This is a draft of Chapter 6, in the book outline

Philosophy in the Resource-Patterns Model of Life

by Richard Ober Hammer, July 2019

This chapter is divided into three major sections:
   6.1, postulates of the Resource-Patterns Model of Life (RPM) explained.
   6.2, models of mind suggested by RPM's postulates.
   6.3, language learning within the model.

6.1 Justification for Postulates of RPM

In Chapter 1, Section 4 we listed the assumptions upon which RPM builds. Now I think that a better name for those assumptions is *postulates*; hereafter I will usually refer to postulates. In this first section of Chapter 6, I will justify some of those postulates.

6.1.1 Basic Postulate: Living Things Exist in a Universe

This Resource-Patterns Model of Life starts in my experience as a human being. I find myself in life having senses, hungers, memories, thoughts, and hopes for the future. I am told and I believe that my life can continue only if I eat enough nourishment, and only if I avoid an array of threats to my life.

I have choices about how to act, and my performance in making these choices seems to affect my success. I would like to make these choices better. Often it seems, if only in retrospect, that I could make better choices.

Fortunately for me there are nutritional resources here on Earth, and I have been taught that I can fill my plate if I choose my actions appropriately. I also have many acquaintances and as far as I can tell I speak for them as well. We humans generally share such experience.

6.1.2 Postulate: Life Advances in Levels

6.1.2.1 Looking to smaller sizes

Now we take an important observation from biology. This has been spelled out in Chapter 4, and perhaps the reader should review that material. But, to review briefly, biologists believe that we humans and other large multicellular organisms grew on Earth from combinations of single-cellular organisms. A long time ago, perhaps a billion years, single-cellular organisms were the largest and most developed forms of life. Then through a long process which we seek to understand those single-cellular organisms combined in cooperating organizations which gave rise to us much-larger multicellular organisms.

Furthermore biologists report a still earlier advance which created the single-cellular organisms mentioned in the preceding paragraph. Those organisms, called eukaryotic cells, appear to be composed of numerous smaller and still more primitive components, called prokaryotic cells, such as simple bacteria.

Thus we see that life in our past has grown from one level to another. At the two higher levels (eukaryotic single-cellular organisms and multicellular organisms) we see that what we normally consider to be a single organism is actually an organization of a multitude of smaller and more primitive organisms. And the same sort of composition from earlier "living" parts may also apply to the lowest level which I have mentioned, prokaryotic cells.

One can speculate that such disaggregation could carry on to smaller and smaller levels. But I will not attempt such a claim. For our purpose here it seems sufficient to notice the two smaller levels which immediately precede our present level as multi-cellular organisms. From those two we may:
  (1) formulate some outlines of how life as we are given ability to observe it has grown to its present stage;
  (2) propose that such level-to-level growth continues now as we humans, and perhaps other multi-cellular organisms, organize ourselves in groups.

6.1.2.2 Looking to larger sizes

I do not suggest that every organization which we humans form constitutes a new Living Thing (LT), because most of our organizations are temporary and fall apart before long. Consider a parallel between organizations at our human level and organizations of single cells at the first smaller level. I suppose biologists could point out numerous organizations of single cells which do not constitute a new multi-cellular organism. Single cells may form millions of organizations among themselves, giving each of these organizations a "test run" we may say, before any one such organization displays all the attributes which we humans might recognize as an individual living thing at a higher level. Similarly on our level, I suppose we humans will form many, many organizations that fall short of the sort of viability which might be necessary to form a new LT on the next higher level.

Of course we will notice many types of organizations if we start to examine and compare them. I have proposed one possible [way to sort organizations into eight types](). This is based upon whether the subject organization has attributes which may be called: member-aware, self-aware, and encoded.

A skeptic may doubt that large organizations composed of many humans might be viewed as individual living things such as we humans experience ourselves to be living things. After all, we humans are conscious, it may be argued, implying that no other organization could enjoy such awareness. I have tried to answer such skepticism in Chapter 4. Let me add here that I believe our most advanced science still has not grasped what exactly consciousness is. Lacking ability to clearly define consciousness, as we do, I would be humble about any confident assertion that other entities must lack it. See more about consciousness in Section 6.2.1.2.

6.1.3 Assumption: Prosperity Is Good (PIG)

When we are working with LTs we normally assume that prosperity (a higher standard of living) among those LTs is a good thing, and that an increase in population along with an increase in prosperity is also a good thing. I am a human, a LT, and my immediate impression is that my prospering, and the prospering of other people, is good.

People who focus upon sustainability may disagree with the PIG assumption. But whether we RPM modelers agree or not, motives to increase prosperity may have been programmed into us through Darwinian selection. If those motives occurred in some individuals by natural variation, and if those individuals succeeded in leaving more offspring, then we see the evolutionary mechanism by which a bias for prosperity may be felt in most of us today.

This assumption gives shape to the whole thrust of this book. Throughout we seek to understand how new sets of rules, which enable prosperity, can be discovered and learned. We assume it will be a good thing if we can understand processes of discovering new rules and thus augment our human prosperity.

6.1.4 The Second Law of Thermodynamics

6.1.4.1 Postulate: We continue to live because we keep on finding new Resource Patterns (RPs).

A cornerstone of RPM comes from physics, as the second law of thermodynamics places limits upon life. The second law tells that perfect sustainability is not possible, on this Earth or within any bounded ecology, for a Living Thing (LT) or any population of LTs. It is not possible to recycle 100% of the energy and raw materials which a LT requires to stay alive. So for a LT (or any population of LTs) to continue living it must consume new resources, energy and raw materials, which come from somewhere outside the LT (or population), and eventually from outside the previously assumed bounds upon the ecology which hosts such a population.

At first look this physics of the second law confirms the views of doomsayers. In every snapshot of life which we may take, we can see that the LTs in this snapshot are consuming resources which are clearly finite in this snapshot. It would seem that life such as this cannot go on indefinitely. Such a doomsayer's view is easily supported from observation and logic.

But another way of looking at the history of life on Earth shows no such gloomy outcome. We humans are the crowning product of that history, such as we like to see it, and we seem to be doing quite well. Our numbers are larger than ever before and each of us on average lives better than ever before. Yes, we use up the supplies of some resources. But we always seem to find other ways to satisfy our needs.

What could explain this seeming contradiction of the 2nd law? Here is what I assume must be the answer. Resources which we humans discover how to employ have been available here on Earth all along. But we did not recognize the potential value of such resources until we reached a level of technical sophistication which enabled us to exploit that value.

So, while a doomsayer may grasp onto the despairing long run implications of the 2nd law, a pragmatist may say that we humans should perhaps be more humble. We need not look to the everlasting and entire-universe-encompassing view. It may be enough, the pragmatist may continue, if we learn from the last three billion years of life on Earth in order to look ahead for only a shorter time, say one million years, as we anticipate our life in only this one galaxy.

There is of course no guarantee that our good luck in gaining access to new resources will continue. Ecological doomsayers might be correct. But in the sky I see the Sun and Jupiter, vast amounts of energy and raw material which we humans have yet to learn how to exploit. I assume that our human good fortune in increasing prosperity will continue for the foreseeable future.

6.1.4.2 Postulate: The universe must be patterned, or we would not be here.

The next implication which we draw from the second law is that the universe must be patterned, or we wouldn't be here. We have seen that a living thing can keep on living if from time to time it discovers and imbibes a supply of each resource which is essential to maintain it's life. This discovery of a resource will in many cases involve a LT moving to the location where a resource exists. As such, the LT must be either:
    (1) lucky, in happening to move where a resource lies at hand, or
    (2) knowledgeable, from either
        (a) experience, in remembering the location of an unexhausted resource or
        (b) education, having been taught where to seek a resource.

But can it be only luck (as just listed above)? We might imagine some circumstances in which a LT could survive with luck alone. This could occur when the resource was all around, predominantly present, so that a LT would just happen to bump into it so often that it would almost surely never run out; it would not need knowledge.

But it seems survival can't rely upon just luck alone. Notice for instance that we assumed motion in the just-described situation where a LT comes into chance contact with a resource frequently enough to allow survival. But if an LT has an option to remain idle, that is to make no physical movement in any given cycle, then such an LT might adopt a strategy of remaining idle most of the time – and might very well starve to death. So in order to survive our LT in this example

must not adopt that stupid idle, strategy, and must "choose wisely" to adopt a strategy of sufficiently frequent moving about.

We living things need to employ some knowledge of how to behave in order to arrive at those circumstances in which we are able to partake of life-essential resources. Said another way, we must adopt behaviors to be able to partake of enough life-essential resources in the environment in which we find ourselves.

Now, if you accept that conclusion, I believe that a corollary is that life-essential resources must exist in patterns in our environment. The opposite of patterned distributions would be random distribution. But if resources were distributed randomly then we living things could not make use of knowledge to exploit those resources; random behavior would be just as productive as knowledge-guided behavior. Since we know, I hope I can claim, that our knowledge serves our survival, I hope that I have your concurrence that our corollary is also true: Most of the life-essential resources upon which our lives rely are distributed not randomly but in patterns, patterns which enable our knowledge-guided behavior, Resource Patterns.

Here is another point which follows the above reasoning. Wherever we see a living thing we will probably be correct to assume that there must be one or more patterns of resources in that LT's environment which sustain that LT. Or, there must at least have been RPs in the past sufficient to have developed and maintained the LT up to the time when we saw it.

6.1.4.3 Postulate: Living Things which we see must be "doing right".

Since continuing life requires behavior patterned to exploit whatever RPs exist in the given ecology, we may conclude that LTs which we observe must, in an overall sense at least, be doing the right things in order to survive. This general rule will not apply to every observation of a LT during a prescribed span of time, as shown by the following exceptions to the general rule:

- A LT which has recently sated itself on a large meal may be able to survive a long time on its internal storage of resources.
- A LT may be young, having been born or launched into life with sufficient internal storage of resources to survive a relatively long time.
- A LT may be only one of a large number of offspring produced by a parent (or parents), in a dispersion of young into the environment in which only one or a few members of this dispersion need to "take root" and survive for the long term, in order to support the larger-view of life of the species of the parent(s).
- A LT may have been provisioned by a supporting organization to go on an exploratory mission during which it is expected to find no resources for itself, but only to bring back information which may prove helpful to the supporting organization.
- The RP which has sustained a LT for a long time may have been finally used up in a recent moment. So while this LT appears to be doomed, it is still surviving on previous productive behavior.

While such exceptions do exist we must be careful to allow the possibility that LTs which we observe may be living thus, in a temporary and unsustainable way. But still, generally speaking, if we observe a LT (or a set of LTs) over a long span of time and a variety of circumstances, at

some point as we extend the scope of our overview we must be observing, in this scope, behavior aligned with the RPs at hand.

6.1.4.4 Postulate: The universe's Resource Patterns must exist in a range of scales.

Recall that we see living things on at least three levels, and apply the conclusion we reached above in Section 6.1.4.2: that the existence of LTs strongly implies RPs which are exploitable by those LTs. We are thus encouraged to suppose that there are patterns of resources on different scales. Prokaryotes, before the time when many of them organized into eukaryotes, must have been sustained by exploiting one or more Resource Patterns (RPs) on a scale accessible to them. The same would apply to single–celled eukaryotic organisms. There must be larger resource patterns, or more difficult to exploit RPs, which gave rise to and sustained these larger organisms. Similar reasoning applies to the next higher level, to us multicellular organisms. There must be in our environment RPs with size or difficulty which placed them outside the reach of single cellular organisms, but which we larger organisms can exploit. Thus, evidence and reasoning support our conclusions that RPs exist on a scale appropriate for us multicellular organisms and smaller.

But what about larger-scale organizations? Can we assume larger resource patterns exist in our environment which may sustain the organizations of us humans which we are constantly testing? Of course there may be debate among us as to whether such larger RPs exist. If those larger RPs do not exist then all multi-human organization will fail. But there seems to be plenty of evidence that larger RRs do exist. Notice that for most of our recorded history several types of our human organizations (families, business firms, religious organizations, and states) have been forming and lasting long enough to suggest the existence of a sustaining RP. Recall the argument developed above that we may assume existence of a RP which sustains any long-lasting organization — whether or not we can perceive that RP. Furthermore, we can easily perceive large and difficult RPs which we have not yet gained organizational sophistication to exploit. Notice the sun, plenty of energy; notice Jupiter, plenty of raw material. To date we have used almost none of these resources. They wait for us, assuming we reach an appropriate scale of size and sophistication.

Concluding, we assert that RPs exist on a wide variety of levels, both smaller and larger than our human scale.

6.1.4.5 Postulate: Living Things have goals.

Here I will argue that LTs must have "goals", or at least the behavioral choices of most LTs much of the time should suggest that they have goals. We have noted that LTs need to adapt their choices to the patterned availability of resources in their environment. They must do so in order to discover resources — in order to survive. So if we see a LT we can assume that it has in the past, unless it has been temporarily lucky, acted as if it wanted to discover resources. We could reasonably judge that it has at the very least this one goal of discovering resources.

If we are modeling LTs with computerized agents, what I am here calling a "goal" will be represented by rules which the agents use to select their actions. For more on that subject see Chapter 5, "The Learning of Rules".

While we can expect a goal of survival to be evident in the behavioral pattern of most LTs much of the time, this goal will not always be superficially evident to an observer. When an LT is doing well, when it has an abundant store of life-essential resources, then the best chance of long-term survival may be served by experimental behavior. Experiments may increase the chance of discovering new as-yet-unexploited RPs. Such a discovery may increase the chance of survival either when presently exploited RPs are used up or in event of some future calamity.

But experimental behavior, perhaps by definition, will include some failing behavior. So outsiders, evaluating such observed behavior, may judge this failing behavior negatively. Failing behavior may be justifiable in the light of a given organization of LTs, when the larger picture and long-term future are considered. Taking chances in order to discover good RPs may necessitate some failures from behavioral choices.

While I hope I have convinced you that individual LTs must have goals, we should not necessarily expect all LTs in a given population to have entirely the same goals. Various goals which led to experimentation and dispersion may help a population of LTs in the long run.

6.2 Models of Mind: Overcoming the 2nd Law of Thermodynamics

As we have reviewed above, the 2nd law of thermodynamics challenges us to explain the existence of Living Things (LTs). We meet this challenge in RPM with a theory which offers to explain the increase in material order in some locations. Those locations, being the bodies of LTs, can be built and maintained with localized gating or control of the down-gradient flow of matter. Of course some order is dissipated overall in each interaction (according to the 2nd law), but properly chosen interventions can create, for a time, locales of increased order. The set of choices implied by these properly chosen interventions becomes the challenge which life must meet. Since we LTs exist, we may infer that life does indeed overcome the challenge. It is a challenge of information processing, a challenge of "mind".

Within our modeling method, RPM, we divide this information processing challenge in two. The division takes place along the line established by our definition of a LT.
 (1) On the smaller side, by defining a LT we give only a general overview of information processing which must go on within the LT. So we modelers are challenged to produce a more detailed and insightful explanation of the information processing within these pre-existing LTs.
(2) On the larger side, we modelers present an information-processing challenge to the LTs (and ourselves) in the initial condition: how can these LTs advance themselves through coordination of their activities in order to exploit a large but difficult RP?

We will address each side of this division below, in Sections 6.2.1 and 6.2.2 respectively.

6.2.1 Information Processing within Existing LTs

6.2.1.1 Flow diagrams of information processing

In this section we will look at how we might model a mind of one of our starting-level agents, by stepping through a series of diagrams which show an increasing level of detail.

For our first step we will simply observe a LT and notice that it seems to be living. It is producing observable actions. See Figure 1. This Figure shows one of our familiar critters, but for this discussion we will generally be thinking of our more general class of LT. This shows what exists, what we hope to model.
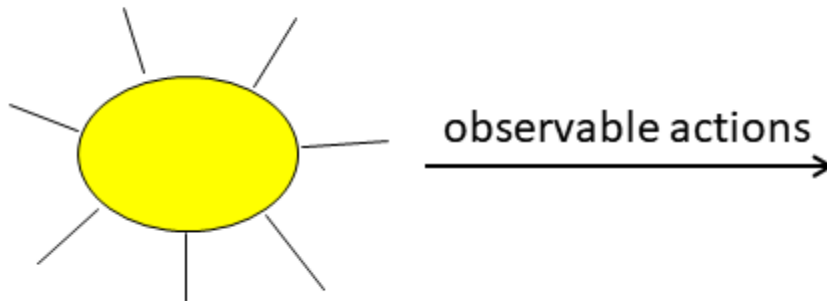
Figure 1. The motions of a LT, not entirely predictable, suggest internal information processing.

Recall the discussion in Section 6.1.4.3. Since we do observe a LT its life history must comply with the 2nd law of thermodynamics. This LT in our view must be doing something right, or many things right. Its survival to this moment shows that it (or its parents or its parent-community) is not acting randomly or otherwise stupidly. As we proceed in sketching a model of this LT's interior decision making process, we should remember that we are trying to capture how this LT does things right in the world in which we see it in order to survive.

We notice that the LT depicted in Figure 1 does not always do the same thing. At different times it does different things. So we can be hopeful that this variability of action gives another clue of the 2nd law compliance, of something environmentally-smart about this LT.

Next naturally we notice circumstances in the surroundings of the LT. See Figure 2. We notice that these circumstances change from time to time. So we may propose that the different activities of the LT are responses of the LT to its different circumstances. As we start to study this possible correlation we try, naturally, to list the LT's external circumstances which we suppose may be influencing its choices.
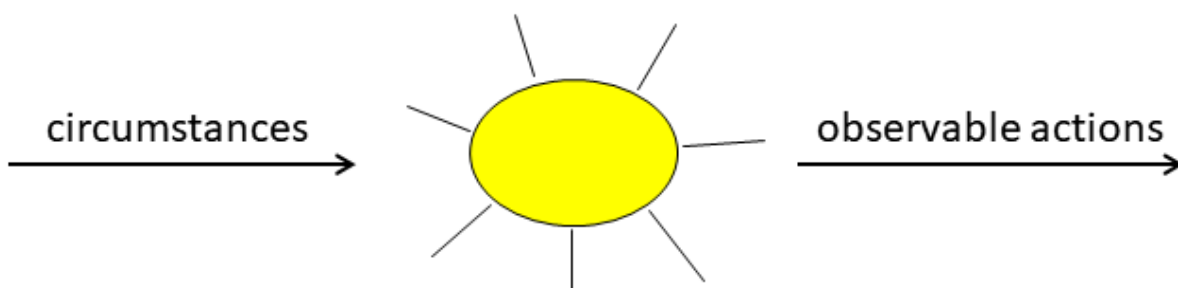

Figure 2. In trying to explain the LT's motions we try to catalogue affective circumstances.

The next step in our effort to model how the LT is deciding how to act may be to start building a model of the information processing system which we suppose must be operating within the LT. For our purpose of modeling the LT's information processing system, we may at first view that information processing system as a black box. See Figure 3. Now the circumstances (of the LT in Figure 2) come into the black box as sensory inputs. And the choices (actions of the LT in Figure 2) come from the black box as its outputs. Our use of a black box here shows that something, perhaps very complex, goes on inside the black box.

Figure 3. We call the LT's information processing a Black Box.

Next we will start to diagram what we might suppose goes on inside the black box. We use diagrams of a sort used by computer programmers to show their top-level description of how a program operates. Three preliminary points should be understood:
- With these diagrams we are modeling the mind, and only the mind of a LT. Other parts of the LT which we assume exist, but which we do not specify here, are one or more senses (abilities to detect aspects of both the external and internal environments) and one or more abilities to act (to make some physical or bodily motion).
- These diagrams should be helpful for thought experiment agent-based modeling (TEABM) modeling as well as for computerized agent-based modeling (CABM). If we work at the shallower depth, of thought experiments, then we might proceed with our thoughts codified no more explicitly than as shown by diagrams such as these. If, on the other hand, we work on the more difficult CABM depth, we will still probably use diagrams such as these as our starting points, that is as our top-level description of what we aim to accomplish with more detailed computer programming.
- We should not forget that many agents at a given level may have among them many different designs of minds. That complexity may be addressed in the future, but at this early stage I will be presenting only one prototypical design as our starting point.

In Figure 4 we model what happens during one moment of the life of a LT by starting at the top and making one pass down through the five boxes. And, since an arrow loops from the bottom box back up to the top box, we model a continuing stream of moments which extends through the entire life of the LT.
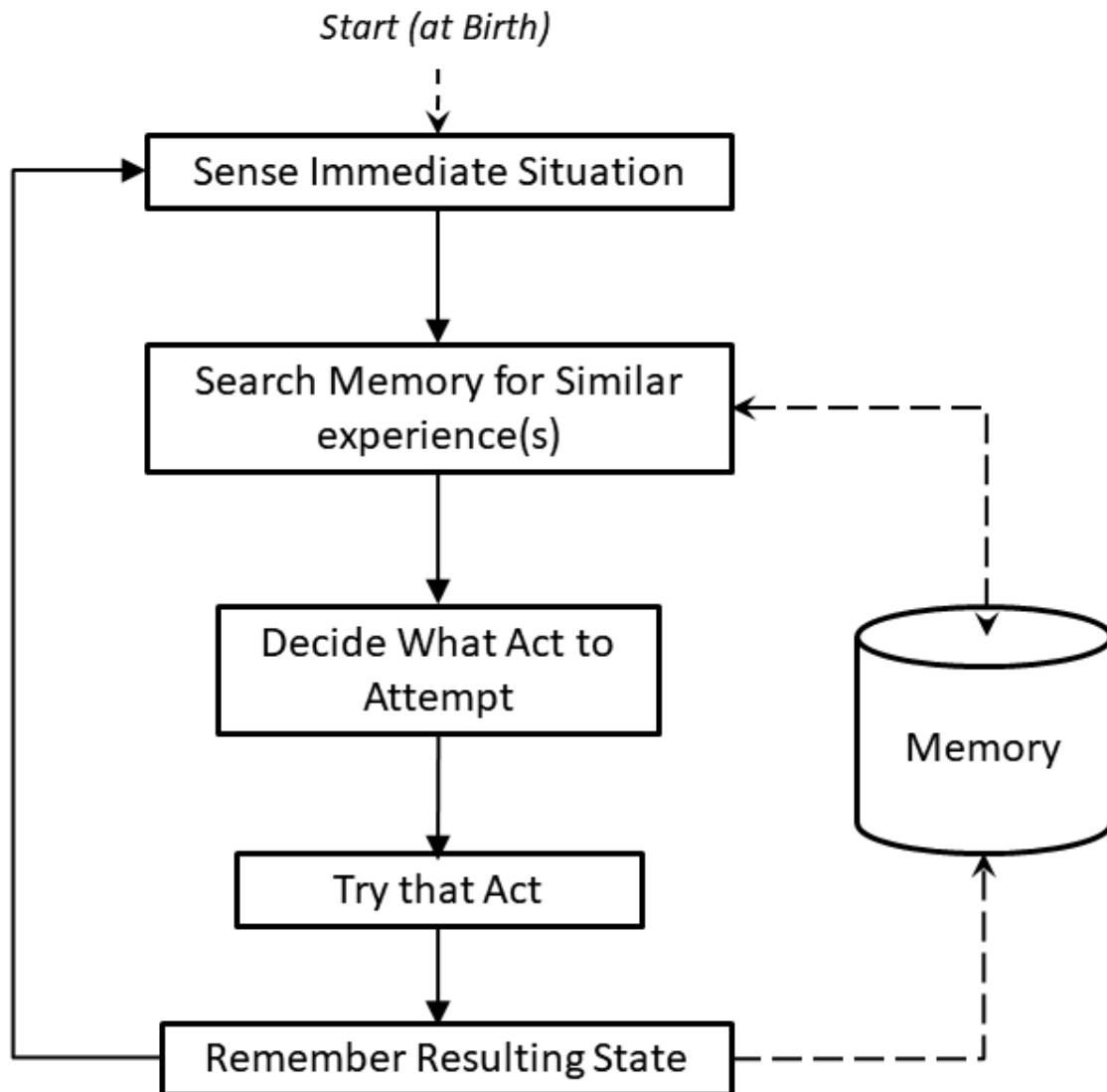
Figure 4. We break down a LT's information processing into these five steps during each moment of life.

In this diagram we include the important concept of learning from experience. We give memory to this mind, showing memory in the shape of a drum since that shape represents a database in computer diagrams. The LT can learn from its prior successes and failures.

The memory of this LT might be empty at the start of life, that is on the first pass down through the boxes. In that case memory could not help with the first choice. But in every subsequent moment there will be some memory, and increasingly more memory, which the LT may use as it attempts to choose an act which might bring it success given the current situation.

We say the LT decides what act to "attempt" rather than to "perform" because such a decision may fail. As the LT tries to perform the chosen act the outside world also acts. For example a decision to step forward may be blocked by sudden insertion of some physical barrier. Note that doing nothing, that is to make no outward move, may be among the acts which we build our LT with ability to attempt during a given moment.

We should keep in mind, as we go forward, that in order to survive the LT needs other things in addition to what we are now diagramming. These other things include:
1. The mind needs a good set of pre-programmed instincts or behavioral rules. These are needed in addition to memory in order to make productive decisions about how to act. The most important built-in rules aid survival by giving high priority to avoiding danger and finding food.
2. The environment must contain a pattern of resources which make it possible for a LT, such as we are creating it, to discover and exploit this pattern, that is to survive.

Next, we will use a diagram which adds explicit display of two functions which were only implicit in Figure 4's broadly-described "Decide What Act to Attempt". See Figure 5.
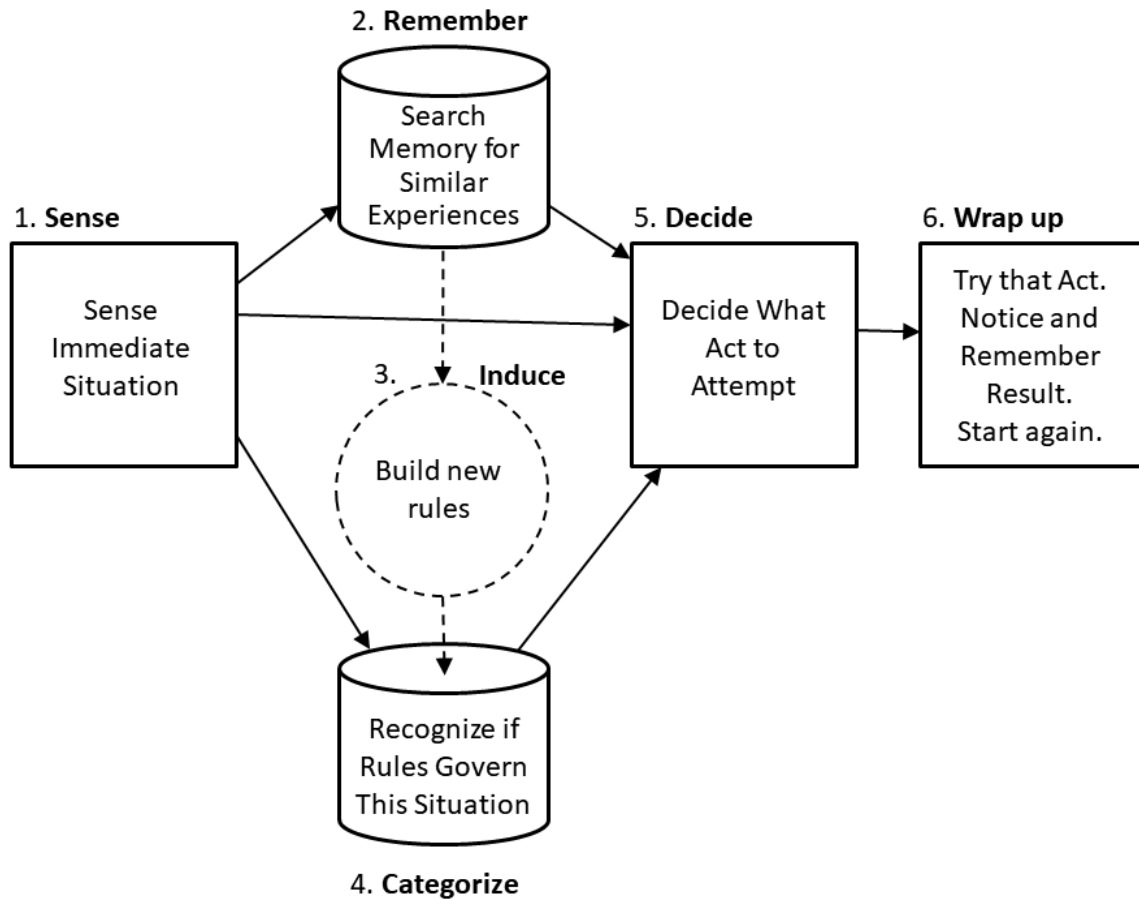
Figure 5. We extend the model by adding the possibility of learning new rules.

The two additional functions are:
- Box 3 Induce, shows an ability to build new rules after repeated experiences. This function operates more slowly, over extended periods of time, and is drawn with dotted lines to signify this difference of pace.
- Box 4 Categorize, which is a memory for storing and searching rules which may pertain to the situation of the moment. These rules, which may either suggest or require a particular action in response, are passed on to Box 5, Decide.

A line is drawn directly from Box 1 Sense to Box 5 Decide to allow immediate responses, like reflexes, in situations where there is no time to remember and consider anything.

It may be natural for a reader at this stage to ask what is the difference between the rules implicitly operating in Box 5 and the rules delivered from Box 4 to Box 5. For now we can speak only in general terms about this distinction while we remember that these box diagrams are only a beginning toward analyzing a process which, when more fully modeled, may require hundreds of boxes for display in such graphical form.

Finally, in this sequence of increasingly complex diagrams, we will add a box which represents planning ahead. See Figure 6 in which we have added, to the boxes in Figure 5, Box 5a Plan. The

function performed in this box is to imagine the future. Before the LT decides in Box 5 what act to take in the present, this new Box 5a shows that the LT can guess what will happen if it makes a particular act.

Figure 6. We add ability to imagine a future, to plan ahead.

The outputs of Box 5a would resemble the outputs of Box 1. While the outputs of Box 1 represent the external and internal environments, as the LT is able to sense those environments, the outputs of Box 5a represent those environments as the LT is able to imagine them.

To produce a guess about what would happen in the world following this LTs imagined (or hypothesized) act, the process in Box 5a may consider that the external and internal environments would differ from the immediate situation, as recently sensed in Box 1, in ways such as these:
- the world will differ insofar as this LT's hypothesized act (the input to Box 5a from Box 5) will have been performed.
- other LTs may have reacted to this LT's act and thus changed the condition which this LT senses.
- physical processes underway in the world may have continued during the elapsed time between cycles, and the situation changed accordingly.
- one cycle's worth of the LT's internally-stored resources will have been consumed.
- and more. This list could be extended of course.

Because the output of Box 5a attempts to represent a future state of the world, Figure 6 shows the three output arrows from Box 5a feeding into the same boxes as the three output arrows of Box 1. After the processes of Boxes 2 and 4, results are fed back again into Box 5. There, in light of these imagined results, a decision may be made on what act to attempt, with this decided act sent to Box 6. Or such a decision may be postponed while another imagined act is sent to Box 5a, to consider another step ahead of planning.

6.2.1.2 An Explanation for Consciousness

I discover a possible explanation for consciousness in the preceding development of flow charts for the mental process of an existing LT. Notice these processes which coexist at Box 5 in Figure 6.
- What comes in: sensory inputs, memories, awareness of rules.
- What it can do: Imagine an immediate future which may follow from the next act of this LT, and also a long term future involving a series of acts by this LT.
- What it must do: Decide what act to undertake next.

I suggest that the combination of those processes describes what I experience as consciousness. In addition to offering an explanation of consciousness this way of thinking offers a definition of consciousness, being the combination of processes described above.

John Searle gives a similar definition (in a YouTube lecture titled Professor John Searle: Consciousness as a Problem in Philosophy and Neurobiology, Cambridge, 2014, near minute 5 of the 53 minute recording). Transcribed here:
> Consciousness consists of all of those states or feeling or sentience or awareness. It is a set of processes that begin when you wake from dreamless sleep and it continues all day until you go to sleep again or drop dead or go into a coma or otherwise become, as we would say, unconscious.

Most other theorists who write on the subject of consciousness dive in without offering a definition. Probably consciousness is very hard to define unless we take this approach of a simple coincidence of processes.

Notice that this combination of processes in Box 5 seems to follow from our basic assumptions in RPM. If a LT is to survive it must process information in the way suggested for Box 5: Moment by moment the LT must choose how to act in response to its sensory inputs and goals. So, if all LTs have such a process, and if this is consciousness, then it would seem that all LTs have a cousin of the sort of consciousness which we humans experience.

6.2.1.3 An explanation for the experience of dreaming

Building upon the above concept of consciousness, we may also find a rudimentary explanation for dreaming within the processes depicted in Figure 6. To find this explanation we would assume that when a LT sleeps some, but not all, of the information processes shown in Figure 6 shut down. We would guess that sensory inputs shut down for the most part, and also the abilities

to act. But the imagining or planning function of Box 5a would continue to operate, along with the loops involving Boxes 5, 5a, 2, and 4.

6.2.1.4 Induction and Deduction in this model

Texts on philosophy and logic often mention induction and deduction. We may find rewarding insights if we search for these two ways of thinking in the model of information processing which we developed above in Section 6.2.1.1.

For readers not familiar with these terms, 'induction' refers to thinking from specific instances to general rules or observations. 'Deduction' is the opposite, thinking from general rules to specific implications. For example, suppose you move to a new region and during your first few months there it rains on Tuesdays but not on other days. You may think, "It always rains on Tuesdays but not on other days". This is induction:

(particular observations → general rule).

After that rule about when it rains has formed in your thinking, if you want to know if it will rain on a particular day in the future you may use the rule to make a forecast. You will think it will rain on that day if it is Tuesday. This is deduction:

(general rule → particular statement).

Deduction
First let us find deduction in our model. We start with this general point about computer programs which give us a way of modeling the information processing of LTs. Computer programs can normally be seen as performing deduction in their operations, because a computer program is a set of rules which are followed rigidly step after step to produce an output. Of course the inputs to the program also affect the output, as well as the rules which constitute the program. But still, once given the inputs, the output is then usually calculated deductively from working the rules of the program on those inputs.

Now, to find deduction in our particular model, refer to Figure 4, and consider the step "Decide What Act to Attempt". Here the program will follow rules in evaluating the present situation in light of relevant memories to produce a single decision on how to act. This is deductive process.

But to be careful here we should recall one of RPM's foundational assumptions, being that LTs act non-deterministically. This assumption might challenge our view that the process of deciding upon an act represents deduction, because normally we would think that deduction, with given rules and given inputs, would always produce the same output. But not necessarily. Suppose a situation in which preliminary calculations have suggested not just one act but a set of acts all of which promise roughly equal hope in these circumstances. For such a situation the program might select one act randomly from that set. With some randomness inserted into the calculations, the LTs behavior may be both nondeterministic and deductive.

Induction
Now we turn our attention to induction as it may appear in our model. At first we might think that no induction is needed in our model. No induction was mentioned, after all, in Figure 4. And

the calculation which we discussed above worked only with already-established rules (deduction) to produce its output. So we have not yet said why induction would be necessary or helpful.

Here is a reason for induction: Reality must limit the amount of information processing which a LT can do in a given moment. If memory has grown large with experiences then a large number of relevant memories, passed into the "Decide What Act to Attempt" step, could require a long time to compute. So if a LT needs to act quickly it will need some computational way to decide quickly.

For an example, suppose we have a critter as introduced in Chapter 2. It has only a simple mind such as in Figure 4. Suppose it has run dangerously low on sugar but now, in this moment, it senses that it has just come into contact with a portion of sugar. It faces no immediate danger other than the danger of starvation for want of sugar. It has to decide how to act in this moment.

Following the procedure which we have sketched above in Figure 4 our critter will:
1. search through all its memories, selecting those memories which resemble to the current moment;
2. evaluate the acts performed in those previous moments based upon the quality of the outcome at the conclusions of those moments;
3. wrap up this deductive comparison by selecting the one most promising act to perform in this moment.

This deductive process might take ten minutes, especially if our critter is mature and has accumulated many relevant memories. But there is risk in taking a long time to make this choice because in our critter's world it is possible that another critter may consume the sugar while our critter is thinking.

Notice that the process of combing through all memories might discover that our critter has chosen the same act in every one of the previous moments which resembles this present moment. If that is the case it becomes obvious, to us modelers at least, that the lengthy process outlined above can be cut short if our critter has a way to recognize the circumstances of this moment as calling for immediate choice of a single act. So our critter has a better chance of survival if we can give it a way to learn a new rule for action in circumstances where quick thinking is important for survival.

How will this induction be done? There are many possible ways, but we should note:
- induction is not strictly logical. One general rule to cover all possible experiences with given circumstances cannot be derived from only a limited number of particular experiences learned in those circumstances. As such the inductive production of a new general rule always gambles with the possibility that the rule will be discovered to be wrong by some new experience. Induction requires a leap of faith.
- variety in method of induction, among the members of a population of LTs, is probably good, since this variety gives the population a better chance of having at least one member which produces a good rule for the given circumstances. That one member at least may survive and its progeny may gain dominance.

We may reasonably guess that the members of a population which seems successfully established in a given environment have induced a set of rules which reflect the features of that environment fairly well. These induced rules enable the LTs to decide efficiently how to act in conformity with their environment. The induced rules are a sort of representation of the environment, a representation created for the purpose of enhancing survival of the host LT in its environment.

Now we have described the need for induction in a model such as Figure 4, so this motivates addition of induction in Figure 5.

6.2.1.5 Interior perceptual map

In this section I will introduce the idea that the mind of a LT carries a perceptual map of reality. The purpose (or use) of this map will be to give the LT an interpretation of its sensory input data. Given any particular set of inputs the map produces a map reading. Map readings help the LT to know where it is and what is happening, and to decide what it should do.

In Figure 7 we see a representation of how most of us most of the time conceive of our relation with the Real World. That is, we LTs are separated from the Real World, but we have a direct view of the Real World. We see the Real World as it is and respond to it accordingly, or at least it seems that way to us.
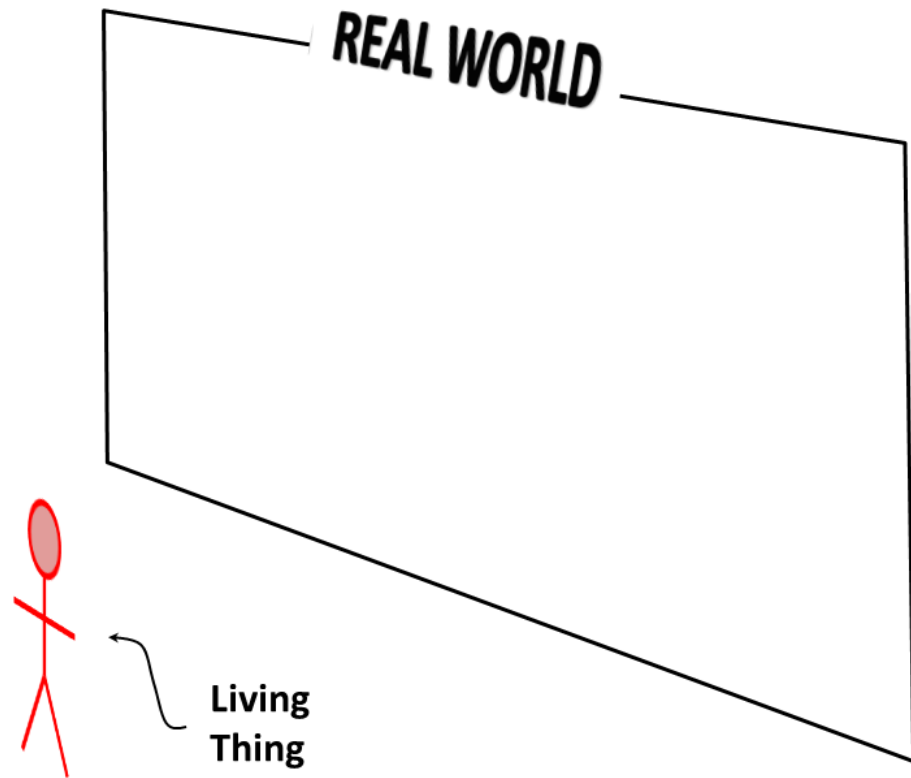
Figure 7. We LTs normally imagine we have unmediated exposure to the Real World.

But as we study ourselves and our world we need to make new distinctions. Now we will divide the information processing capabilities of the LT. In Figure 8 we see:
1. the LT's perceptual map of the Real World. It represents subconscious, automatic perceptions and judgements.
2. the conscious decision-making center of the LT which we will call the Homunculus.
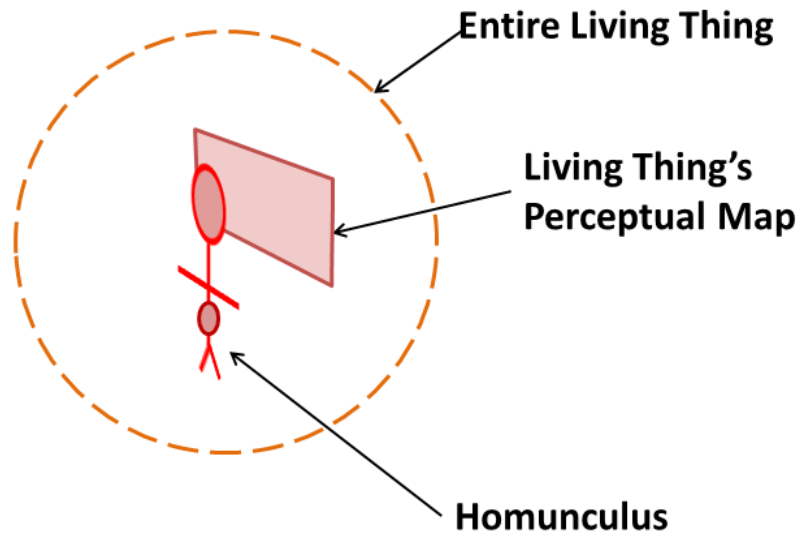
Figure 8. Clarifying the division in our nervous systems between perceptions and consciousness.

Now we move on to Figure 9 where we combine the ideas from Figures 7 and 8. On the left near to us we see the LT divided into perceptual map and homunculus. The homunculus can "see" only its perceptual map. All of the homunculus's inputs about the "real world" come from the perceptual map and not directly from the Real World.
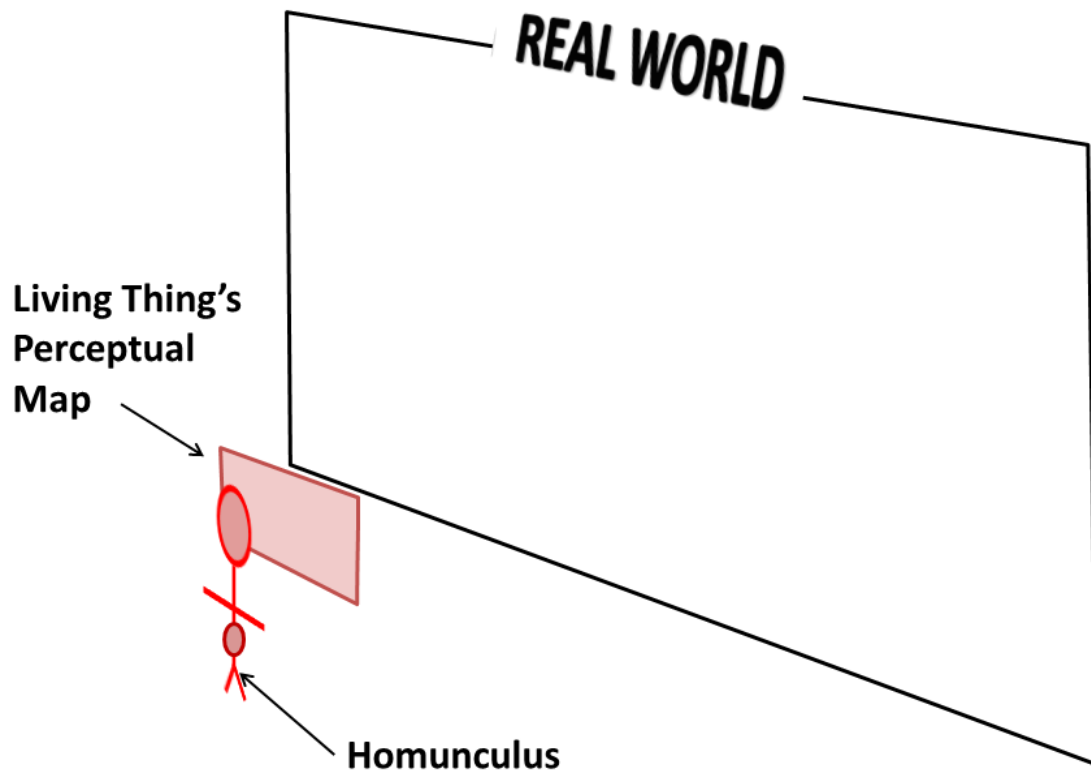
Figure 9. Showing our modelers' conscious recognition that a LT's perception of the external world is mediated by a Perceptual Map..

Some confusion may be caused by my use of the name "map" here, because in everyday human life we are conscious of using a map, such as we would be conscious of using a road map. But for our purpose in these discussions I want to separate consciousness, which resides in the homunculus, from all the nervous system processing which goes into perception, usually subconscious, which resides in the perceptual map.

You might also be confused by this division of a LT's information processing capability into only two parts, map and homunculus, since in Section 6.2.1.1 we divided this capability into as many as seven parts, as in the seven boxes in Figure 6. But these are two separate models of a LT's information processing capability, models invented for different purposes. Figure 6 may help a computer architect who is structuring the overall task of developing programs to mimic information processing in a LT. Whereas Figure 8 will help us as philosophers working with more general problems, such as the language-learning which we take up in Section 6.3.

Even though these two models represented in Figures 6 and 8 serve different purposes, we may gain some clarity by comparing them. How, we ask, do the seven boxes of Figure 6 relate to the two components of Figure 8? It seems clear that Box 1 Sense of Figure 6 would be mostly in the perceptual map of Figure 8, while Boxes 5 Decide and 5a Plan of Figure 6 would be in the

homunculus of Figure 8. But the other four boxes of Figure 6 (Remember, Induce, Categorize, and Wrap up) cannot be so simply thrown into one or the other of the two components of Figure 8. The computation performed in each of those four Figure 6 boxes is, in most normal circumstances performed in the perceptual map. But in certain demanding circumstances consciousness, such as I experience it, enables me to examine what goes on in those four boxes, to question and perhaps override the outputs of those boxes. So the computation performed in those four boxes may be divided, in abnormal circumstances, between the perceptual map and the homunculus in Figure 8.

To say more about the purpose of this map, it is, like any other attribute of a LT, to enhance survival and reproduction. A LT can survive if the map-readings from its interior map provide good-enough input data for the LT's decisions on how to act. Moreover those LTs which have the best maps for a given environment will probably produce the most offspring. After a few generations these "best maps" will be common in a population which, to our observation, seems to be surviving quite well.

Also we should be careful to think in evolutionary terms. The quality of the map is judged by success in leaving offspring. The quality is not judged by our human values. We cannot expect a LT to evaluate a given circumstance as we humans would evaluate that circumstance.

Perception is Interpretation
Most of us quite naturally trust our senses to bring us accurate impressions of our surroundings. And we might believe that these impressions, which we receive consciously from our senses, are true and unbiased. Our senses, we might believe, are free of any possible errors which might have been introduced by subjective interpretation of sensor data. But in what follows I question this presumed accuracy. We will consider only one sense, that of sight, but similar arguments can be made for other senses.

When I become conscious of seeing a face before me, that impression seems to come into my consciousness as a single, whole impression. But my perception of the face starts in my eyes, and we know that each of my eyes contains millions of sensor cells (rods and cones). Each one of these sensor cells is positioned on the retina so as to catch and respond to the light coming from a tiny fraction of the visual scene in front of the eye. So the single impression of a face before me must have been constructed somehow from the millions of output signals from individual sensor cells. A great deal of subconscious interpretation goes on in my nervous system between the outputs of the individual sensor cells and the single impression of a face which I experience in my consciousness.

Notice that my eyes sometimes deceive me. I see something – or think I see something – which turns out to be something entirely different when I have looked for more time or moved closer to get a better look. I assume that you have had a similar experience. What we can conclude from such experiences is that our visual perception works with sketchy data, with incoming light impressions which are insufficient to support a definite conclusion of the name or category which needs to be assigned to a pattern perceptible in the incoming light. But our survival as LTs requires that we choose acts based upon the best information available to us. An immediate, although possibly mistaken, identification of what we see will often enough give us an advantage

that we should not be surprised that our sense of sight interprets whatever data it has, giving our deeper, decision-making nervous process such an identification.

Pixels of Map
The map's representation of the world is not perfect, as we have just argued. We can use an analogy of pixels which, as you know, make up any modern digital image. The pixels in a digital image, while not representing the world perfectly, do hopefully give a sufficiently useful idea of the content of a scene for a viewer to derive a helpful understanding.

Considering the size of the pixels in this map, or the graininess of the representation provided by the map, the pixels will get smaller with experience. We might consider each new experience (each cycle of the model) as adding one pixel. Each new pixel in the map provides more detail in that area of the map. For an example which develops this idea see Section 6.3.4.1.

Notice that the addition of experience (of pixels) does not increase the size of the overall area represented by the map.

Concerning the overall area represented by the map, we will assume that the area of the LT's map covers its universe of possible experiences, as suggested in Figure 9. The map is there to help the LT decide how to act in any circumstance it might encounter. This universal coverage exists from the first moment of the LT's life.

But as I just described, the detail within the map improves with experience. And now we should clarify that the resolution of the map, that is the density of pixels, varies from one part of the map to another, depending upon the number of experiences in any sub-area of its world. The density of pixels will become high in the frequently recurring situations which a LT will encounter. On the other hand the density of pixels will be low in situations which the LT has never encountered, or encountered only a few times.

Ways other than experience to populate the map
Education: We should mention education as it may pertain to pixels in a LT's map. In RPM models more advanced than we have yet developed senior LTs purposely educate junior LTs in safe teaching environments where the seniors can impart education-pixels, which would have useful resemblance to experience-pixels, in the maps of growing junior LTs. A senior might say "I hope you will never encounter (such-and-such) a dangerous situation, but if you do, this is how you must react…". In this way a LT may be equipped to learn from the experience of its elders, with one or a few pixels implanted in regions of the LT's map, even for situations completely unlike all the living experiences of this LT.

Inheritance: We will also mention the inheritance of instincts or dispositions which we humans and other fancy LTs get from birth. Instincts and dispositions may also be regarded as pixels in a LT's map, pixels acquired not from direct experience but from some process at birth.

6.2.2 Information Processing Toward a Higher-Level LT

In the use of RPM, once we have established the initial condition, with a population of LTs surviving by feeding upon a barely sufficient RP, we turn our attention to discovering how members of that population can overcome the information-processing challenges of exploiting larger and more difficult RPs. This relates to our assumption that life advances in levels, as we developed in Chapter 4. The challenges of information processing for groups of LTs will be the subject of the next Chapter, on Public Psychology. Here I will make just a passing observation about these problems.

In a typical experiment in RPM we will try to show how the LTs of the initial condition can solve a single problem of coordination (for examples, see the challenges in Section 2.4). Then if we succeed, if we and our LT-agents can solve that problem, we will have demonstrated one of the abilities which a becoming-more-effective organization of the LTs will need. But this one ability must represent only a small fraction of all the advances which would be necessary for us to judge that the organization had gained all the abilities which, taken together, would lead us to judge that the organization had gained the status of a LT on the higher level. So in this way of modeling we may advance only one step at a time. We will need to make many such experimental advances before the combination yields a LT.

If that ambitious aim is achieved then we will be able to claim that our efforts have created the model of the mind of the new, higher-level LT. And, as we work on the humble steps toward achieving that aim, we can claim that we are building that model of mind.

6.3 Language

In the course of modeling how groups of LTs may discover modes of cooperation, RPM promises to show us many things about natural language, since language provides one of those modes of cooperation. By 'natural language' I mean a human language, such as English, as opposed to a computer language, in which programmers write specific instructions to be carried out exactly by a computing machine.

In RPM, the development of a set of mutually helpful signals may help our LTs coordinate their actions. Along the way we will encounter philosophical implications about the nature of language.

6.3.1 A Simple Agent-Based Language-Learning Experiment

I will start by describing a simple language-learning experiment (LLE), since this concrete example may help readers understand more abstract points I will make later on about language. Before we jump in, it is worth noting that this language-learning experiment is one which I have carried out on a computer. As such, this experiment differs from most of the other experiments which I describe in this book, since those other experiments have been carried out only as thought experiments (TEs), which require less specification of details.

Two simple agents, a consumer and a producer, exist in a computer program which runs in cycles. In each cycle:

- The consumer wants one of five commodities randomly, one of: wheat, oats, chocolate, beer, or nothing. But the consumer has no ability to get such for itself. The consumer can only act by displaying one of ten symbols: A,B,C,D,E,F,G,H,I,J.
- The producer can see the consumer's symbol and can produce any one of the five commodities. But at the outset the producer has no idea what the consumer's symbol might mean.
- Success is awarded to both agents when the producer delivers what the consumer wants.

Each agent has an internal memory which starts out empty but which remembers all past experiences. In each cycle each agent performs the following steps in sequence:
1.   notices its input (commodity wanted for the consumer, or consumer's symbol for the producer),
2.   looks in memory for previous experience with that input,
3.   decides upon an act by repeating a successful experience, avoiding an unsuccessful experience, or by acting randomly if experience offers insufficient guidance,
4.   performs the act and remembers the result, that is whether this combination of input and act succeeded or failed.

As you might guess the two agents in this experiment eventually discover a language through which they successfully coordinate their actions. In this language the consumer always displays a specific one of the ten symbols which uniquely correlates with its want, and the producer upon

seeing that symbol always delivers the commodity the consumer wants. This accomplishment, with success in every cycle, has usually been attained within a few hundred cycles in my computer runs.

In case you did not guess that the two agents would eventually stumble upon a perfectly successful language, let me explain. It happens because there are only a small number of ways that each agent may respond to each of a small number of inputs.

- If by luck the producer delivers what the consumer wants then both receive positive feedback. In that case each of the two agents remembers its experience, the stimulus-act combination which led to this positive feedback, and each will forevermore repeat that act when presented with that stimulus.
- But we must recognize that at the outset bad luck is more likely than good luck. In the first cycle of the model there is only one chance in five that the producer will deliver what the consumer wants. But each agent remembers each failure. In succeeding cycles each agent randomly selects an act from among a smaller number of remaining possibilities. Eventually it has to happen that the producer will deliver the commodity wanted by the consumer.

6.3.1.1 Language Implications

With this experiment, with the producer and consumer discovering a mutually supporting set of signals, we can proceed to enticing philosophical speculation.

First we will be bold and call that set of symbols which come to be used regularly a language. Admittedly it is a simple language, but it provides a good starting point for our research program into more realistic language-learning experiments.

Concerning the meanings of words, we should notice why each symbol (word) comes into use. This coming-into-use happens because a symbol helps both of the agents accomplish mutually beneficial coordination of their actions.

- The consumer is aided by knowing a symbol which, when it feels hunger for a specific item, it can display to the producer with confidence that the producer will then deliver that specific item.
- The producer is aided by knowing which food, among those which it can produce, will win a reward for it when delivered in response to a particular symbol from the consumer.

So the coming-into-use of a word has everything to do with the needs and abilities of the consumer and producer (the agents in the model), but has nothing to do with the needs and abilities of other agents – including notably us modelers – unless the needs and abilities of those other agents somehow translate within the model to needs and abilities of the consumer and producer.

Another lesson we may take from this experiment is that we should not be surprised by the growth of natural languages. This growth can be seen to flow from the RPM assumptions. We modelers can expect growth of natural language where we arrange the following:

- Make sure the agents have: (a) memory, (b) tendency and ability to repeat favorable acts and avoid unfavorable acts, (c) ability to display various signs to one another.
- Place agents in an environment where there is a large or difficult RP, exploitation of which is beyond the capacity of an individual but within the capacity of an appropriately organized group of the individuals.

If we structure the experiment appropriately we will expect the agents to discover cooperation which employs their signals. They should discover a cooperative resource-exploiting language just as surely as a computer could find a needle in a haystack if we give the computer sufficient abilities and enough time.

Creation of a new organization, but not a new LT
At the completion of this experiment, with the two agents acting together in harmony for mutual benefit, outside observers may perceive the pair of agents as two halves of a single entity. Whether an observer perceives the two trading partners as a single entity may depend upon the priorities of the observer or upon what the observer has been trained to see.

In the jargon of RPM we may call the cooperating pair an *organization*, which we see as resulting from any degree of cooperation, but not as a LT, which we see only if an organization has all the properties of a LT which we listed in Section 1.4.

Precise meanings are not realistic
In this consumer-producer language-learning experiment, the words have exact meanings. 'F' for instance might come to mean oats, exactly and for ever more. But, as we all know, in the real world our natural language terms can take on a variety of meanings. Later in Section 6.3.4 we will run through a language-learning experiment in which words can and will always have fuzzy or ambiguous meanings.

6.3.2 Language Stories

6.3.2.1 "Wha?", "Duh!" story

Now I make a point with a little story.

> During the night at Empower Designs the IT guy installed an update to the software used by the staff. First thing the next morning Rachel arrived at her cubical and, as she started up her computer, Bob arrived at his neighboring cubicle and likewise touched the start button on his computer. Soon Bob heard Rachel say "Whaaa?"
>
> A minute later Rachel heard Bob carry the conversation forward, "Duh!"

Both employees are having difficulty with something in their environment. Both have expressed frustration, but what is it about? Are they having the same problem? We do not know yet, of course, and neither do they.

But I think this exemplifies the way in which many of our productive conversations start: Using terms which are so broad as to be almost meaningless we reach out to someone who might have a problem similar to our own. As a conversation continues we find more specific terms with which to share our experiences.

I want this example to throw cold water on the idea that our language should be precise. Of course we strive to reach ever more precise understanding of our problems, but vague words and vague sentences help us find standing as we step toward clarity.

6.3.2.2 The ambiguity almost universally inherent in natural language

Consider a three-word sentence, for example "Bill hit Bob." Suppose this sentence exists in a context where each word has five possible meanings: there are five "Bills", five meanings offered in the dictionary for "hit", and five entities called "Bob". Then our three-word sentence has 5x5x5 = 125 possible meanings.

Suppose Susan sees something which she describes as "Bill hit Bob". What Susan saw aligns exactly in our context with one of the 125 possible meanings. The other 124 possible meanings describe something else — which Susan has no intent to convey. Suppose Alex hears Susan and his mind goes to work, subconsciously, trying to make sense of her three-word sentence. Alex's mind has 125 options, all of which might make sense to Alex at the conscious level. But the meaning delivered up in Alex's mind, from the subconscious level to the conscious level, will be the one meaning of the 125 which makes most sense in Alex's subconscious calculation. Suppose we ask Alex if he understands what Susan said. He might answer "yes" if his mind has been able to achieve a reasonable level of confidence in any one of the 125 possible meanings.

Will the one meaning which Susan saw be the same meaning which Alex comes to understand? Of course we do not know. We need more information to answer that question.

6.3.3 Formation of verbs

In this chapter we deal with nouns more than verbs, but we will speculate a little bit here about formation of verbs.

In Section 6.2.1.1 we have suggested a model of information processing for a LT that decides upon an action for a given circumstance by searching memory for actions taken in previous, similar circumstances. This search of memory might produce evidence that particular acts produce predictable effects. This change of state of the world might be represented as a triple:

prior state $\rightarrow$ my act $\rightarrow$ present state.

This triple has much of the semantic content of a verb. Verbs as such might be used in planning activity (Box 5a of our Figure 6).

A verb, as just suggested, might represent change in the world during a single cycle of our model, but a verb might also represent a sequence of changes brought about during many cycles. A sequence of steps may often be required to achieve a single objective during the life of a LT.

Imagine for example a village of primitive people live in huts in a clearing in a forest. From the clearing there is a path which leads through the forest down to a stream. As such the overall activity of 'go to the stream' may be broken down into three steps: go out of the hut; walk across the clearing to the path; walk down the path to the stream. This sequence, having been accomplished one or more times successfully, may be remembered as a single set of connected act-choices through which a single objective may be achieved.

Such a set of act choices may later come to be named among the villagers with a single verb meaning go-to-the-stream. Such a name could be selected almost arbitrarily from among signals available to villagers, since what matters is that agreement on the meaning can be discovered or created. The mapping between signal and set-of-act-choices can be invented on the fly by advanced LTs such as we imagine our villagers to be.

In this way a single word, a verb, can come into use and come to mean to each villager what it means to that villager. Of course different villagers may have different sub-acts, somewhat different connected sequences of act choices, to accomplish the same goal (go to the stream). For instance, occupants of different huts must take different routes from their huts to the path since each hut must be in a different spot. But success and reuse of the single word (verb) depends upon the success of the overall sequence and not upon detailed equivalence in the component sub-acts.

6.3.4 Formation of a Single Noun

6.3.4.1 Language-map thought experiment (TE) learning wheat vs. non-wheat
We will now build a thought experiment which suggests how we LTs might learn the meaning of a noun. In this experiment we will name our agent the "producer" since this agent has a role somewhat like the producer in our earlier experiment in Section 6.3.1. Sometimes our producer will be asked to select "wheat" from its environment, other times it will be asked for "non-wheat". It will be trying to learn the meaning of only this one word, wheat, as distinguished from everything else.

The world in which our agent, the Producer, lives
We give the producer a two-dimensional world from which it will be given opportunities to select objects. In this world objects are depicted by black dots and the character of an object, whether wheat or non-wheat, is determined by the object's location. See Figure 10. The area of wheat is shown by an oval. Everything inside the oval is wheat, everything outside is non-wheat. But only we modelers know about or can see the oval drawn on Figure 10. The producer cannot see that oval and will learn the character of an object only after making a selection. The producer will be trying to learn which area in the world correlates with wheat.

To keep this simple, the distinction shown by the line of the oval is absolute: Thus object A is definitely wheat, even though it is near the line, and object B is definitely non-wheat.
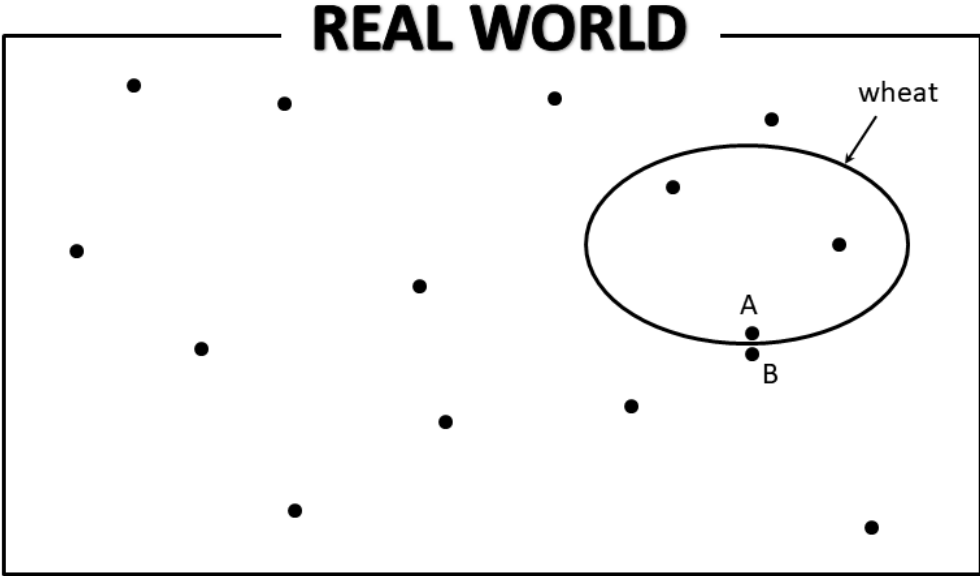
Figure 10. Producer's Real World with black dots representing objects, both wheat and non-wheat

We work with an image we call the map
We will use the idea that a LT has a map of its world in its calculating capacity and we will show how an agent might use its map to converge upon an understanding of the word "wheat". You may recall that when we introduced the idea of a map, in Section
6.2.1.5, we emphasized a distinction between the Real World and the LT's map of the real world. But that distinction, while crucial for some insights, would add unnecessary complexity to our current thought experiment. So we will use a map which is rectangular like the Real World and which we assume represents the real world perfectly enough for our purpose in this experiment. The map, which will keep a record of the producer's experiences, starts out empty as shown in Figure 11, showing no experience at all.

producer's map of real world
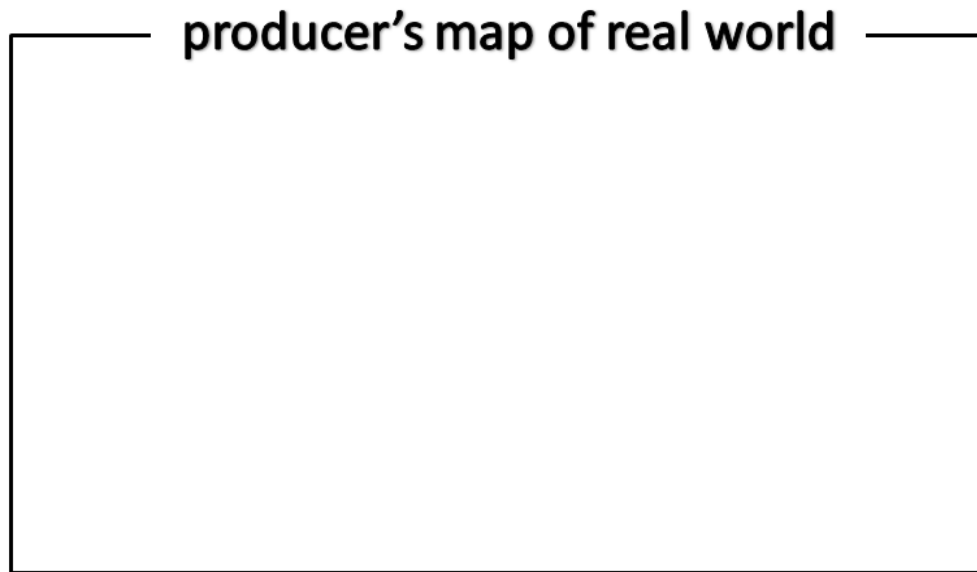
Figure 11. Producer's map at the start of cycle 1.

The cycle in this TE
Once again in this experiment our producer makes a choice during each cycle. At the start the producer has no experience and can hypothesize only that wheat may exist anywhere in its world. But as cycles pass the producer can form and then improve a hypothesis about which area of its world contains wheat.

Each cycle consists of these steps:

1. The producer gets a request for one of two types of objects: wheat or non-wheat.
2. The producer is given three options selected randomly (by the operator of the model) from among all the objects in the world. There is no guarantee that any one of these three may satisfy the request.
3. The producer may update its hypothesis concerning what part of the world contains wheat, if, given what it learned in the previous cycle, an updated hypothesis will help with the current decision.
4. The producer chooses one of the three options.
5. The producer learns the actual type of the object it selected. It writes this new knowledge into its memory for the location. Thus for the future it knows what kind of object to expect at this location.

Now we run the experiment
We will operate this model for a few cycles while we observe what is going on from the viewpoint of the producer's homunculus. That is, we observe the producer's map.

In the first cycle at the outset the map is void of experiences, as in Figure 11. We assume that the producer receives a request for wheat and then, as shown in Figure 12, the producer is given three choices among which it must choose one in hope of finding wheat.
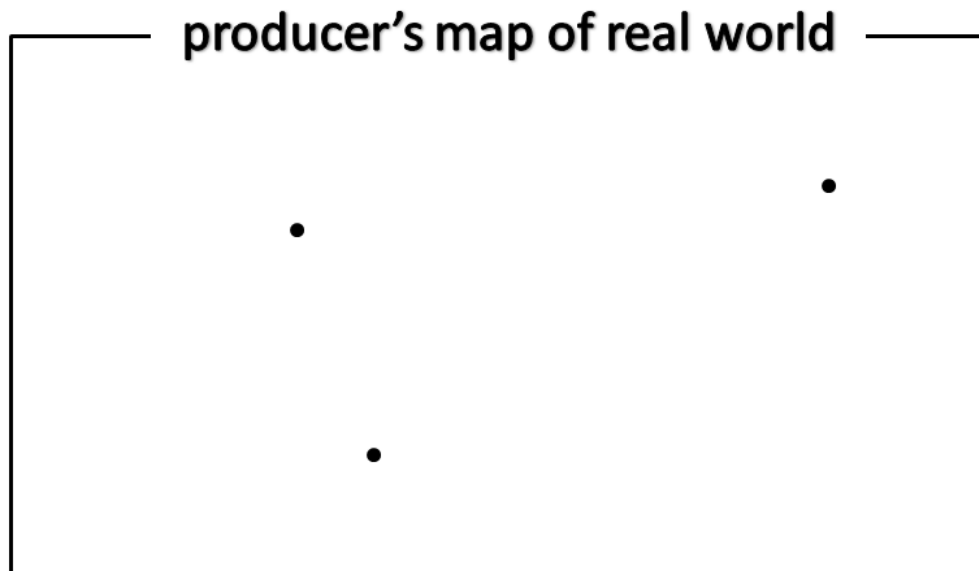
Figure 12. Three choices offered to producer in cycle 1.

We will assume that the producer chooses (guesses) the object farthest to its left and then learns that object has turned out to be non-wheat.

So after the end of the first cycle the producer's map will show one experience as in Figure 13. The producer knows that an object at the location of the red dot was non wheat. We will use red dots to show non-wheat experiences.

Notice that the two objects not chosen by the producer in cycle 1 are forgotten. There is now no trace of them on the map, since the producer has learned nothing about the objects at those locations.
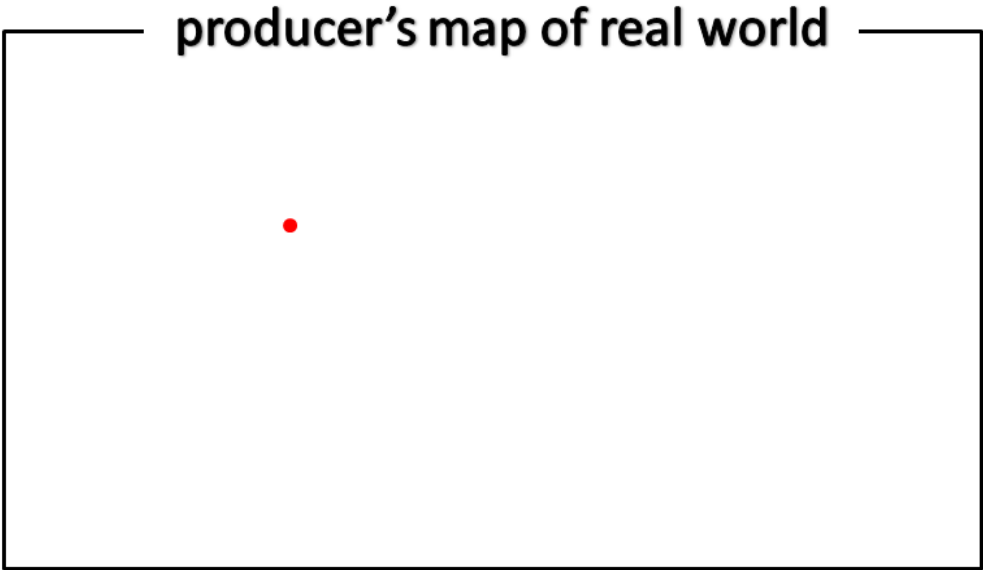
Figure 13. What the producer knows after Cycle 1.
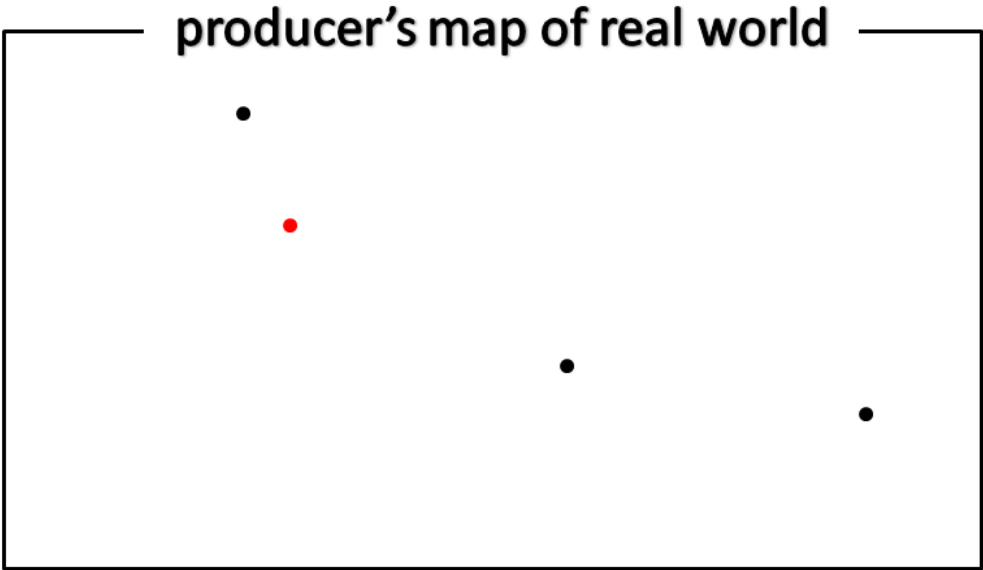
Figure 14. The producer's options in Cycle 2.

Now in cycle 2 assume that the request-symbol is "wheat" once again. The producer's three choices for cycle 2 are shown in Figure 14 as black dots, along with the single red spot of knowledge so far. The producer will try to use this single spot of knowledge to decide which of the three dots might be wheat. We will suppose that this producer forms a hypothesis — that its world is divided in half, down the middle, with wheat on the right and non-wheat on the left. See Figure 15 which shows a vertical dotted line where the producer guesses the division occurs. This guess makes sense of what it already knows and also helps it to decide what to do in the present cycle.
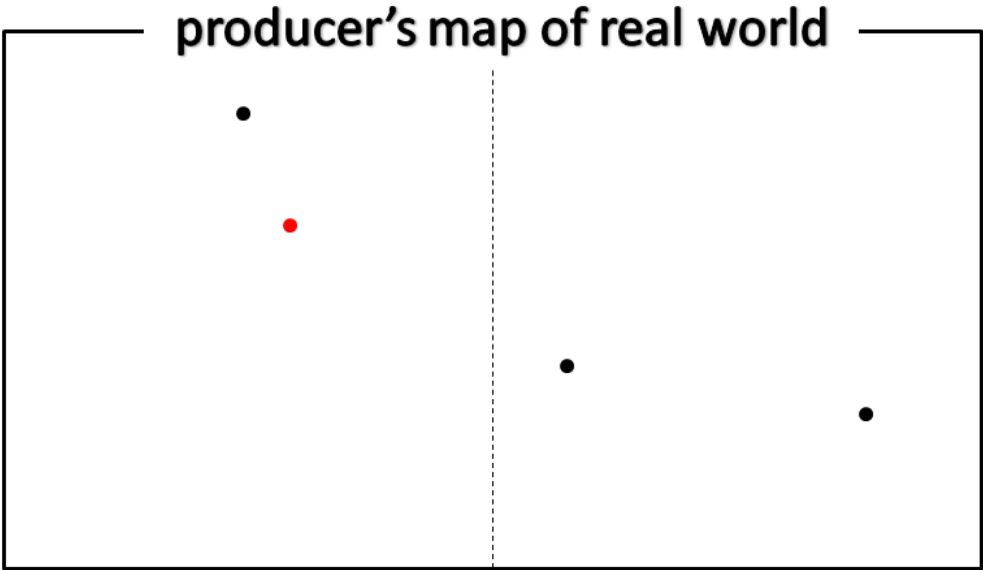
Figure 15. During cycle 2 the producer hypothesizes that non-wheat lies left of the dotted line and wheat lies right of the dotted line.

So the producer will choose one of the two objects on the right side of this dividing line. We will suppose the producer chooses the object at the far right. Once again the producer receives negative feedback, since this turns out to be non-wheat. So the producer starts out in cycle 3 with the knowledge shown in Figure 16. Each red dot shows that an object selected from that area was non-wheat.
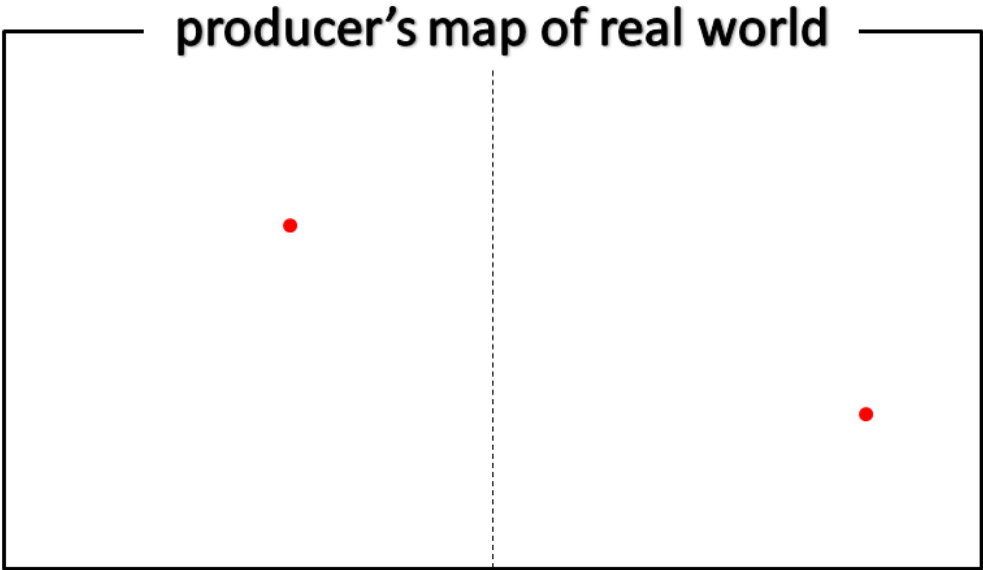
Figure 16. What the producer knows after completing two cycles.
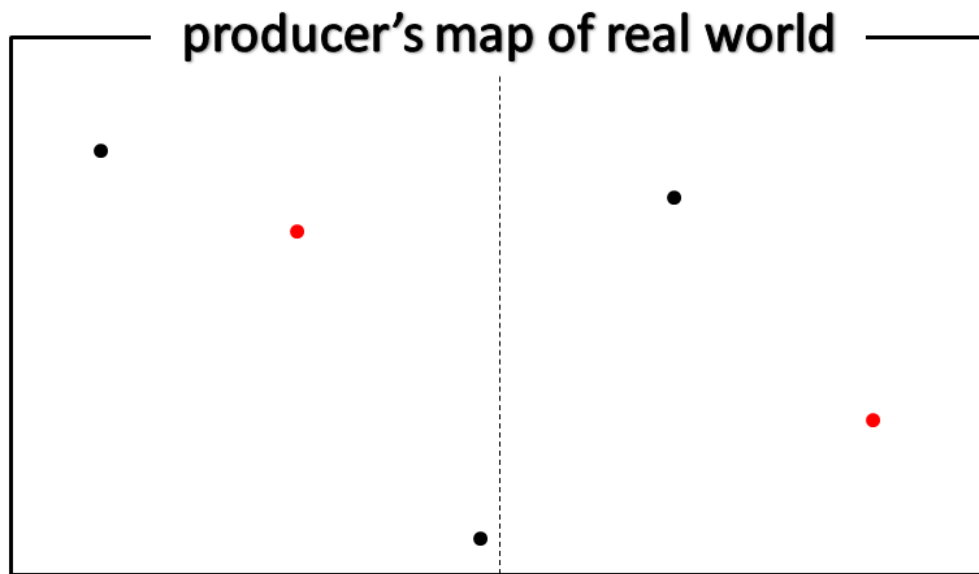
Figure 17. In cycle 3 the producer is offered the three objects shown with black dots.

In the beginning of cycle 3 we will assume that the producer is asked for wheat once again. The three choices offered in cycle 3 to the producer are shown as black dots in Figure 17. Figure 17 also shows the two red dots of non-wheat experience as well as the vertical dotted line representing its previous guess about the layout of its world.

Assume that the producer now guesses that only the upper right-hand quadrant of its world contains wheat. This quadrant is marked off with dotted lines in Figure 18.
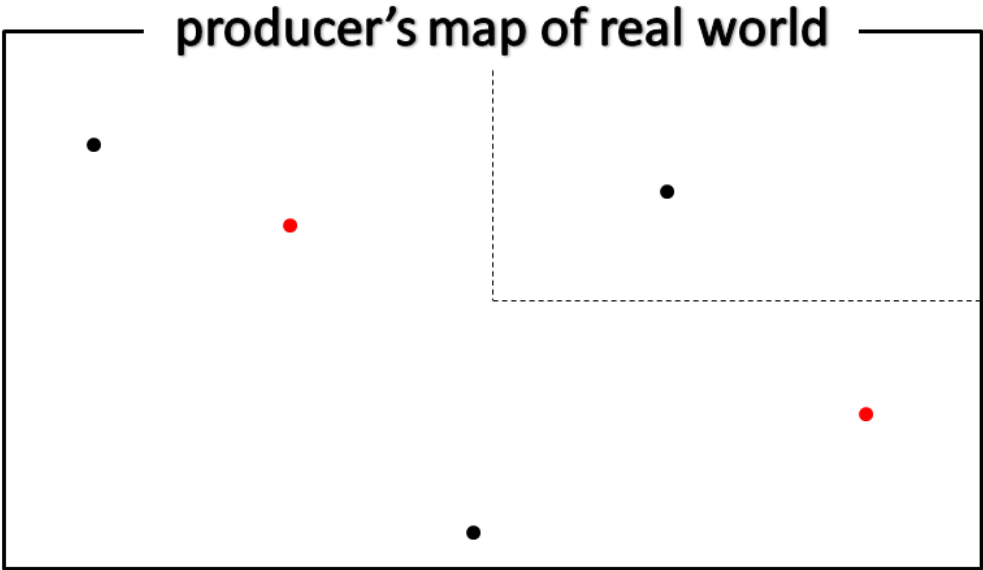
Figure 18. In cycle 3 the producer refines its hypothesis of which area (the upper right) contains wheat.

In conformity with this guess about the way things are, the producer chooses the one dot which lies in the upper-right quadrant. This time finally the producer has made the correct choice. It learns that wheat was found at the location chosen. In Figure 19 we show the producer's map as it exists at the start of cycle 4, with two red spots indicating where non-wheat has been found and one green spot indicating where wheat has been found.
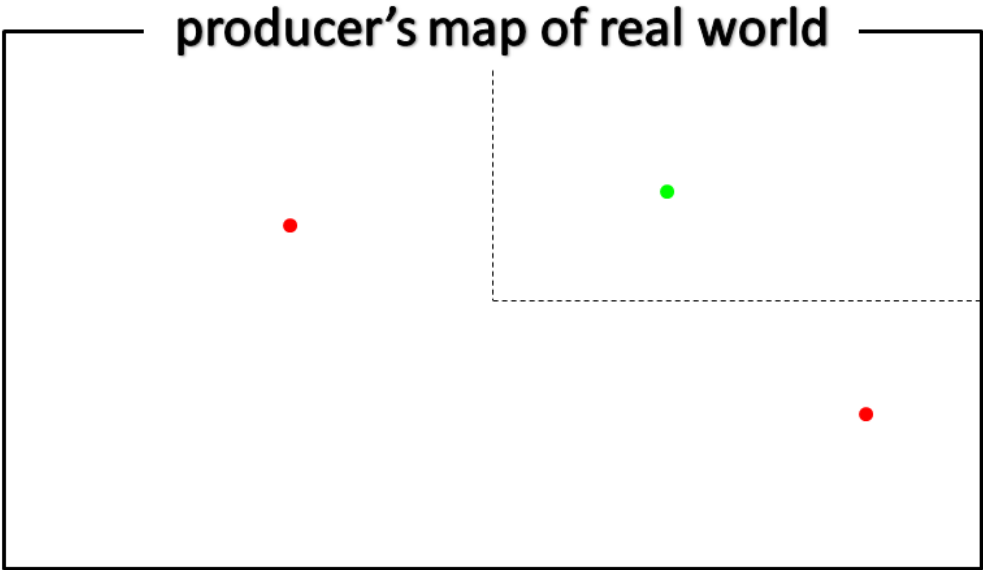
Figure 19. What the producer knows after three cycles.

Now that we have completed three cycles of examples showing how the producer's map gradually gains experience, let us step ahead to the end of 25 such cycles. See Figure 20. In this condition, with 20 dots of experience showing non-wheat and 5 dots of experience showing wheat, the producer may hypothesize that the rectangle shown in dotted lines delineates the part of the world which contains wheat.
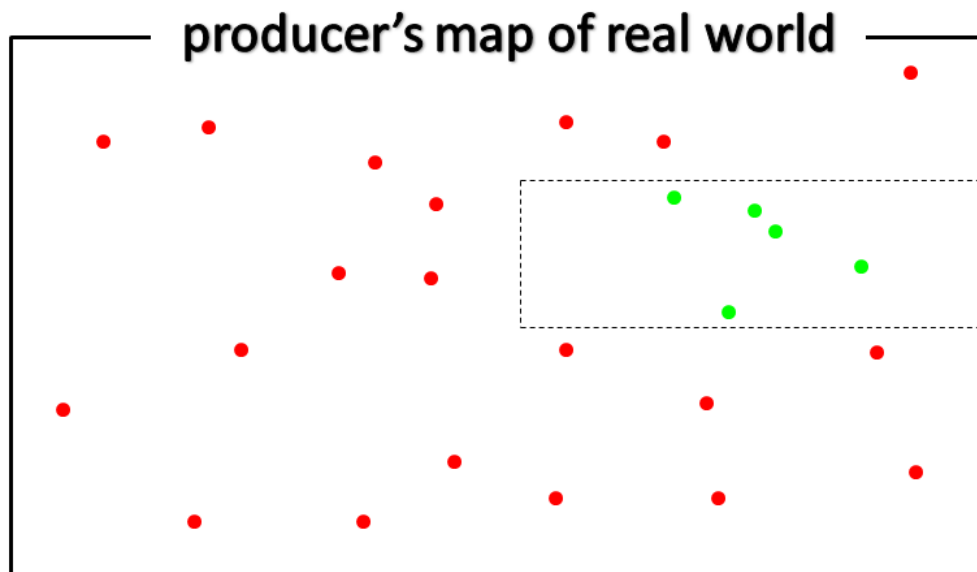
Figure 20. Step ahead to the end of cycle 25 to see both the producer's experience and hypothesis of the area of wheat.

Claims at conclusion of this thought experiment
(1) No matter how much experience the producer gets it never attains perfect knowledge of what is wheat in its world. The producer never hypothesizes an area for wheat which entirely equals the oval with which we modelers defined the wheat area; there will always be cases near the border of the oval upon which the producer might err in its guess about the identity of an object. The producer's knowledge of the meaning of wheat will always be ambiguous.

(2) The experience of the producer in this TE seems analogous to many human experiences of learning the meanings of symbols. Any human agent sent out to acquire an object, whether named with only a single word or specified with a sheaf of documents and pictures, may fail to make a choice which satisfies the principal who sent him.

But of course the chance that the human agent may satisfy his principal increases with the amount of working-together experience shared by this agent-principal pair.

(3) We may say that the producer always has an operational definition of wheat in that it will always act to deliver its best guess. It is never stumped.

(4) With enough experience our producer can get very good at selecting wheat in its world. It might go one million cycles without a single error! But we modelers must be careful about how we describe this accomplishment. We might imply that the producer has more capabilities than it actually has. We humans are biased to quickly and easily perceive things for which we already have names in our natural language. And we are likely to use those names when we talk about those things, talking either to ourselves or others.

For example, we modelers may "see" that our producer has learned a good working definition for wheat. If we say "it has learned a good working definition for wheat" then it may seem a natural step for one of us humans to ask the producer to tell its definition of wheat. But this shows the danger of anthropomorphism. In our outline of this TE, we have given the producer no capacity of language! It certainly does not have any way to receive (to sense) our word "definition". True, it will have built a good set of lines, or rules for selection of objects in its world, but we have given it no capacity to summarize and describe these rules. So we see a stark contrast in these two views of the mental capacity of our producer, between acting as if it knows the definition of wheat and being able to produce a definition in some language of "wheat".

(5) Further this producer has no notion of truth. The primitive operations of this producer select an object which best matches some geometrical constraints on a plane. Nothing in the calculating capacity which we have specified for our producer requires it to form a statement which can then be submitted to a logical truth test. This producer, operating at its primitive level, has not yet come close to needing a fancy concept like truth.

Self awareness, or more specifically an ability in the producer to recognize about itself that it has discovered a rule, may come only much later in the advance of nervous system processing.

(6) In the computational processing of each of our producer's cycles we modelers can recognize computations which we may label as induction and deduction. We can see induction when the producer first hypothesizes which area of its world contains wheat by drawing a line to divide the world, and further induction in the later cycles when the producer draws or moves dividing lines to focus its hypothesis more closely. We can see deduction when the producer employs this current-best hypothesis in choosing one of the three candidate objects offered in a given cycle.

6.3.4.2 Confusion when two try to learn the same "fact"

In the TE just completed in Section 6.3.4.1 we saw how a good working idea of the meaning of a word such as wheat may be gained by an agent we called a producer. Now we will run a similar TE with a different producer which we will call the "other producer". The other producer will differ from the first producer in the ways that it:
- chooses a single object when more than one object remains possible after the range has been narrowed by the current hypothesis for which area contains wheat.
- revises its hypothesis for the wheat area with new experience.

We will see how this different way of thinking leads the other producer to a somewhat different understanding of wheat.

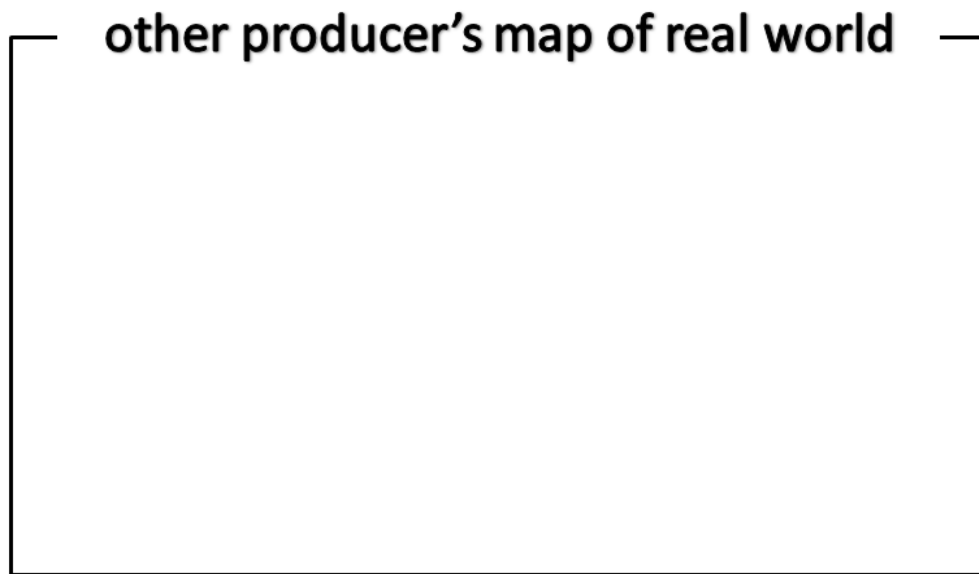**other producer's map of real world**

Figure 21. The other producer's map at the start of cycle 1, showing no experience.

We assume that the other producer starts out with no information in its map, as shown in Figure 21.
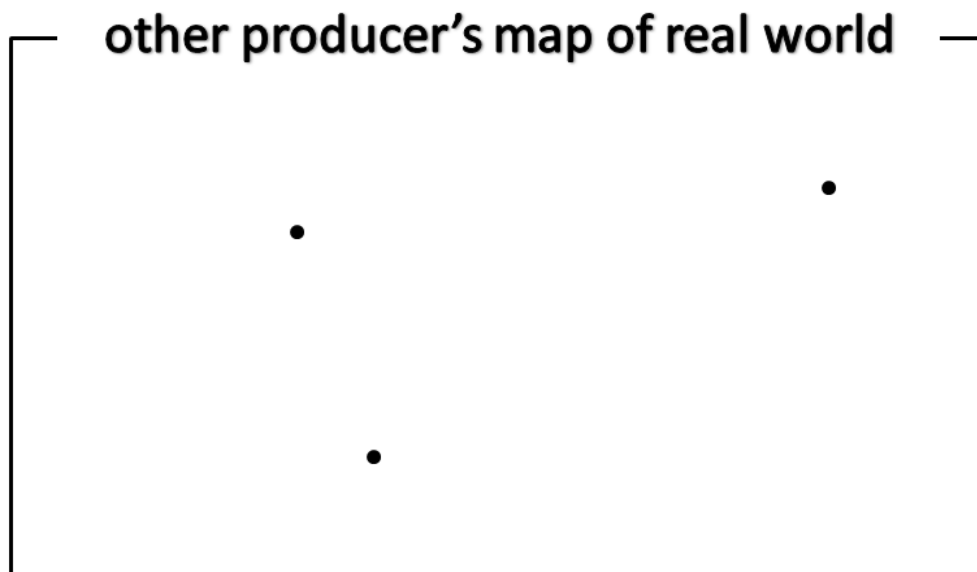
Figure 22. The other producer's three options in cycle 1.

At the start of the first cycle the other producer gets a request for wheat. Then it is given three choices, the same three choices as the first producer got, see Figure 22. But while the first producer guessed that the spot farthest to the left represented wheat this other producer guesses the spot on the right.

This guess turns out to be correct. So the other producer starts cycle 2 with the knowledge shown in Figure 23. The green spot represents a location at which wheat has been found.
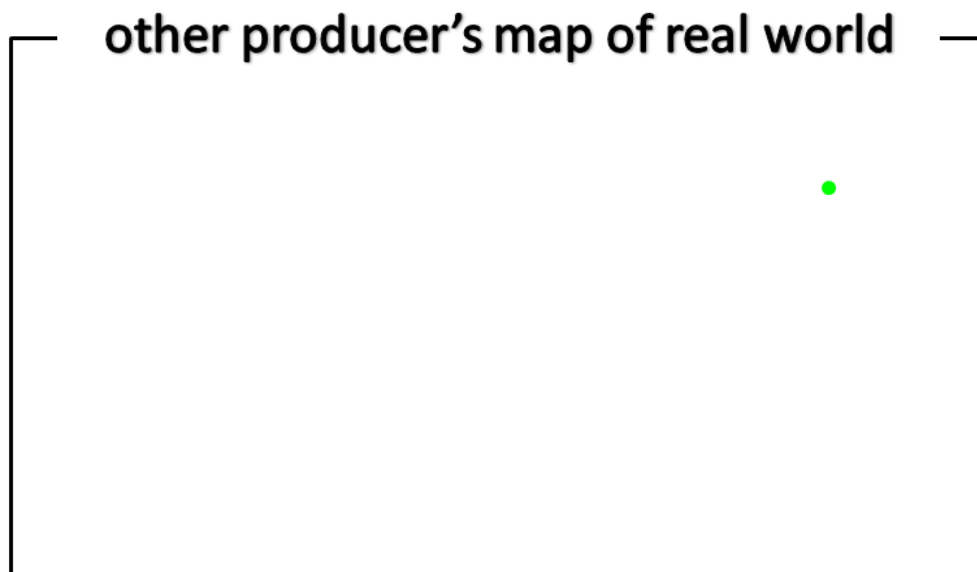
Figure 23. Showing what the other producer learned in cycle 1.

Since this other producer has satisfied a request for wheat in its first cycle we will assume that in the second cycle it gets a request for non-wheat, and it is given the three choices which are shown in Figure 24.
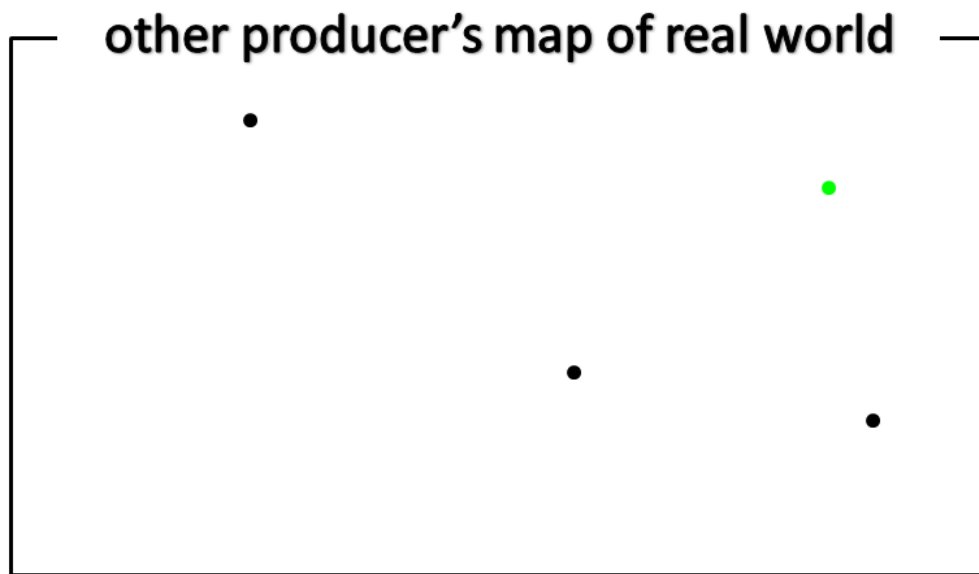
Figure 24. In cycle 2 the other producer gets these three options.

In order to use what it has learned, with its one spot of information, this other producer hypothesizes that the world may be divided into the two classes, wheat and non-wheat, by a horizontal line as shown in Figure 25.
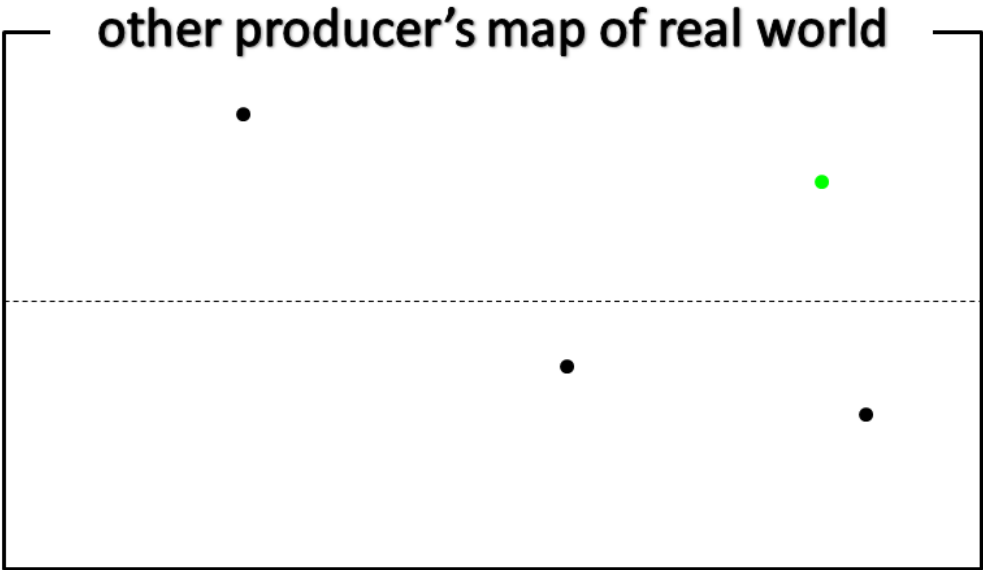
Figure 25. In cycle 2 the other producer hypothesizes that wheat lies in the upper half of its world.

Then, since the hypothesis places non-wheat in the lower half, it guesses one of the two objects in the lower half; it guesses the object at the right. That turns out to be correct. It is non-wheat. So our other producer starts cycle 3 with knowledge depicted in Figure 26.
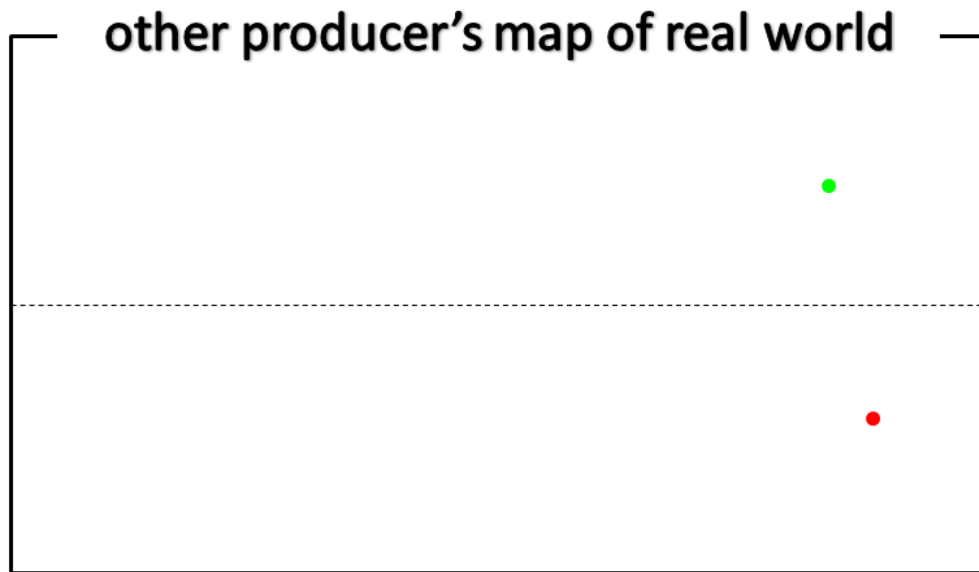
Figure 26. Showing what the other producer knows at the end of cycle 2 (and the start of cycle 3).

In cycle 3 the other producer receives a request for non-wheat, and it is given the three choices shown in Figure 27.
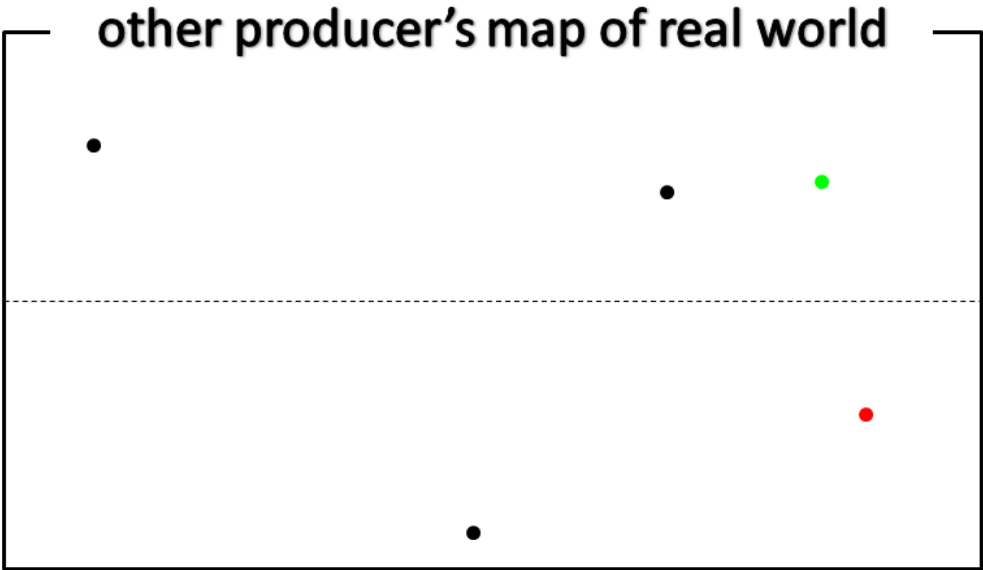
Figure 27. In cycle 3 the other producer is given the three options shown.

The new information learned in cycle 2 fits with the hypothesis formulated from what was learned in cycle 1, so the other producer has no reason to move the horizontal dotted line or to draw another dotted line. It chooses the one option which satisfies its hypothesized rule. This is also non-wheat. So again our producer has succeeded and it enters cycle 4 with the knowledge represented in Figure 28.
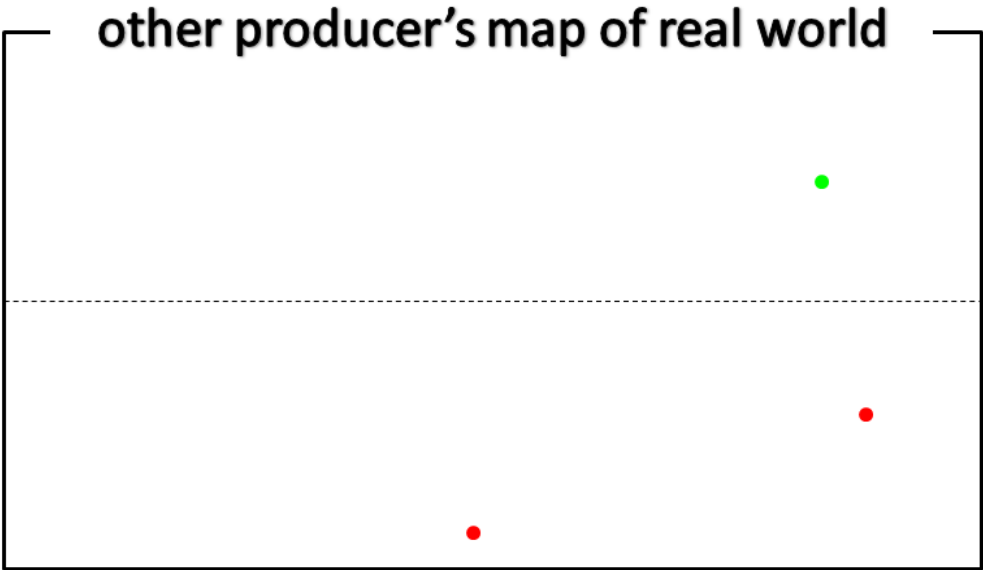
Figure 28. What the other producer has learned after three cycles.

Now, as we did with the first producer, we will jump ahead again to end of the 25th cycle. We will assume that our other producer has repeatedly revised its hypothesis about which area of the world contains wheat. See Figure 29 which shows 25 dots of experience and the dotted lines to show the other producer's current hypothesis concerning which area contains wheat.
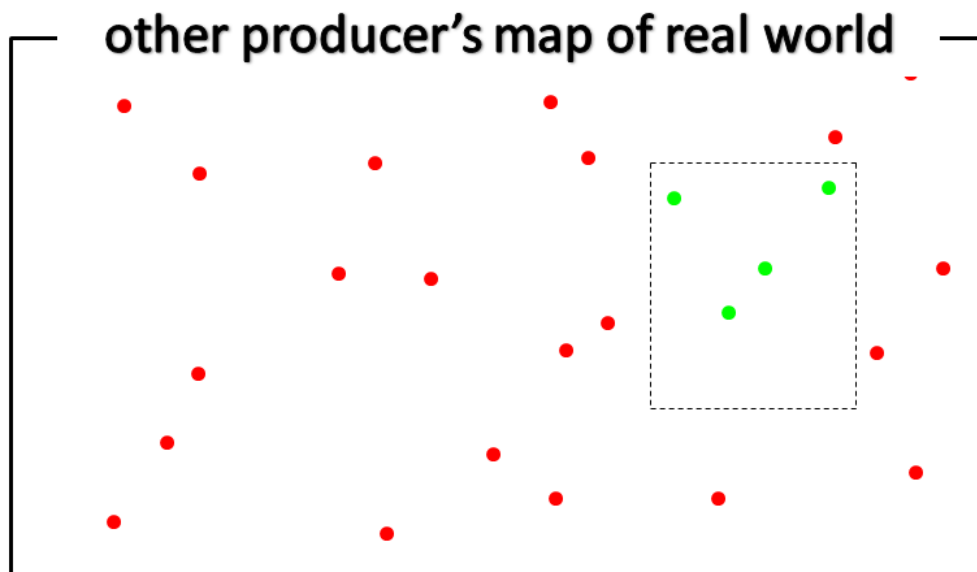
Figure 29. Step ahead to the end of cycle 25 to see the other producer's experience and hypothesis of the area of wheat.

Now we can see that the two producers' ideas (or hypotheses) of wheat differ somewhat. In Figure 30 we see the two areas hypothesized by the two producers, copied from Figures 20 and 29, along with the true area which we modelers established by definition, copied from Figure 10.
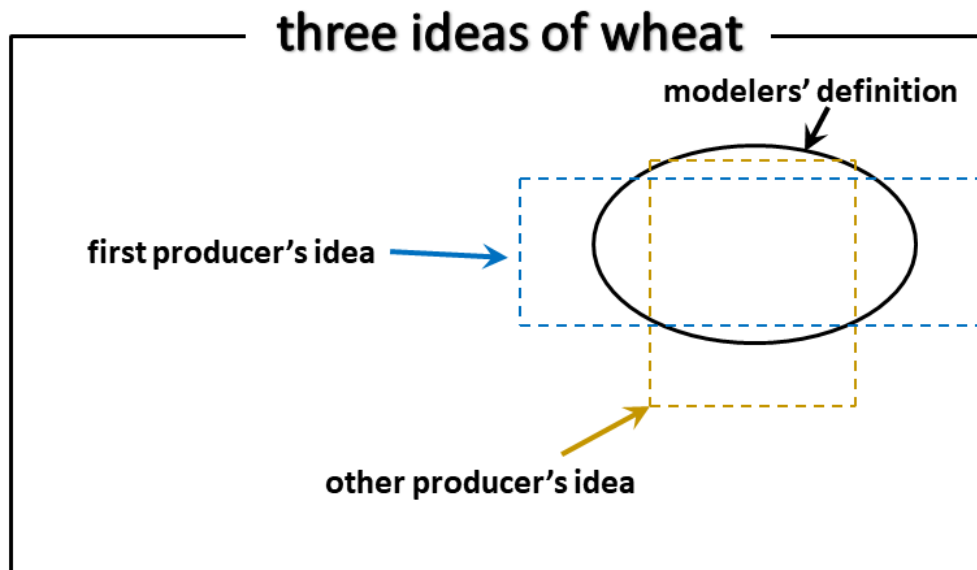
Figure 30. Our two producers have learned much but have different ideas of wheat and still might err when selecting wheat.

Claims at conclusion of this thought experiment
(1) To the extent that this LLE models how we learn words in natural language, it is natural and inevitable that different people will have different ideas of the meanings of words.

(2) The difference between these two producers' hypotheses of the area for wheat has come about because the two producers have different ways of thinking. Even in identical circumstances sometimes they make different choices. As a result they accumulate different experiences with which to make future choices.

(3) Yet the difference between the two producers will narrow as they each accumulate experience because, as we constructed these TEs, judgement concerning whether a chosen object is wheat or non-wheat comes from a standard (the oval) created by us modelers.

(4) We, as users of natural language, should understand that other users of natural language always have their own individual needs, abilities, and experiences, so unless we have completed the perhaps impossibly difficult task of learning of those others' minds, we should not believe that we perfectly understand the natural language expressions of others. Rather, we should understand that our grasp of the meanings of others' natural language expressions is limited by

the scope of our own needs, abilities, and experiences. We may be able to act as they would wish in response to their statements, but we have to be careful in assuming we understand why they said what they said.

6.3.5 Concluding Claims about Language Learning

These LLEs get their significance from RPM's approach to realism, to a realistic ontology
Some readers may judge the above language-learning experiments to be simple and insignificant, so I want to emphasize how the meanings of these experiments are amplified by the context of RPM. If you accept RPM's postulates outlined in Sections 1.4 and 6.1 then you can see how agents in our world find themselves, like the agents in our LLEs, in circumstances in which they gain a clear advantage if they can learn to successfully signal one another. RPM lays down a circumstance in which we should expect LTs given appropriate powers to develop mutually helpful signaling.

Good enough understanding
One result which follows from RPM's postulates is that signals between agents do not need to be perfect. All the agents need is that their attempts at communication succeed frequently enough so that the overall benefits of those attempts exceed the costs. Words can and probably do have fuzzy meanings because that is what works in the practical world of RPM. Figure 31 summarizes this argument, showing that our understanding may be good enough for various particular

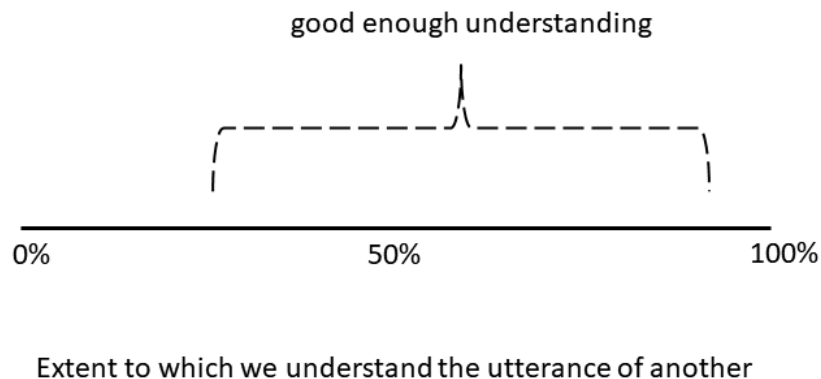circumstances if we understand each other only partially.

good enough understanding

Figure 31. We manage to survive with good enough understanding.

6.3.6 Future Directions for Language-Learning Experiments (LLEs)

Rewards should come from exchange with other agents, as well as from the environment.
The LLEs presented in this chapter have included us modelers in that we:
(1) set the standard for what learning was to be accomplished by the agents in the experiments, and
(2) provided feedback to the agents concerning the how we judged their behavior.

While this involvement has enabled us to create a few simple but instructive LLEs, future LLEs will do well to remove us modelers farther from the action in the model.

We could create LLEs in which agents are:
(1) motivated by their own hungers or ambitions (hungers or ambitions which we modelers have built into the agents, but which we modelers do not control directly because these hungers or ambitions arise in response to ongoing circumstances in the model), and

(2) rewarded by what they gain in exchange with other agents (as they learn to coordinate their actions with other agents in order to harvest from resources which we modelers have placed in their world).

Let the agents learn which classes of objects deserve distinctive names.
Furthermore we modelers could learn by removing ourselves from specifying the objects for which the agents need to learn names. In the experiment of Section 6.3.1, you may have noticed, we modelers set up the experiment so that the agents would learn one symbol each for each of the five types of objects which we made available. We did not know what symbols would become associated with the objects, but we knew there would be five distinct symbols for the five objects which we built into the experiment. We could step closer to realism by building an experiment with these components:

- agents with hungers (metabolic requirements) which may be satisfied in varying degrees by different resources found in the world
- resources which are distributed in the world in concentrations which sometimes vary gradually from location to location
- agents capable of communicating with a substantial variety of symbols.
- agents restricted in their individual capabilities so that cooperative ventures are likely to benefit at least some of the cooperators.

With such components we may expect to see signaling develop but we may be incapable of guessing the size of the vocabulary which ensues.

For examples of these issues, consider these TEs with generalized need for food.

> (Example TE 1) As we LTs interact with the world, there will be many objects in the world which we cluster together in our thoughts because those clusterings have some common feature as seen from the viewpoint of the LT.

> When for example I feel hungry I think of food. Many varieties of food may serve this need of mine. I search, not for the one specific double cheeseburger which I may purchase but rather for any of the many particular meals which I may subsequently find before me. I could not perform such a generalized search without having some sort of conceptual category which encompasses all the possibilities for the meals which I might eventually consume.

> (Example TE 2) Imagine agents with some similarity to the consumer and producer in Section 6.3.1. Suppose the consumer can feel hunger for grain and this hunger might be satisfied by either wheat or oats, so that a different symbol may be discovered to convey this more-general wish, and both producer and consumer rewarded by the environment when they accomplish productive signaling for this circumstance.

> Extending this thought experiment, it may happen that the producer gains access to rice and that the consumer would find rice to satisfy its hunger for grain, even though the consumer does not even know about the existence of rice.

Although the producer does not know under what circumstance if any he will be rewarded for supplying rice to the consumer, as time passes a circumstance may arise in which the producer's best available choice is to try passing rice to the consumer — and see what happens.

If the consumer favors the rice it might want more and thus repeat the signal which had previously gained rice from the producer, but the producer, finding a different set of objects available in its world in the present cycle, may choose to deliver an object which previously had been favorably associated with the consumer's symbol. So we have a situation in which the two agents would do best to learn a new signal for rice, but also to modify the meaning of grain to include rice.

6.3.7 Conclusion: Ambiguity in Language is Unavoidable but Often Helpful

These TEs exemplify what I claim is a general truth. We LTs, in roles as consumers, routinely experience needs for which we are ignorant of all the specific ways in which a given need may be satisfied. And, when we are in roles as producers, we routinely have options for actions which, given particular inputs, may bring us rewards; but we don't know till we have more experience or some sort of informative guidance. Not only the experimentally-simple producer and consumer, but also we fancy human LTs, live in circumstances which require us to signal with terms which have, for our initially and necessarily ignorant positions, many possible meanings – general meanings.