading*. These are threshold-based rules encoded in computer programs. Rules such as *sell all stocks if the market index falls below some set price* were carried out automatically, with no human supervision. Most analysts now believe that program trading contributed to the 1987 crash but was not the primary cause. More detailed analyses of the 1987 crash reveal that a large number of portfolio insurers—traders who guarantee a rate of return to the portfolios of their investors—produced strong positive feedbacks that were not tempered by negative feedbacks. As the market fell, these portfolio insurers sold off stock to prevent losses. As the crash unfolded, these insurers sold more and more stock. In effect, insurers acted as if they were populations of individuals with diverse thresholds. One portfolio insurer sold over $1 billion in stock. To put that in perspective, only $20 billion in stock was sold that entire day.

The second crash, the May 6, 2010, “flash crash” dropped the Dow Jones Industrial Average by 5% in three minutes. It was the result of algorithmic trades. Owing to the complexity and speed of modern financial markets, no one knows for certain what exactly caused the flash crash. We know that a large mutual fund made a huge sell order, dumping over $4 billion of stock futures into a market containing high-speed trading algorithms that try to exploit beneficial trades. The algorithms sensed a price trend and starting executing trades at breakneck speed. Imagine the riot model at high speed. This produced a toxic market, in which traders worry that large institutional investors

know something that they do not know and so they exit the market.[6] Many of the algorithms stopped trading given the abnormal market behavior; other algorithms kept selling, and a crash ensued, all in the span of a couple of minutes. Within twenty minutes, the bargain rules went into effect and, as predicted by the ping-pong model, brought prices back up (nearly) to the original price.