PART II

Agent-Based Computational Model

IN AGENT MODELING, we essentially build artificial societies of software individuals who can interact directly with one another and with their environment according to simple behavioral rules. On agent-based modeling in general, see, for example, J. M. Epstein and Axtell (1996), Axelrod (1997a), Resnick (1994), J. M. Epstein (2006), Tesfatsion and Judd (2006), Miller and Page (2007), and the large literature cited in these works.[118]

I developed this model in *NetLogo* 5.0. Source Code for the canonical[119] Parable 1 run is given in Appendix IV. A table of parameter values for every run is also provided. As earlier noted, all movies are posted on the book's Princeton University Press Website. Interactive Applets for each movie run are provided there as well. The Applets allow the user to alter various assumptions with "sliders," movable bars on the Interface. These user-adjustable parameters include the attack rate, search radius, extinction rate, memory length, and damage radius, for example. This offers nonprogrammers an extensive basis for experimentation with the model. For programmers, the Applets also include the full source code for every run. Hence, all results are certainly replicable. However, the English-language exposition that follows is meant to be sufficient to permit replication by reasonably adept programmers (who are also good readers).

Replicability

Apropos of this, I am not sure replicability is an attribute of models proper. Leaving aside the case of authors who are literally pretending to have a model, one could always "replicate" model output by running the same model on the same inputs. So, when a person says a model was not replicable from some article, they are really asserting that the author's *English-language exposition* of the algorithm was insufficient to permit a reimplementation by that particular reader. If so, it would appear to measure the author's facility in English—or the reader's lack thereof—but it has nothing to do with the actual computer program or mathematical equations, which—if provided, as here—are replicable ipso facto. In any event, such ambiguities as may arise can be resolved by reference to the code provided in Appendix III and on the book's Princeton University Press Website.

Present Interpretation

Later, I will offer a number of alternative interpretations of the model in the fields of health behavior, economics, network science, and law. But for expository purposes, we imagine a conflict, indeed a guerilla war like Vietnam, Afghanistan, or Iraq. As discussed in the Introduction, events transpire on a 2-D population of contiguous yellow patches, each of which represents an indigenous agent. Specifically, we imagine that a single stationary indigenous agent occupies each patch. This expository grid is 33 by 33 (the default *NetLogo* dimensions), so there are 1089 Yellow agents.

These indigenous patch-agents do not move. They have two possible states: inactive and active. At any point in time, they occupy only one of these states. Inactive agents are yellow. I have given them slightly different shades of yellow just so they are visually distinguishable squares, as shown in Figure 34. Active agents are orange. These agents activate randomly, at a rate (the attack rate) adjustable by the user. They will be discussed shortly. The three Blue agents represent occupying forces and are of the full *Agent_Zero* type. They are mobile. Every cycle through the (randomized) agent list, the agent adopts a random heading and takes one step in that direction. So, they do not jump to random distant sites but move to random neighboring ones. They execute a 2-D random walk, in short. It is not a perfect mixing, or mass-action kinetics, process. The space is a finite bounded square lattice.[120] These Blue "rovers" are connected to one another bidirectionally, as indicated by the two-headed arrows shown in Figure 34. (Hence, there are in-degree and out-degree distributions, and so forth). This exactly parallels the mathematical networks developed above. Some rovers give high weight to other rovers; some do not (see the Appendix IV table, or *NetLogo* Code in Appendix III for the values employed).

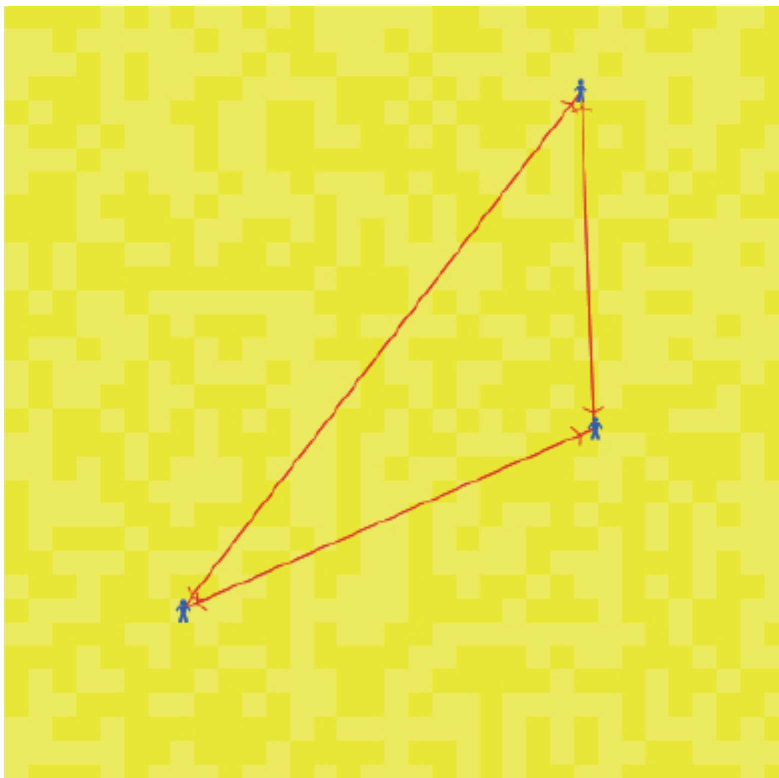

FIGURE 34. Indigenous Population (Stationary Yellow Squares) and Occupying Rovers (Mobile Blue Agents) [Movie 1]

Minimalism

I have hand-coded three agents to ensure complete control over specifics in the smallest possible model that can exhibit majorities. For large *n*, one would of course initialize the agents with random weights and other parameters drawn from distributions. The agents and their connections are shown in Figure 34.

Movie 1 (on the Princeton University Press Website) simply shows the three agents in random motion connected to one another.

Now, let us posit a distinguished region of the space—in this version, the northeast quadrant—where yellow patches "activate" at a user-specified random rate (a global constant, implemented as a user-adjustable slider in the *NetLogo* interface).[121] Think of these agent activations as insurgent attacks. These explosions are shown as orange patches in Figure 35. I make no assumptions whatsoever as to the comparative legitimacy of occupying or insurgent agents. **Movie 2** shows these, with fixed Blue agents.

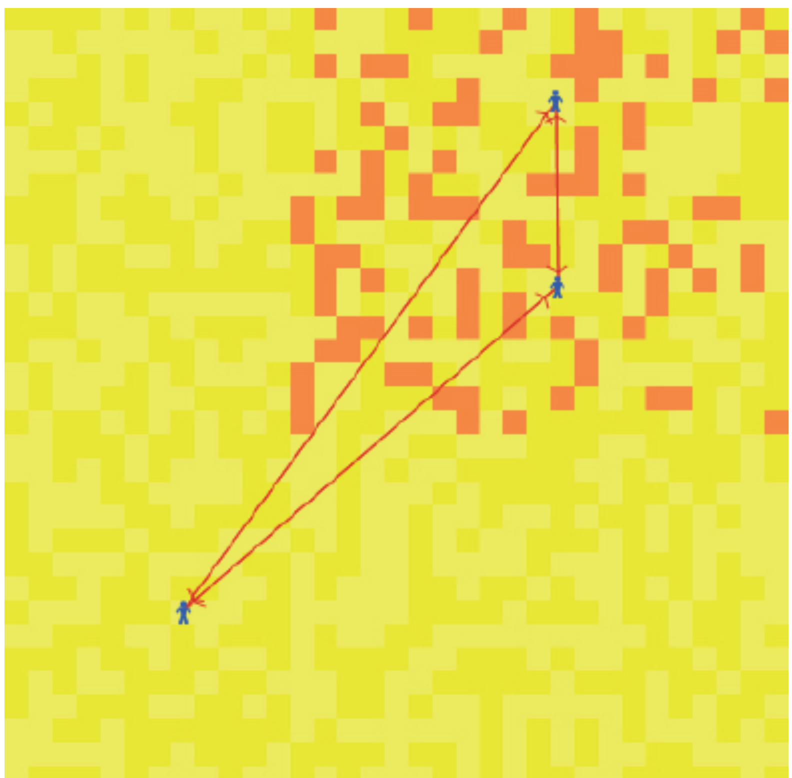

FIGURE 35. Adverse Event Activations [Movie 2]

It is central to distinguish between a Blue agent's separate affective, "rational," and social components and to understand how they are combined to form the agent's overall disposition in the wake of attack. As before, once this disposition is formed, it is compared to the agent's threshold. If the overall disposition exceeds threshold, action is taken; otherwise it is not. In this interpretation, *action is the destruction of indigenous (Yellow) agents within some user-specified damage radius* (again, a slider in the *NetLogo* Interface). Now each component is described.

Affective Component

These orange activations (explosions) are the conditioning trials for the Blue agents. When a Blue agent "steps on" an orange patch, he updates his affect through the generalized Rescorla-Wagner equations.[122] Learning rate parameters (the α's and β's), limiting values for associative strength (λ), and the exponent (δ) all affect individual learning curves and can vary across agents. An extinction rate is applied at every iteration except those in which an active patch is encountered. So, hostile affect toward the indigenous population evaporates at a user-specified extinction rate in the absence of local attacks.[123] This extinction rate can be zero since extinction, as noted earlier, is by no means assured simply by cessation of trials. This is the *affective* component of the Blue agents' disposition.

A suitable extension would be to include the well-established contextual conditioning that Blue agents would presumably undergo in the course of their spatial movements. They would come to associate the northeast quadrant itself with danger, and this would amplify the estimates made purely from event sampling. The hippocampus is central to this well-established contextual conditioning in space. In animal models, Knierim (2009) and Knierim and McNaughton (2001) have used "multi-electrode arrays to record the extracellular action potentials from scores of well-isolated hippocampal neurons in freely moving rats. These neurons have the fascinating property of being selectively active when the rat occupies restricted locations in its environment. They are termed *place cells*, and it has been suggested that these cells form a cognitive map of the environment (O'Keefe and Nadel, 1978). The animal uses this map to navigate efficiently in its environment and to learn and remember important locations" (from Knierim Research

Page, Johns Hopkins Mind/Brain Institute site). *Agent_Zero* agents do not have a mental map of the area and condition only on the event stream, not also on position, though an Agent 1.0 could certainly have this endowment.

"Rational" Component

Turning to the evidentiary/ratiocinative component, Blue agents have a *spatial sampling radius* (which can be heterogeneous but is also a slider in the *NetLogo* Interface), within which they conduct local sampling of *the landscape*,[124] here interpreted as an indigenous population.[125] As discussed earlier, they estimate the probability that an agent is a hostile agent (e.g., the probability that an agent is a terrorist given that he is Muslim) by comput-ing the relative frequency of orange patches within their sampling radius.[126] Obviously, this probability estimator exhibits *sample selection error*—the local ratio may be a poor estimator of the global one.

Some readers may feel that this simple computation is putting "reason" at an unrealistic disadvantage to "the passions." In fact, while this sample estimate is crude statistically, its computation is remarkably sophisticated cognitively. Indeed, this imputes to the Blue agents more cognitive capacity than untrained humans possess. In *The Mathematical Brain,* Butterworth (1999) makes a powerful argument that among our innate universal endow-ments is a *number module*, giving us the capacity to make crude numerosity judgments; and he provides evidence that the parietal lobes are centrally implicated. So, just to be shamelessly phrenological, while *Agent_Zero* walks into an ambush, his amygdala is activated and so he registers fear, but his number module is also making a very crude frequency judgment: *enemy/total*. Butterworth argues that even this simple relative frequency is very hard for humans to compute, which suggests a neural basis for one of the best-documented biases in all of psychology: base rate neglect (Kahneman and Tversky, 1973; Tversky and Kahneman, 1982). As he puts it, "we ignore base rates because we ignore rates" (Butterworth, 1999). So, simple as it seems, *Agent_Zero*'s computation of a local ratio is far from trivial.

Another factor that would corrupt the Blue agent's estimate of the actual local ratio (itself a biased estimator of the global one) is the specific areal pattern in which the activations present themselves. In experiments, human subjects are quickly shown two spatial arrangements of dots: one has them spread over a wide area, and the other has them tightly packed. We will judge the former pattern to be the more numerous (Krueger, 1972, 1982). One can imagine how this areal bias might have conferred a selective advantage—we are more vulnerable (and so more alert) if surrounded by predators than if they are all clustered within our vision (giving us more escape routes).

It also happens that, even if our dots occupy the same total area, random patterns (as in this model) are typically overcounted as against regular ones, again perhaps because unpredictable predator patterns are harder to anticipate and evade than regular ones.

Related mechanisms may explain why we involuntarily complete patterns like those shown in Figure 36. The seminal example is the Kanizsa triangle (Figure 36A), after Italian psychologist Gaetano Kanizsa (1955).

This "phantom edge phenomenon" (seeing an outline that is not actually there) is due to what neuropsychologists call the "T-effect."

> Groups of neural cells see breaks in lines or shapes, and if given no further input, will assume that there is a figure in front of the lines. Scientists believe that this happens because the brain has been trained to view the break in lines as an object that could pose a potential threat. With lack of additional information, the brain errs on the side of safety and perceives the space as an object. The circle is the most simple and symmetrical object, so the mind usually sees a circle unless active effort is made to see an alternate shape. This illusion is an example of reification or the *constructive* or *generative* aspect of perception, by which the experienced percept contains more explicit spatial information than the sensory stimulus on which it is based (Ehrenstein illusion, n.d.).

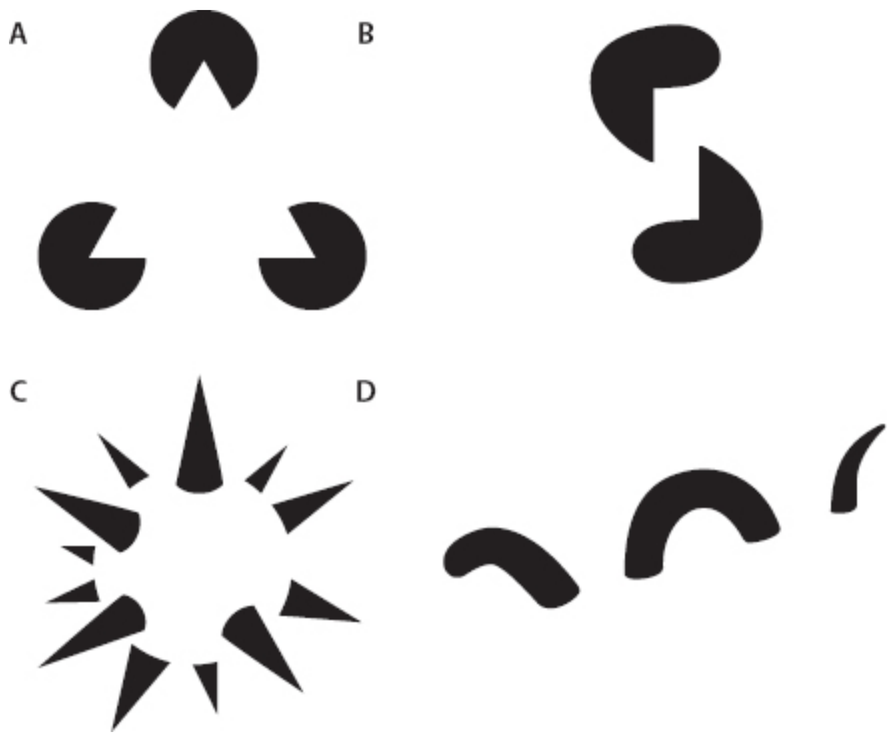

FIGURE 36. Phantom Edges

This is yet another source of potential Blue agent "threat inflation" that we shall ignore. I would say that propaganda generally—in "completing" political patterns that aren't there or inviting their completion—traffics on this same apparatus. Indeed, the entire art of propaganda is to offer as little of the picture as possible, leaving it to the audience to "fill in the blanks" opened by vague outlines of subversive "others."

Finally, the base model treats the affective and statistical estimates as independent when they are almost certainly entangled. There is interesting experimental work on the classification (as hostile or peaceful) of inconclusive data, specifically under circumstances of threat (Baranski & Petrusic, 2010). Affect, in other words, colors one's probability judgments, particularly in settings of the sort we have posited.[127] I introduce this in an extension of Part III. By contrast, in the basic model, these are superposed but decoupled—neither is a mathematical function of the other.

In sum, Blue agents simply compute the relative frequency of orange patches within their spatial sampling radius to estimate the likelihood that patches are immanently violent. This initially appears to be a very crude algorithm. In fact, it would probably be way beyond most humans, particularly in the stressful circumstances of interest here. But, since we want to give reason a "fighting chance" against passion, we'll start here. So, we now have an elementary type of bounded rationality, in addition to a simple representation of affect.

Social Component

The third *Agent_Zero* ingredient is social. At any time *t,* the total disposition of each agent is the sum of her affect, $V(t)$, and her local probability estimate, now a function of time, $P(t)$, plus the sum of each other agent's weighted solo disposition (each the sum of their own V and P), all minus her threshold.[128] Unless otherwise noted, the term *disposition* will denote *net disposition* in all *NetLogo* graphical output.

Sampling and Dispositional Radii Mathematically Independent

It is important to reiterate that the mechanism of influence in the model is not behavioral imitation, even if agents are within the narrow spatial sampling radius of one another. In Part III, an extension offers a way to introduce this distinction. But we do not use it in the main development. Agents can influence each other (have dispositional weight) at *any* range, by a large variety of avenues (e.g., auditory and textual social media), and the binary actions of other agents can alter the landscape (by destroying sites), which can affect one's frequency calculation. But binary action proper is not registered or, therefore, imitated. The spatial sampling radius is typically a cluster of contiguous sites on the landscape proper, such as a Von Neumann neighborhood. This spatial sampling radius is bounded and landscape specific. The radius of dispositional contagion is neither; the two are mathematically independent.[129]

Action

If overall disposition is greater than the threshold, action is taken: the agent destroys all patches within a user-specified damage radius (another slider).[130] Destroyed patches (indigenous agents) are colored *very* dark (i.e., blood) red and cannot be active (they are dead).

Pseudocode

So, for each Blue agent, the algorithm (pseudocode) is as follows:

```
Compute own affect (with orange explosions as conditioning
    trials);
Compute own local probability (relative frequency of orange
    within spatial sampling radius);
For each other agent in network
    Compute the weighted solo disposition;
Add the above-computed numbers;
Subtract own threshold;
If the result is positive, Act; otherwise don't;
Apply own extinction rate to own affect;
Move;
Repeat.
```

What Is Time?

Finally, it is worth noting exactly what we mean by "time" in this model. In the agent model (as against the continuous-time differential equations) time is discrete. Here, time advances by one unit with every complete updating cycle of every agent and every patch.[131]

II.1. COMPUTATIONAL PARABLES

Science begins as parable, and ends as probability.[132] As this is a very young science, the runs that follow are closer to parables than to mature scientific claims of any sort. They arguably qualify as explanatory candidates in a broad sense, in that they *generate* certain qualitative behaviors (see J. M. Epstein, 2006). But they are computational parables—fables if you prefer. Of course, some fables endure.

Parable 1: The Slaughter of Innocents through Dispositional Contagion

For the base case run of the agent model, we will immobilize one of the agents. Call him Agent 0. *Netlogo* begins subscripting agents from 0, so this numbering assures consistency with the code provided. But "*Agent_Zero*" is the name of a class, while "Agent 0" is the name of an individual instance of that class. Lest any confusion arise, all the agents are of the general *Agent_Zero* type, just as all the diverse actors in a classical economic model would be of the *homo economicus* type (with different parameters, for instance). Agent 0 will be stationary in the southwestern quadrant of the landscape. The other two agents, Agent 1 and Agent 2, will execute random walks on the landscape but will begin in the hostile northeast quadrant. Agents can sample only the four sites to their immediate north, south, east, and west—their Von Neumann neighborhoods. All agents update their affects and local probability estimates, with dispositions updating over the fully connected network. For the Base Case, there is no affective extinction.

Crucially, Agent 0 is never attacked. As he is subjected to no direct conditioning trials, his immediate direct *affect is zero throughout.* Because he encounters no orange attack events, his estimate of probability (hostile given indigenous) is also *zero throughout.* Yet, he wipes out a "village"! How? As shown, the two rovers are encountering attacks (orange events). They are updating both their affect and their local estimate of the attack probabil-ity. When their total disposition values exceed their thresholds, they retaliate within their destructive radius (here equal to their sample radius). Destroyed sites are dark red, as shown in the left frame of Figure 37. Their destruction and the escalation of their affects and probabilities continue. At all times, Agent 0 is weighting these (i.e., their solo dispositions) and adding them to his own destructive disposition.

Finally, these push him over his own threshold and he wipes out innocents, despite having a sample probability of zero, and no direct emotional grievance against the population, as depicted in the right frame.[133] *Also, Agent 0 is not imitating the destructive behavior of either other agent.*

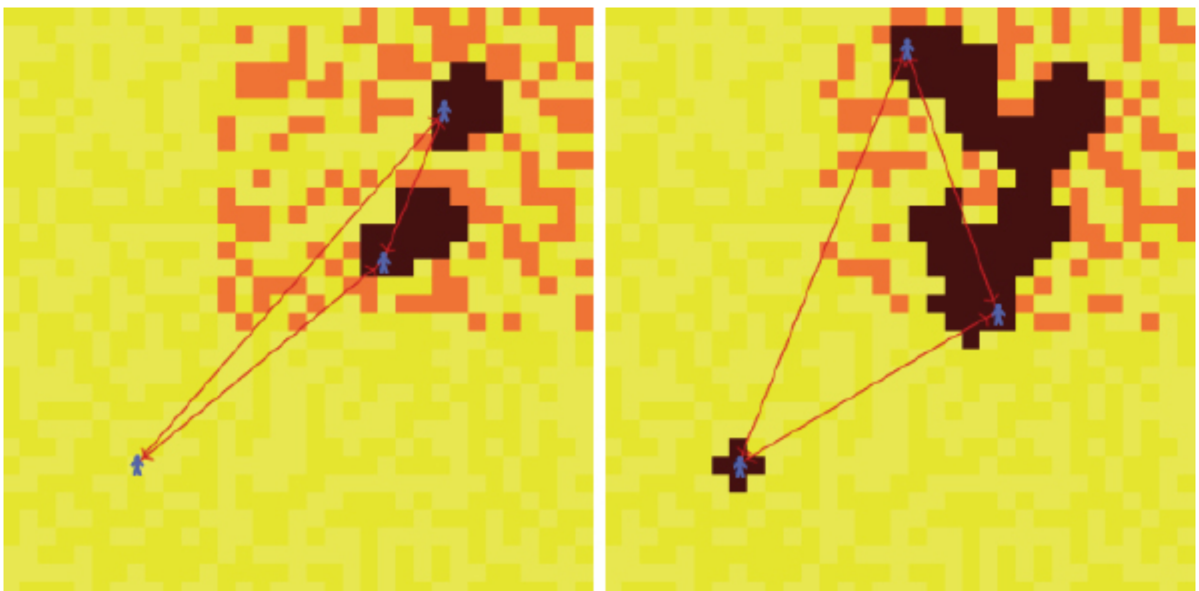

FIGURE 37. Activation by Dispositional Contagion [Movie 3]

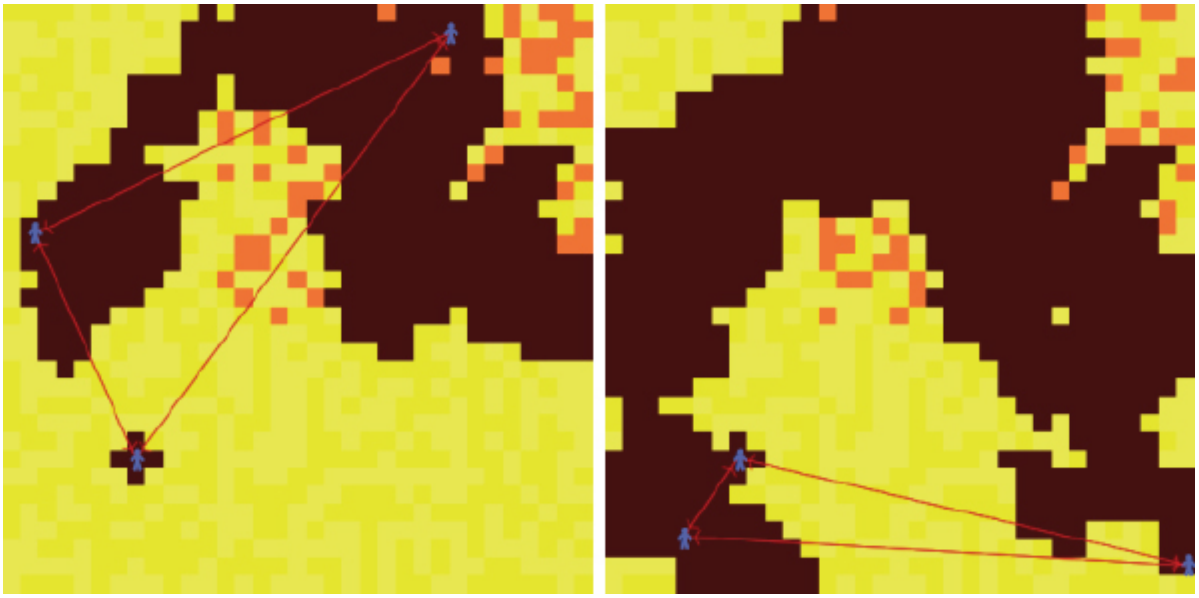

FIGURE 38. Slaughter of Innocents Continues [Movie 3]

In this particular case, he cannot even observe their destruction of the landscape because “vision”—the sample radius—is set to one patch in each direction.[134] With no extinction of affect, the mobile rovers go on to wreak vast destruction in regions that have never done them harm either, as shown in the two frames from **Movie 3** shown in Figure 38.

Specifically, having wiped out many of the insurgents (in the northwest quadrant) and having now drifted out of that quadrant, the rovers are, in fact, encountering mostly yellow (innocent) patches. Accordingly, their estimated probability of a hostile patch (the local relative frequency of orange) falls to zero. Yet, without any evaporation of affect—with no extinction of the conditioned affect—their dispositions remain high, and the killing continues with no direct empirical (observational) basis and no new conditioning trials. This is shown in the time series of net disposition and probability in Figure 39. Disposition and destruction remain high, despite falling probability for Agents 1 and 2. That is, their rampage continues as all empirical basis for it—their probability estimate—evaporates. This behavior is consistent with seminal laboratory psychology work of Zillmann et al. (1975).

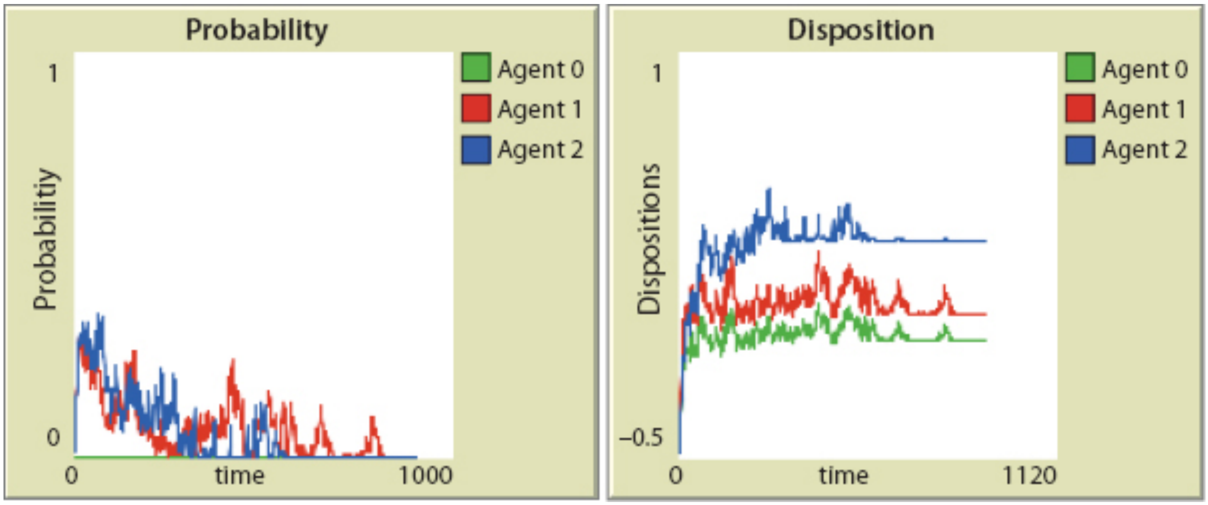

FIGURE 39. Rising Disposition Despite Falling Probability

Zillmann's Experiment

The article describing this experiment in detail is aptly entitled "Irrelevance of Mitigating Circumstances in Retaliatory Behavior at High Levels of Excitation." In sum, Zillmann et al. (1975, p. 282) showed that "Under conditions of moderate arousal, mitigating circumstances were found to reduce retaliation. In contrast, these circumstances failed to exert any appreciable effect on retaliation under conditions of extreme arousal." Specifically, "the cognitively mediated inhibition of retaliatory behavior is impaired at high levels of sympathetic arousal and anger." These conditions of affective arousal are certainly met, and agent behavior is entirely consistent with Zillmann's result.

Again, Agent 0's probability (the green curve of the left panel of Figure 39) is zero throughout. He would never have acted alone. And, he would never have acted even in a model of behavioral imitation, because he literally cannot "see" the others, and he need not, if they are in other forms of communication, such as auditory and social media.[135] The entire *NetLogo* Code for this parable is provided in Appendix III and again on the Princeton University Press *Agent_Zero* Website. This is a disturbing run,[136] but it is not yet our canonical central case, because Agent 0 does not act *first*.

Parable 2: Agent_Zero *Initiates: Leadership as Susceptibility to Dispositional Contagion*

Having developed all this apparatus, we can now generate that case, in which the first agent to act is not the one with the highest affect or the highest empirical estimate of indigenous hostility. Indeed, Agent 0 (again stationary) is subject to no direct aversive stimuli (orange explosions), so his individual (i.e., directly stimulated) affect and probability are both zero throughout, as shown in the corresponding plots of Figure 40. By contrast, the mobile rovers are subject to attacks, are accumulating affect, and are in-creasing their estimates of the probability that an indigenous patch is hostile (that a random patch will turn orange). All thresholds are equal at 0.5,[137] but neither rover's disposition exceeds this, so neither of them acts. Through their weights, however, their dispositions elevate Agent 0's to the highest of levels (see disposition plot), which exceeds the common threshold first. So, he is the first to act, as shown in the Figure 40 screen shot and **Movie 4**.

This is the situation I aimed to generate: *The agent at the front of the lynch mob has no particular grievance V or evidence P, and left to his own devices would never act. Notice that this is not "the banality of evil." Agent 0 is not "just following orders," because none are issued. And he is not imitating the behavior of others, because he is the first to behave!* The deeper point, as emphasized throughout, is that no agent is imitating *the behavior* of others, regardless of the order in which they activate. Thus, the model can generate important group dynamics without recourse to the copying of behavior (which is a binary variable that doesn't enter into the disposition calculus). Action (i.e., behavior) occurs if total disposition exceeds threshold, which occurs first for Agent 0.[138]

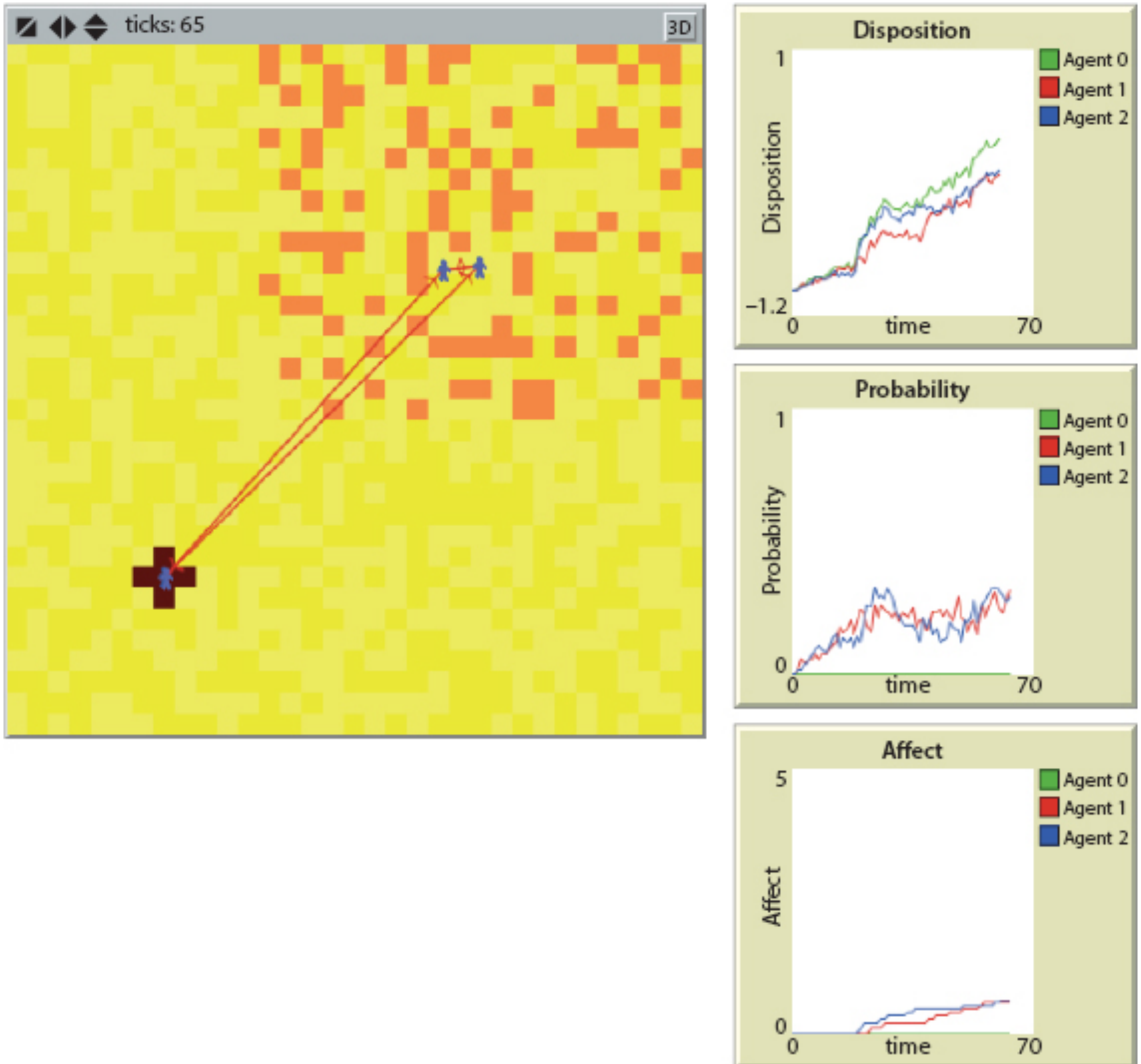

FIGURE 40. Agent with Zero Affect and Probability Acts First [Movie 4]

So, is Agent 0 a "leader," or is he simply the most susceptible to dispositional contagion? Which is the more compelling picture: the "great man" theory, or merely the susceptible one?[139]

This is Tolstoy's *swarm-life of man* in its most virulent form. Speaking of Bonaparte, Tolstoy writes:

> Though Napoleon at that time, in 1812, was more convinced than ever that it depended on him … he had never been so much in the grip of inevitable laws, which compelled him, while thinking he was acting on his own volition, to perform

> for the swarm-life—that is to say for history—whatever had to be performed." (*War and Peace*, p. 648)
>
> "To perform for the swarm-life. …" What a phrase! And this is the sense in which Tolstoy wrote, "A king is history's slave" (*War and Peace*, p. 647).

Complex Contagion Revisited

It is worth noting that Agent 0 does not act based on either one of the others alone. Here, he requires the swarm, the weighted sum, and multiple dispositional exposures, to go.

Run 3. Information Cuts Both Ways

In the runs thus far, the agents' spatial "vision" (landscape sampling radius) has been limited to a von Neumann (N, S, E, W) neighborhood of radius 4 patches. What is the effect of increasing the agents' vision? Let us begin with Agent 0 in his usual fixed position in the southwest quadrant, which we assume to be peaceful. Now let us give the other two agents fixed positions as well, but in the violent quadrant, as shown in Figure 41. What is the effect of increasing everyone's vision? More peace? More violence? Neither?

Let's consider Agent 0. His vision is his spatial sampling radius. As this extends into the red zone, he is seeing more violence. Hence, his estimate that a random patch is violent grows, as will his violent disposition. The agents in fixed positions in the red zone, however, have the reverse experience. Rather than seeing more violence as their vision grows, they see more yellow—peace! Accordingly, their probability estimate falls. Finally, when vision increases to the point where they can all see the entire landscape, their probability estimates converge, because their samples are now identical. Notice that the sample selection biases were very great at the low-vision outset, with Agent 0 underestimating—and the others overestimating—the global probability. Now they converge on the correct global probability, as shown in Figure 42, where I simply increased the sampling radius midrun with the program's slider.

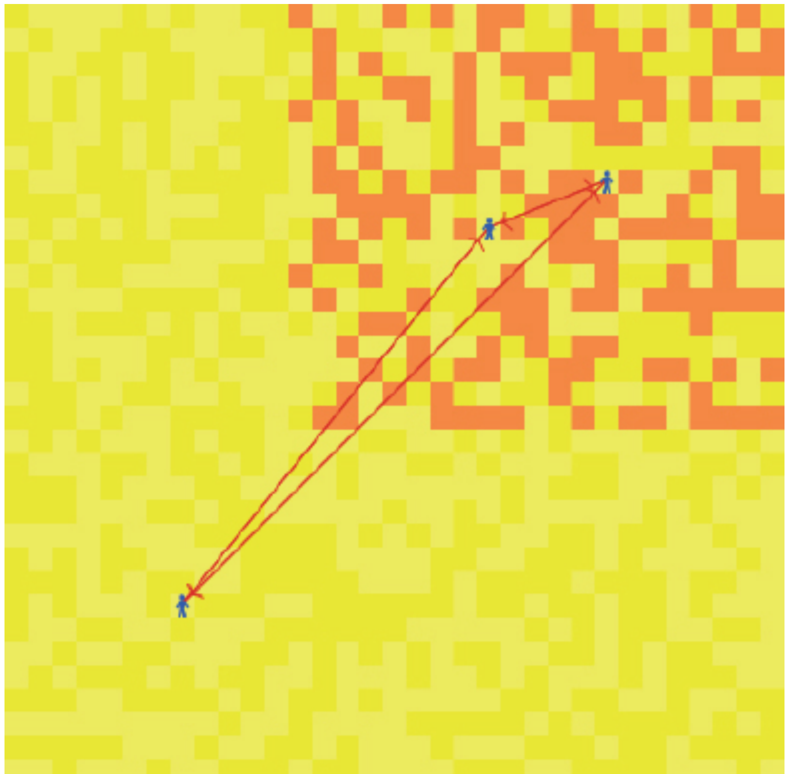

FIGURE 41. Fixed Agent Positions

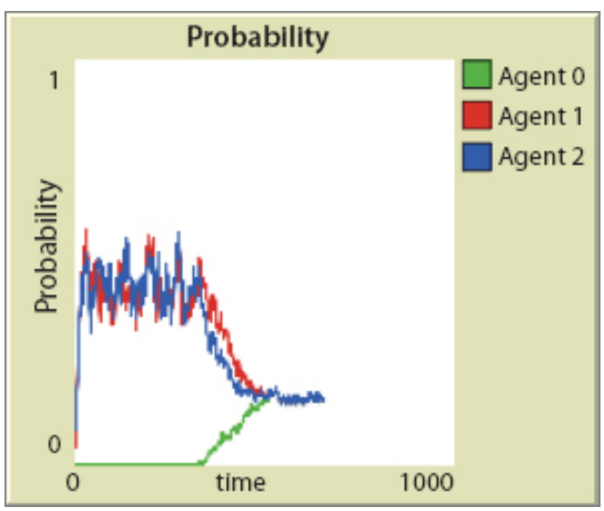

FIGURE 42. Probabilities Converge

This is an example of how sensitivities can be explored “on the fly” (midrun) in *NetLogo*, which readers are invited to do using the interactive Applets posted on the book’s Princeton University Press Website.

Heterogeneous Vision

In this experiment (and in this variant of the model), vision was the same for all agents. Again, I am using the term *vision* figuratively, to denote the agent’s search space. This could be entirely local, or global, or spread over a network, or confined to an organization. It could be literally ocular, auditory, olfactory, or text based, and so forth. The model permits high heterogeneity of vision. And it would be reasonable to explore this in future research, since people, in fact, differ widely in search spaces, and for a variety of reasons. Some types of information are expensive, for example. Some individuals are simply more inclined to acquire and process information than others. Cacioppo and Petty (1982) dub this “the need for cognition” and present experimental research that could be imported into the agent population. Instead of using a single global value for the sampling radius, one could use an empirically based distribution of vision as a crude analogue of this need for cognition. This could be a nice example of *computational social neuroscience*, where individual agents are based on experimental neuroscience, but then interact with one another in simulated populations.

Run 4. A Day in the Life of Agent_Zero***: How Affect and Probability Can Change on Different Time Scales***

Before we take up the topic of memory in Part III, which also involves time scales, I would like to show how the model can capture three easily recognized spatially explicit examples in which affect and probability change on different time scales. Obviously, many other examples will come readily to mind.

Case 1: Daily Grind

We've all (presumably) had the following experience: we begin the day in a perfectly good mood, go to work, have a lousy day, and come home in a rotten mood. Can we grow this prosaic example? Yes. In Figure 43, Agent 0 starts the day at 7:00 a.m. in his pleasant yellow neighborhood. His affect is zero and his appraisal of the probability of annoying demands is also zero. (Figure 43 also shows the *NetLogo* Interface with its user-adjustable sliders).

Then, as shown in Figure 44, he spends an aggravating and aversive nine-to-five day at the office (located in the upper right quadrant), where he is peppered by annoying demands (orange events). Within hours, his expectation of further annoyance and his aversion (affect) increase until, at quitting time, he is in an absolutely foul humor.

He arrives home, in Figure 45, where he is utterly free of harassment. He knows (since located in the yellow zone) the likelihood of further badgering to be zero, so his *P*-value drops to zero. But unless extinction is very fast, he is still in a foul humor (high *V*) when he arrives home. (Maybe his disposition to have a drink even exceeds his threshold!)

Exactly the same run can be interpreted variously.

Case 2: Emergency Responder

For example, one could interpret this as a story about first responders who enter a burning building—a terrifying experience during which the probability of being burned is high in proportion to the frequency of flames (orange squares). Once out of the building, the responder knows that the probability of burn injury is zero, but this fact does not extinguish the fear, which endures.

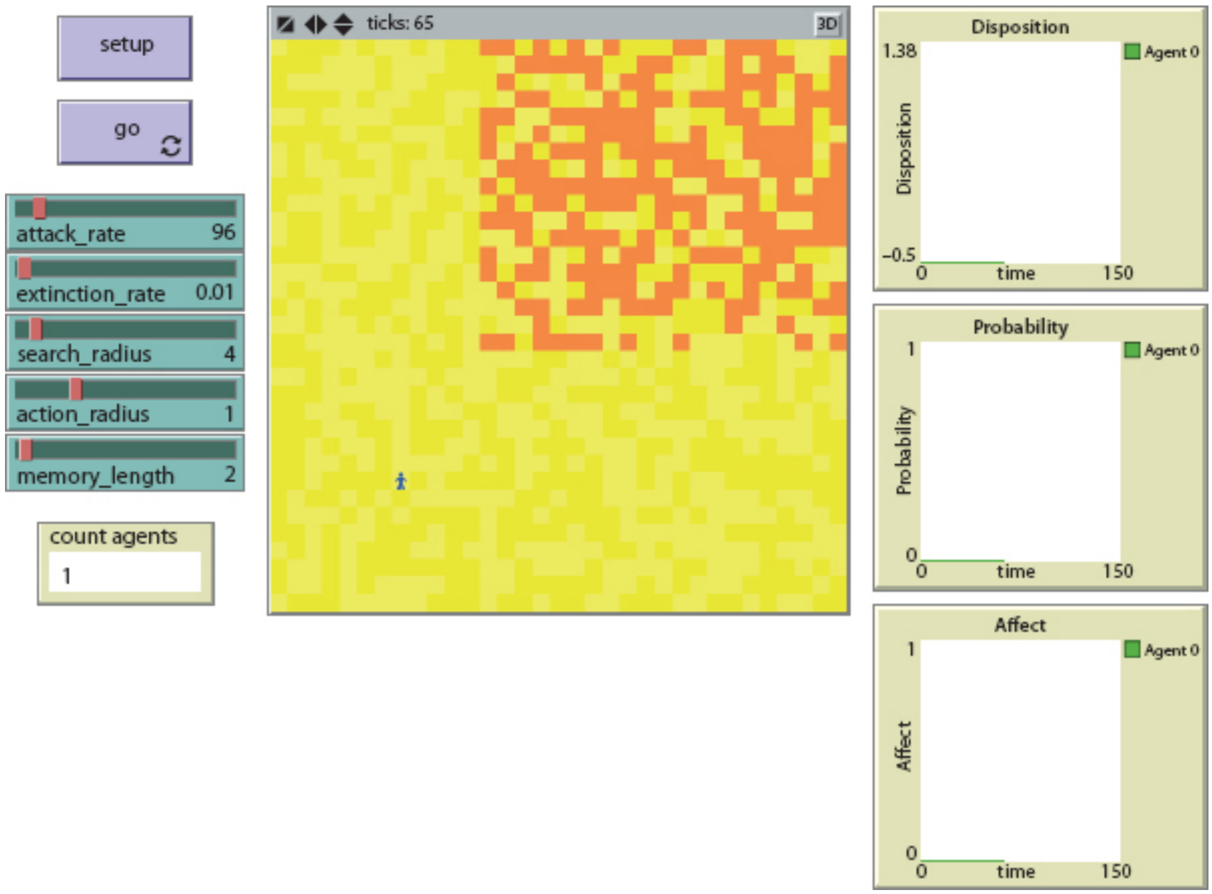

FIGURE 43. 7:00 A.M. Morning Coffee

Case 3: Combat

As the most extreme example, one thinks of entry into a war zone with orange bursts as enemy fire. During battle, fear and the probability of being hit are at their maximum. Upon withdrawal from the field, the probability drops to zero, but the posttraumatic stress can endure.

In all three cases, our agent begins the story in the placid yellow zone and in the affectively neutral state: that is, $v\,(0) = 0$, as shown in Figure 43.

After 150 periods, he ventures to the northeast quadrant—variously interpreted as rife with annoying office demands, flames, or enemy attacks. In this zone, both his affect and his estimate of the probability of aversive events quickly rise to high levels, as shown in Figure 44.

At period 400—quittin' time—our protagonist departs this zone and heads back to home/base. He recognizes that the probability of further adverse events is zero (the sample probability curve falls to zero). But, this is insufficient to reverse his bad feelings, due to a low extinction rate, yielding the results of interest. As shown in Figure 45, Probability drops to zero, but the aversive affect persists.

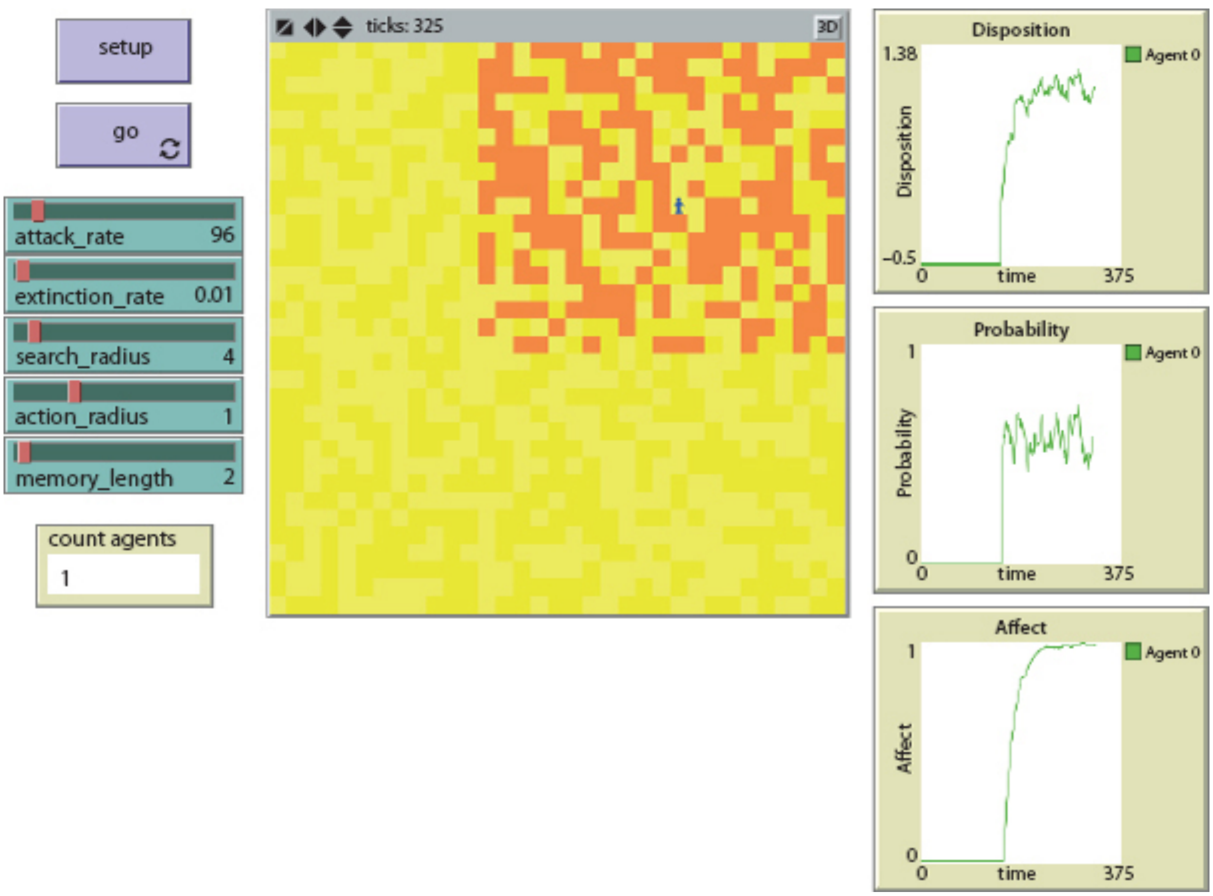

FIGURE 44. Nine to Five: Rising Demands During the Day

Of course, as discussed briefly before, by situational conditioning, the agent will come to associate the workplace itself with aggravation. So, *Agent_Zero* is already aggravated (or afraid, depending on the interpretation) when he walks in the door.

Case 4: A Happy Day

Clearly, the preceding fire and combat interpretations would involve fear and the amygdala, among other regions. But the same general associative learning *model*—though not the same brain regions—could apply to happy days, where one's disposition to break out in song is low at the start (Figure 43). So, suppose Agent 0 leaves home in the southwest for her college reunion somewhere in the northeast. On campus, happy singing breaks out all around (the orange outbursts of Figure 44). Agent 0 is rather shy (has a high sing-along threshold) so would never join in, except that her two best (high-weight) college friends (Agents 1 and 2) join in. Their dispositions to sing have weight, so she joins in. Finally, the party ends, and she heads home. And yet, even as the probability of direct musical encounters is zero, she remains aglow and sings the old college songs all the way home, as in Figure 45.[140]

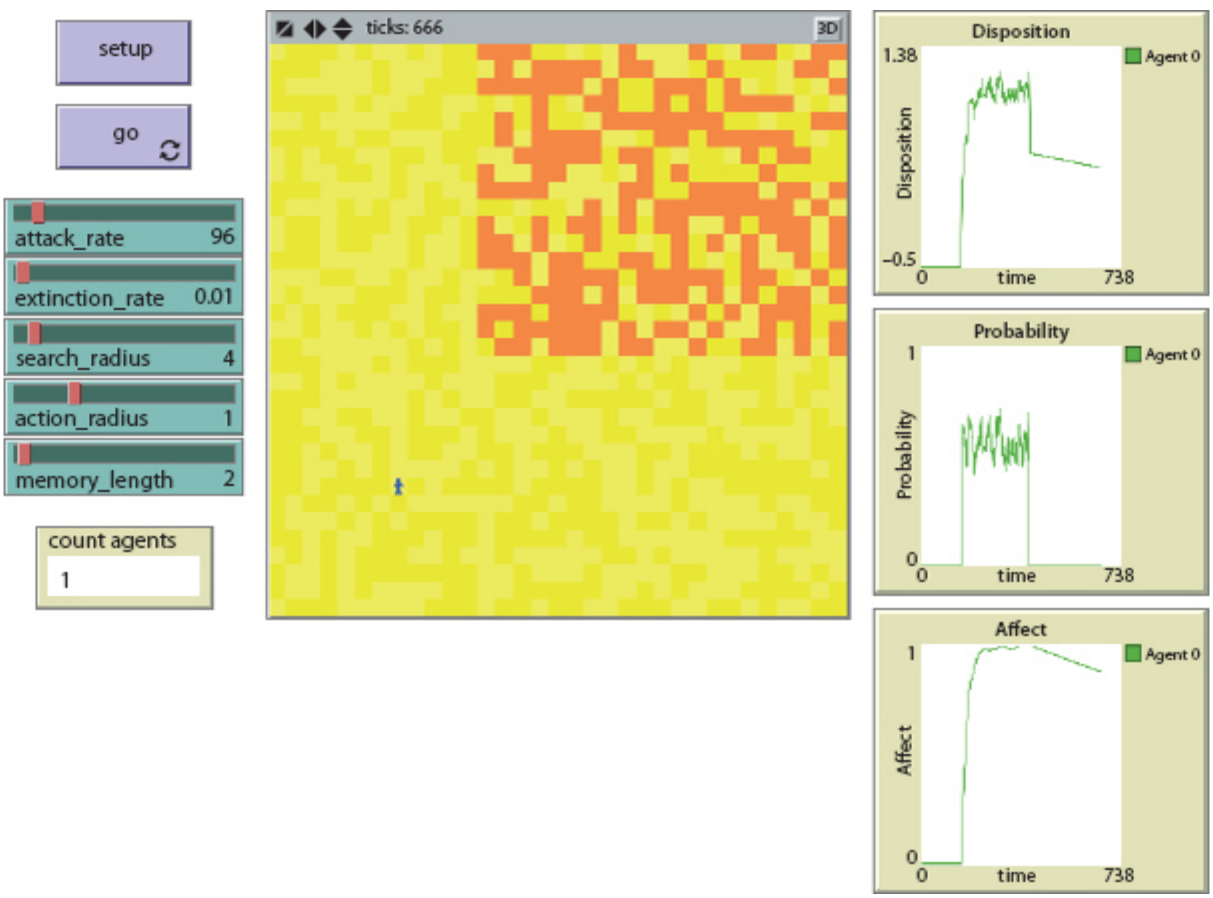

FIGURE 45. Direct Stimulus Stops, but Affect Continues

We have been exploring cases where *Agent_Zero's* sample probability rises and falls abruptly with his or her location in space, but her affect persists long after stimuli (trials) end. It would be interesting to devise cases in which affect evaporates before evidence does. We will do so below, when memory—along with much else—is introduced in Part III.

However, perhaps we have shown that the unadorned basic model —the basic *Agent_Zero*—does generate the intended central parables and much more. Specifically, most modeling focuses on extreme events. But the everyday life of people is equally worth modeling and, like the cases we've just developed, can be seen to "ring true" in the model, which is a good start.

Another game one can play with the base model is to explore the effect of one person's deficit on other individuals in her network. Earlier, in connection with posttraumatic stress, we used the mathematical version of the model to explore how one individual's experience affects others. Now, using the agent-based version, we will (I believe for the first time) *lesion* an agent and see the result, not only on her, but on others.

Run 5. Lesion Studies

My limited exposure to the literature suggests the utility of a purely logical dissection of the claims one might make about the amygdala and lesions.

Logic and Lesions

Ever since Klüver and Bucy's (1937) path-breaking work with primates, it had been conjectured that disabling the amygdala virtually eradicates fear. Recalling the rat's apparently hard-wired fear of even cat urine, "Large amygdala lesions dramatically increase the number of contacts a rat will make with a sedated cat.

In fact, some of these lesioned animals crawl all over the cat and even nibble its ear, a behavior never shown by the non-lesioned animals" (Davis and Whalen, 2001). More recent lesion studies—or contemporary studies using animals with genetically engineered deficits, such as "knock-out mice"[141]—establish that disabling or eliminating the amygdala indeed eliminates fear (along with much else). So, recognizing many nuances, just for logical precision, let's write this as[142]

$$\neg A \rightarrow \neg F.$$

If no amygdala, then no fear.[143] It follows logically that where there is fear, there is amygdala activation.[144] That is,

$$F \rightarrow A.$$

I have never understood why nature should ever respect our paltry rules of deduction,[145] but this is an observed regularity also (LeDoux 2003). Neither of these entails that excitation of the amygdala causes fear … that is, that

$$A \rightarrow F.$$

Lesion (or knockout) studies alone show *necessity, not sufficiency,* in other words. However, a history of experiments has shown that, "In humans, electrical stimulation of the amygdala elicits feelings of fear or anxiety as well as autonomic reactions indicative of fear. While other emotional reactions occasionally are produced, the major reaction is one of fear or apprehension." (See Davis and Whalen, 2001, and references cited there.)

While granting, then, that there are experimental grounds for an inference that $A \rightarrow F$, as a general proposition, this is equivalent to

$$\neg F \rightarrow \neg A.$$

which clearly *fails* since fear-inducing stimuli (e.g., snakes) are not the only inputs stimulating the amygdala (*A*). For example, erotic nude pictures and loud music can activate it[146] (Holland and Gallagher, 1990). So, we are not yet at the point where, given a subject's imagery (even accompanied by many other readings), we can infer their emotional state, or self-reported feeling, if there even is one!

In sum, while the amygdala does exhibit high functional specificity for fear (Kanwisher, 2010), the amygdala is *not* the only brain region involved in fear (Lindquist et al., 2012), *nor* is it the case that the only stimuli that activate the amygdala are fear inducing.

Lesioning *Agent_Zero*

Obviously, lesion studies on healthy humans are unethical. But lesion studies on software people are not (at least not yet). We can knock the amygdala out of *Agent_Zero,* as it were, and explore not only how it affects her behavior, but also how it affects the behavior of all others in her social group!

Here is *Agent_Zero*'s *NetLogo* amygdala, speaking very figuratively:[147]

```
[
if pcolor = orange + 1
  [set affect affect + (learning_rate * (affect ^ delta) * (lambda −
affect))]
if pcolor != orange + 1
  [set affect affect + (learning_rate * (affect ^ delta) *
extinction_rate *(0 − affect))]
]
```

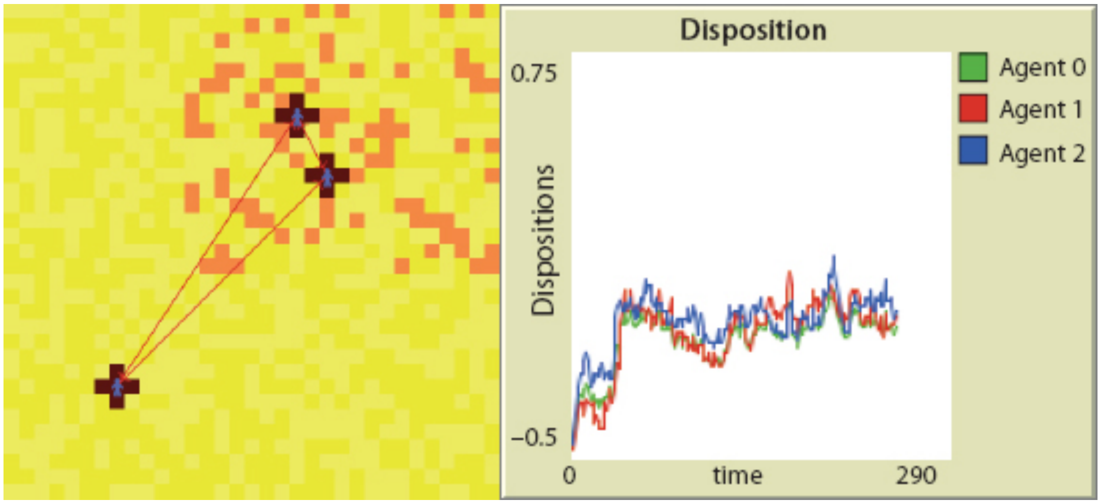

FIGURE 46. All Agents Normal

It is the agent's update-affect routine. In English, it says, "If the patch you're on bears the adverse event color, then set your new affect to your old affect plus the product of: (a) your learning rate, (b) your old affect to the delta power, and (c) the difference between lambda and your old affect. Otherwise (i.e., if the patch does *not* bear the adverse event color), do as before but replace (c) with the extinction rate times the negative of old affect, all of which follows the Rescorla-Wagner scheme.[148]

To knock out an agent's amygdala, we simply knock out this *NetLogo* Code block.[149] We are interested not only in how this lesioned agent behaves, but also in how her neurocognitive deficit affects the whole network. Depending on the agent's weight, the effect can be dramatic. One agent's deficit can have far-reaching ramifications. In Figure 46, we see a Run with all agents functioning normally.[150]

If we now knock out the amygdala of Agent 2 (the upper-right rover), it eliminates her fear (and her violence) and her transmission of fear to both Agents 0 and 1. Agent 1 (upper left) still acquires fear directly from events and transmits this to Agent 0. But lesioned Agent 2 is no longer contributing to Agent 0's fear (either directly or through Agent 1). As a result, the total fear acquired by Agent 0 is now beneath her action threshold, and she never engages in violence. This is shown in Figure 47.

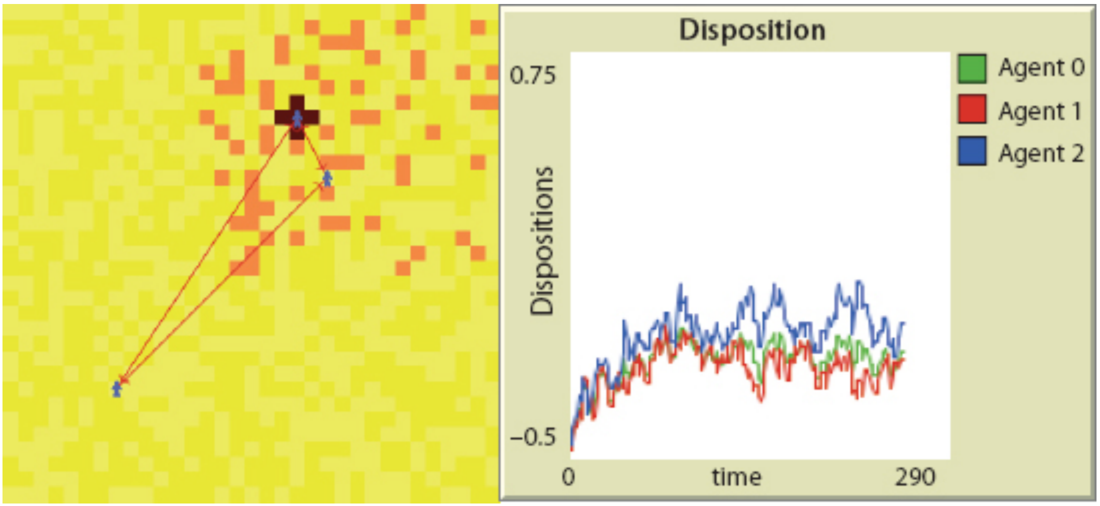

FIGURE 47. Agent 2 (Upper-right) Lesioned

Emotional contagion dynamics are affected if one agent's *direct* fear acquisition is disabled. But, as we now discuss, *observational* fear acquisition may also be impaired by amygdala damage.

Patient S. M.

The famous subject S. M. suffered from Urbach-Wiethe disease. In their classic paper, Adolphs et al. (1994, p. 670) write that her condition "caused an nearly complete bilateral destruction of the amygdala, while sparing hippocampus and all neocortical structures, as revealed by detailed neuroanatomical analyses of her computed tomography (CT) and magnetic resonance imaging (MRI) scans." The result was that S. M. was unable to recognize fear—and emotion generally—in the faces of others. To represent S. M. in the *Agent_Zero* framework, we would add to the disability just discussed the further inability to acquire fear *observationally* (as discussed earlier). Mathematically, this would be arranged by zeroing out the weight S. M. assigns to the affect of others.[151] In the social setting, this will further damp network transmission because she will not pick up emotion; and so she will not pass it on either. So, the damping social effect would be even more pronounced. That, at any rate, would be the hypothesis.

Generative Minimalism

The runs and discussions presented thus far involve no extensions to the basic *Agent_Zero* model. While the agent specification is quite minimal, considerable generative capacity has been demonstrated. While much more exploration of the basic model is warranted (and is easy given the Applets and Source Code posted on the book's Princeton University Press Website), we turn now to 14 significant extensions.

---

[118]A nice "Guide to Newcomers" is available in Axelrod and Tesfatsion, Appendix A of Tesfatsion and Judd, eds., *Handbook of Computational Economics: Agent-Based Computational Economics, Volume 2*. Among the closest things to a textbook on agent modeling is Railsback and Grimm (Princeton 2011). The best hands-on way to get started is to do the three excellent agent-based modeling tutorials that download with *NetLogo* (http://ccl.northwestern.edu/netlogo/).

[119]By canonical, I mean simply the base model for this development.

[120]A torus topology is readily available in *NetLogo* but would have been visually confusing for most of the runs explored here.

[121]*NetLogo* offers a wide variety of distributions from which to draw random numbers. Here, $U(0, 1)$, the uniform distribution on the unit interval, is used.

[122]*Update-affect* is the relevant *NetLogo* code block. See Appendix III. My code extends Rescorla-Wagner in allowing extinction rates different than the classical model, which, of course, is an available setting.

[123]Properly speaking, this extinction-rate slider is a multiplier. If it is 0, there is no extinction. If it is 1.0, we obtain classical Rescorla-Wagner extinction curves. Typically, we use a value in the interval (0,1). So, this is a second extension of the original model (beyond *S*-curve learning), permitting yet another type of flexibility.

[124]Hence the adjective "spatial."

[125]Later, we will interpret the set as a space of financial assets, a family of vaccines, or opportunities for unhealthy eating, over which a *local relative frequency* is being computed and updated.

[126]This is the number of orange patches over total patches within the spatial sampling radius.

[127]An anxiety-provoking context without question (Behrens et al., 2007).

[128]Lest there be any replicative or other confusion, the agent source code and *NetLogo* graphical output use the name *disposition* for *net disposition*. This should occasion no confusion. The *NetLogo* code block (see Appendix III, p. 218) governing this calculation is:

```
to update-disposition
  ask turtle 0 [
    set disposition affect + probability + [weight] of red-link 1 0 * ([affect] of turtle 1 + [probability]
    of turtle 1) + [weight] of red-link 2 0 * ([affect] of turtle 2 + [probability]
    of turtle 2) − threshold]
  ask turtle 1 [
    set disposition affect + probability + [weight] of red-link 0 1 * ([affect] of turtle 0 + [probability]
    of turtle 0) + [weight] of red-link 2 1 * ([affect] of turtle 2 + [probability]
    of turtle 2) − threshold]
  ask turtle 2 [
    set disposition affect + probability + [weight] of red-link 0 2 * ([affect] of turtle 0 + [probability]
    of turtle 0) + [weight] of red-link 1 2 * ([affect] of turtle 1 + [probability]
    of turtle 1) − threshold]
end
```

Terms could be collected in a variety of ways, all equivalent computationally but different conceptually. This form seems expeditious for expository purposes. *NetLogo*'s name for a generic agent is "turtle." I choose to imagine that this is in honor of a famous exchange between Bertrand Russell and an audience member who told Russell that the earth was supported on the back of a great turtle. Russell asked, 'And what, pray tell, is supporting *that* turtle?' The answer was immediate. "Oh, another turtle … it's turtles all the way down."

[129]As noted, agents can be in the same dispositional network even if they are not within one another's spatial sample radius. In such cases, communication (and dispositional contagion) could be by voice, by text message, by iPhone, by field radio, or other social media. Below we offer an ex-tension allowing one to change weights step-functionally when others enter (or exit) one's sampling radius. We do not exploit that in the main exposition.

[130]Later, I endogenize this radius as a function of affect.

[131]Technically, time is measured in *ticks*, a reserved word in *NetLogo*. In this model I advance *ticks* by 1 with each cycle through the *NetLogo* "go" routine, which corresponds to *main* in C or C++. In this case, the full "go" code is a follows:

```
to go
  if ticks > = maximum-stopping-time [stop]
  move-turtles
  activate-patches
```

```
    update-event_count
    update-affect
    update-probability
    update-disposition
    take-action
    deactivate-patches
    do-plots1
    do-plots2
    do-plots3
    tick
end
```

[132]For example, the primordial fire god parables have been displaced by the uncertainty principles of quantum mechanics.

[133]One might well say that Agent_Zero betrays himself in that his solo disposition is below his threshold, whereas his total (in the group) disposition exceeds it. In the nomenclature of the Introduction, $D^{tot} > \tau > D^{solo}$.

[134]As noted earlier, emotion and disposition can be communicated by numerous routes beyond immediate vision.

[135]It is important not to muddy the distinction between the spatial sampling radius and the distance over which dispositional contagion may occur. The two are completely independent in the model. Weights do not increase with spatial proximity or shrink with distance. An extension allowing this is offered under Future Research.

[136]Even 10% extinction alters this considerably. A little forgiveness, or counter-learning, can go a long way.

[137]Notice that this makes Agent 0's solo net disposition negative in fact, since it is v (here 0) plus P (here 0) minus τ (here 0.5). The others' dispositions begin negative but rise quickly with aversive stimulus. Notice also that we do not arrange the activation order by giving agents different thresholds.

[138]There is an asymmetry in the model as developed to this point. In defining the binary action to be X (equal to 1) rather than not-X, one induces a reference direction. In the cases just described it is positive. One acts when one exceeds the threshold, not when one drops below it. As we see, the solo dispositions of other agents can indeed move one's net disposition in a positive direction (e.g., from negative to positive). But, since solo dispositions are nonnegative, they cannot move net disposition in a negative direction, that is in a direction contrary to the reference direction. To permit this, various mechanisms present themselves. One is threshold imputation, which I intro-duce in Part III, to replicate the Darley-Latane experiment. Another would be to introduce negative weights. A third variation, for which I thank Julia Chelen, would be to have agents assign weight to the *average* of others' Vs and/or Ps. I also thank Jon Parker for discussions of this issue.

[139]See Tolstoy (1869; 1998 ed.)

[140]Less frivolously, the model captures cases where the individual has a good impulse but simply needs the support of others to act on it. I thank Julia Chelen for this observation.

[141]See Mayford et al. (1997).

[142]We employ the logic symbols ¬ (the negation symbol meaning *not*) and → (meaning *implies*).

[143]However, see Cunningham and Brosch (2012).

[144]If p implies q, then not q implies not p. Each implication is the so-called contrapositive of the other, with $\neg A$ playing the role of p.

[145]For example, nature respects Newton's second law, that $F = ma$. But, evidently, it also respects every proposition deducible from this law. But deduction is an entirely human invention. Why should nature select the deducible claims as the ones to which it will physically conform? I find that mysterious.

[146]This indicates that the amygdala is, in fact, not specialized to fear. It is implicated in many kinds of arousal.

[147]Again, I am not modeling brain regions.

[148]My code actually generalizes Rescorla-Wagner extinction slightly by introducing the variable named extinction_rate, which is a user-adjustable slider in the *NetLogo* Interface. If extinction_rate = 1, then the scheme is exactly the classical Rescorla-Wagner model. If 0 < extinction_rate < 1, slower extinction trajectories can be explored. Typically, extinction_rate $\in [0, 1]$.

[149]This is the sense in which *Agent_Zero*, as a software object, is "modular," which is to make no claim whatever regarding the modularity (however defined) of the human brain.

[150]The slider settings in this case are: Attack Rate 25, Extinction Rate 0, Sampling Radius 4, Action Radius 1, Memory 1, with Seed 2, which are also given in the Table of Appendix IV.

[151]Technically, weight has thus far been assigned only to the sum of V and P.