UNIT 7: CONSTRUCTING THEORIES IN BIOLOGY

7.1 Beyond Geometry: Theory in Biology

In Units 1–6, we outlined some of the methodological strategies for constructing and evaluating theories, where we used geometry as our primary terrain to exemplify and practice theory building. In this unit, we will move to a different terrain for exemplification and practice — the terrain of biology, but using precisely the same methodological strategies. Needless to say, the migration from mathematics to science requires an additional norm: *the logical consequences of a scientific theory must not only correctly predict the observational generalisations relevant to the theory, but also explain the asymmetries in those generalisations.* (§6.4)

We hope that this migration from geometry to biology would shed light on those transdisciplinary modes of thinking and reasoning that are ***transferrable*** from one domain of knowledge to another. With that purpose in mind, this unit will focus on two different theories in biology:

a theory of *anatomy; and*
a theory of *habitat.*

These theories, combined with theories of biological *function*, biological *change,* and biological *development* are essential to the construction of a theory of biological *evolution*. But the challenges of constructing each of these theories is beyond the scope of this monograph.

7.2 Observing, Reasoning, and Two Formalisms for Reasoning

Suppose you look out of a window and see what is given in this photograph.

Think of the border of the photograph as the frame of the window. You see only a part of the entity, not the whole entity. Will you be able to answer the following multiple choice questions?

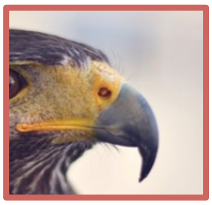

Figure 7-1

1. Is the entity whose part you are looking at:
   a. inanimate, animate, or neither?
   b. a man, a woman, a child, or none of these?
   c. an animal, a plant, or neither?
   d. a frog, an insect, a tree, a snake, a bird, a bat, or none of these?
   e. a hawk, an eagle, a parrot, a crow, a humming bird, a book, or none of these?

Write down your answers to these questions. And now answer the following 'YES-NO' questions. It would be a good idea to write down your answers.

2. Does this entity have the following? Put a Y(es) or N(o) in the box to the right of each.

a.	legs		e.	blood		i.	a tail		m.	fur	
b.	lungs		f.	bones		j.	a beak		n.	fins	
c.	hair		g.	cells		k.	nostrils		o.	fingers	
d.	heart		h.	claws		l.	eyes		p.	feathers	

3. Write down your answers to these questions as well:
  a. If your answer to question (1a) is yes, how many?
  b. If your answer to (1q) is yes, how many?
  c. What else can you say about this entity?

If you now compare your answers with the answers from a random collection of high schoolers, you will find something surprising: almost all of themtheir answers would be the same as yours. And interestingly, all this just because you looked at the picture in *Fig. 7-1*, and recognised it as a beak (question (2j)).

Now, only two, or perhaps three of your answers (the answer to (2j, k, l) came from your ***observation***. The rest came from your ***reasoning***.

To arrive at an answer through reasoning, you used a set of statements of the form:	If an entity has X, then it ……
Logicians refer to statements of this form as ***conditionals***:	If X, then Y.
To express conditionals with a symbol in formal logic, logicians use an arrow: It expresses the relation of ***implication*** (X implies Y).	$X \rightarrow Y$
In the physical sciences, the corresponding relation is that of ***equality***, expressed by the symbol '='	$X = Y$
The equality symbol is equivalent to a double headed arrow (if and only if) in logic.	$X \leftrightarrow Y$

In the formulation of laws in the sciences of inanimate entities (i.e., the physical sciences), the most common relation is that of equality. For the formulation of the laws in the sciences of animate entities (i.e., biological sciences), we will rely instead on the relation of implication.

7.3 A Theory of Parrots

Notice that the questions in (1) are about the ***categorization*** of an entity, one part of which you can observe. In contrast, the questions in (2) and (3) are about the ***correlations*** between biological traits. Based on these correlations, you can infer information about one part of the organism from information about another:

4. a. If the entity has a beak, then ……
  b. If an entity is a parrot, then ……

Chances are that your answer to question (1e) is that the picture in *Fig. 7-1* is that of a parrot. Needless to say, having a beak is not sufficient for you to infer that the entity is a parrot. To infer this, you also need some other features that you observed in the photograph.

The set of ***if-then conditionals*** on the basis of which you gave answers to (2) and (3) constitutes a ***theory*** of the anatomical properties of the taxon we call *parrot*. But notice that it also is the same as a theory of the anatomical properties of the taxon we call *bird*.

7.4 Theories of Birds and Bees

In Units 1-6, we illustrated one of the methodological strategies of theory construction: we construct a theory of X by taking a ***description*** of X and converting it into a Premise-Derivation-Conclusion structure; treating some of the description propositions as ***premises,*** others as ***conclusions***; and deriving the conclusions from the premises through ***deductive reasoning***. In §7.3, we extended that strategy to construct a rudimentary theory of the anatomy of parrots.

This theory begins with the premises in (4), now filled in (in (5)), and proceeds to a number of other premises of the form: If X, then Y:

5. a. If the entity has a beak, it is a bird.
b. If an entity is a parrot, it has feathers.
c. If an entity is a parrot, it has two legs.
d. If an entity is a parrot, it has two wings.
and so on.)

[Note that the statement, "If the entity has a beak, it is a parrot." is false.]

This is not a particularly interesting theory because it postulates a separate premise for each property of the bird. Remember we saw some conditions on axiomatic systems in §6.4? One of the conditions was the logical connectedness (of premises), and other conditions were that we minimise the number of premises (simplicity), and maximise the number of conclusions (generality).

And remember the idea of the logical inheritance of properties?

The properties of a category are inherited by their subcategories. (§1.4; §3.2))

Given the general principle that *subcategories inherit the properties of their mother categories*, one way of minimising the number of premises is to state the conditionals on higher level categories. Thus, instead of (5c), we may postulate (6):

6. a. If an entity is a parrot, then it is a bird. (= 'Parrot' is a subcategory of 'bird'.)
b. If an entity is a bird, then it has two legs. (= Birds have two legs.)

We can now deduce all the properties of parrots we have described above from parrots being a subcategory of birds. This allows for a powerful reduction of premises, because from a single set of axioms on birds, we can now predict the anatomical properties of not only parrots, but also hawks, eagles, vultures, crows, ravens, sunbirds, and so on,when combined with the subcategory statements in (7):

7. a. Parrots are a subcategory of birds.
b. Hawks are a subcategory of birds.
c. Eagles are a subcategory of birds.
d. Vultures are a subcategory of birds.
e. Crows are a subcategory of birds.
f. Ravens are a subcategories of birds.
g. Sunbirds are a subcategory of birds.

Given these subcategory statements, there is no need to duplicate the statement of properties for each of the subcategories separately.

Now, if you compare the properties of birds with the properties of snakes, reptiles, fish, frogs, and mammals, you will find that they have a number of properties in common. Such shared properties across taxa are called ***homologies*** in biology. For instance:

8.	a.	Snakes have vertebrae.	Fish have vertebrae.	Mammals have vertebrae.
	b.	Snakes have blood.	Fish have blood.	Mammals have blood.
		Snakes have mouth.	Fish have mouth.	Mammals have mouth.
		Snakes have eyes.	Fish have eyes.	Mammals have eyes.
		And so on.	And so on.	And so on.

We can achieve even greater economy of premises with the following statements:

9. Vertebrate (DEF): The category of organisms with vertebrae.

10. a. Birds are a sub-category of vertebrates.
b. Snakes are a sub-category of vertebrates.
c. Fish are a sub-category of vertebrates.
d. Reptiles are a sub-category of vertebrates.
e. Frogs are a subcategory of vertebrates.
f. Mammals are a subcategory of vertebrates. And so on.

Needless to say, the different subcategories have not only ***similarities,*** but also ***differences***. That is what the concept of homology implies: similarities among differences, unity in diversity, invariance under variability, and the like. We account for both similarities and differences by postulating the similarities on the mother category, and the differences on the daughter category. That leads to the tree structure of biological categorisation:

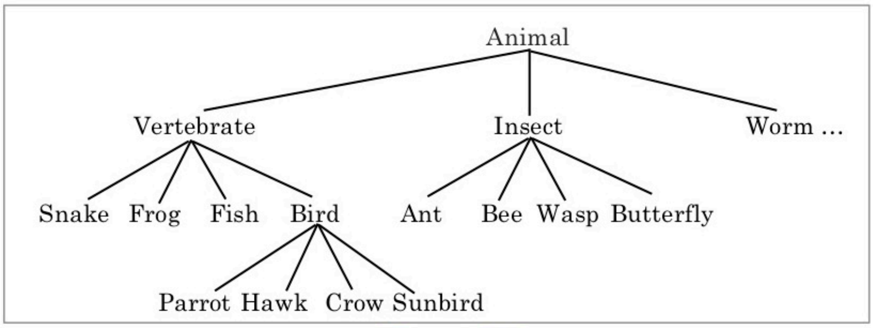

Figure 7-2

To take an example, the wikipedia entry on butterflies gives us a list of the properties of butterflies. (https://en.wikipedia.org/wiki/Butterfly) .

Adult butterflies
- have large, often brightly coloured wings
- have conspicuous, fluttering flight
- have a four-stage life cycle
- undergo complete metamorphosis
- lay eggs on the food plant, on which they grow into caterpillars, which
  - feed on the plant and
  - metamorphose into butterflies

This is just a small list. The entry does not mention that butterflies

- have compound eyes
- have six legs,
- have a neural system
- have a food canal
- have eukaryotic cells, and
- are multicellular

The reason why such properties are not mentioned in the entry is because these are predictable from the classificatory tree of categories and subcategories — a theory of the category of butterflies as a sub-theory of the category of animals, which in turn is a sub-theory of the category of animate entities.

Exercise 1:

TASK 1: Read the Wikipedia entries on snakes, frogs, and fish to make a list of the anatomical properties of each. Derive as many of them as possible by setting up the shared properties as properties of vertebrates. Those that are not shared, state as properties of the daughter categories.

TASK 2: Read the Wikipedia entries on bees, butterflies, insects, worms and animals to make a list of their anatomical properties. State properties of animals to deduce the shared properties of the daughters from the properties of the mother.

If you now go back to Unit 6, you will see that the methodological strategies we have used in §7.2 - §7.4 are those of description, abstraction, categorisation, generalisation, reasoning, and prediction. It is some of these strategies that you pratised in *Ex. 1*.

In the tree diagram in *Fig. 7-2*, we have used yet another strategy, that of representations. Having seen such visual representations in Venn diagrams (rectangles and circles), and in geometry (diagrams of straight and curved lines, triangles, rectangles, circles, …), it should be easy to connect the different kinds of representations in academic knowledge, and reflect on the IDEAS that the representations express (or are intended to express).

Before proceeding, it may be a good idea to reflect on this question: Do circles and rectangles in Venn Diagrams represent the same concepts as circles and rectangles in geometry? Your answer is likely to be no. If so, it would be a good idea to reflect on what exactly the conceptual differences between them are, and what the conceptual similarities are, and write down your answers.

Exercise 2:

TASK: Read the Wikipedia entries on unicellular life forms, prokaryotes, and eukaryotes, and construct a theory of the anatomical properties of life forms. To do this, you will need to incorporate the categories of animate entities, unicellulars, eykaryotes, and prokaryotes into the tree in *Fig. 7-2*.

7.5 A Theory of the Anatomy of Animate Entities

The theory that we have constructed so far is of the ***anatomical properties of animate entities***. It says very little about anatomical structure as such.

For instance, for a ***theory of anatomical structure,*** it is not sufficient to say that vertebrates have legs. We also need to say something about the structure of legs. What does this mean? Well, a human leg is composed of the upper leg, the lower leg and the foot. A human arm is composed of the upper arm, the lower arm, and the hand. We can unify legs and arms by stating these structural properties as the properties of limbs. We also need to say something about hands being composed of the palm and fingers, and something similar about feet. Finally, we need to say something about the three part structure of the digits on the hand, and also, the number of digits.

Constructing a theory of the anatomical structure of animate entities as a project is likely to take a few months. And if we want to make it comprehensive, perhaps two or three years. If one is adventurous, this would be a great challenge for a Master's or PhD thesis.

7.6 A Theory of Fitness: Survival and Extinction

Central to Darwin's theory of biological evolution is the idea of ***Natural Selection***, which he treats as being synonymous with the idea of ***Survival of the Fittest***. The term 'Survival of the Fittest' was coined by Herbert Spencer, famous for his idea of social Darwinism that says that superior physical force has a hand in shaping human history. Darwin adopted the phrase as an alternative to 'natural selection' in the fifth edition of the *Origin of Species*, published in 1869, intending it to mean "better designed for an immediate, local environment." He says:

> " ...can we doubt (remembering that many more individuals are born than can possibly survive) that individuals having any advantage, however slight, over others, would have the best chance of surviving and of procreating their kind? On the other hand, we may feel sure that any variation in the least degree injurious would be rigidly destroyed. This preservation of favourable variations, and the destruction of injurious variations, I call Natural Selection, or the Survival of the Fittest. Variations neither useful nor injurious would not be affected by natural selection, and would be left either a fluctuating element, as perhaps we see in certain polymorphic species, or would ultimately become fixed, owing to the nature of the organism and the nature of the conditions." (p.93)

To understand what Darwin says in this excerpt, let us formulate two related laws/axioms of Natural Selection, which we will call ***Survival of the Fittest*** and the ***Extinction of the Unfit*** respectively. In these statements, 'species' includes 'varieties.'

11. ***Fitness***: There are various degrees of fitness.

- **A)** ***Survival of the Fittest***: Only those species that are *fitter than all other species* are selected for survival. All other species become extinct.
- **B)** ***Extinction of the Unfit***: There is a *threshold* below which a species is unfit. Species that are unfit become extinct. All other species survive, regardless of their *relative fitness.*

Whether 'fitness' is elaborated as in (A) or (B), it is clear that we need a theory of fitness to critically evaluate and compare the predictions of (A) and (B).

Given these concepts of fitness, survival and extinction, what a theory of fitness needs to explain and predict are the ***asymmetries*** of survival and extinction. To illustrate, take yeast. The first yeast emerged on earth hundreds of millions of years ago. Various taxa of yeast still survive on earth, in contrast to many taxa which have become extinct. A theory of fitness ought to explain this asymmetry between survival and extinction.

If we expand the scope of the theory of survival and extinction from macro biology to micro biology, the theory needs to explain and predict the asymmetry between the survival, for instance, of the gene called cdc2 and the extinction of those genes that no longer exist today. This asymmetry between survival and extinction is relevant not only to molecules and taxa, but also cells, tissues, and organs.

In what follows, our focus will be on a theory of Extinction, as part of the theory of Natural Selection.

7.7 A Theory of Viability Selection and Habitat

The research literature on fitness talks about two kinds of selection, namely, ***fecundity selection*** and ***viability selection***. (e.g., "Role transformation of fecundity and viability: The leading cause of fitness costs associated with beta-cypermethrin resistance in Musca domestica" at (https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0228268))

Fecundity is a matter of the number of offspring that an organism, variety, or species produces in the reproductive process. Viability, on the other hand, is a matter of continued existence from the conception till the organism/variety/species reach the reproductive stage. If a unicellular zygote dies before it develops into an embryo, if the embryo dies before birth, or if the infant dies before it becomes capable of reproduction, the issue of fecundity selection becomes irrelevant.

Given this situation, we will choose viability selection as the crucial theoretical concept on which we will construct our theory of fitness. We will also choose extinction (Idea B) as our criterion for the testing of predictions. Given that during the last hundred years, a large number of species have become extinct, it should be easy to test the predictions of the theory in terms of the data available on species that have become extinct during the recent years, and continue to become extinct. We do not need to count the number of offspring an organism or a variety produces.

With those remarks, let us define fitness in terms of the concepts of adaptation and habitat as follows:

12. ***Fitness*** (DEF): For a species/variety to be fit, its anatomy, physiology, and behaviour must be adapted to (aligned with) its habitat; a species/variety is unfit iff it is ill-adapted to (misaligned with) its habitat.

13. ***Law of Extinction***: A species/variety that is unfit becomes extinct.

The next step is to specify the parameters relevant to the fit between the anatomy, physiology and behaviour of a variety/species on the one hand, and the habitat.

For this, let us define habitat as follows:

14. ***Habitat*** (DEF): The habitat of an organism/variety/species X is the local three dimensional space in which X exists.

Let us take a few examples.

15. Imagine the habitats below, each of them being a space enclosed in glass:

	filled with	SURVIVAL / EXTINCTION		
		Fish	Mice	Earthworm
A	water, but no air or soil	survive	die	die
B	air. but no water and soil	die	survive	die
C	moist soil but no water or air	die	die	survive
D	moist soil below and air above the surface	die	die	survive

In order to provide theoretical explanations for such *experimentally testable correlations* in organisms(/varieties/species), *between*

a) an organism and the space surrounding it, on the one hand, and
b) its extinction on the other,

we need to understand

i) its anatomy, physiology and behavior, and
ii) the properties of the space surrounding it, and what exists in that space.

For instance, what are the anatomical and physiological properties that make fish survive in habitat A but not in B-D? What are the cellular and molecular bases of these macro level anatomical and physiological traits? These are some of the types of questions that would need to be answered for a coherent theory of fitness that makes testable predictions in why certain varieties/species become extinct in some habitats, and why some become extinct globally.

The parameters of water, air, surface of the earth, and under the earth as broad categories of habitat are not sufficient. For each of these, we also need to specify:

16. a. Temperature: The upper and lower boundaries for the range of temperature in which the species/variety can survive. (Bacteria can survive in temperature ranges in which humans and mice would not survive.)

b. Composition: The molecular composition of the substances. (Oxygen, Nitrogen, Carbon Dioxide in the air, water and soil)

c. Nutrients: Availability of nutrients (Plants for cows; animals for lions…)

Exercise 3:

TASK: Read the Wikipedia entries on the habitats of as many taxa as you can deal with, and construct a theory of fitness in terms of entity-habitat pairing that specifies the threshold below which the entity becomes extinct.

This task may take three hours, three months, or three years, depending on whether it is a learning task for an undergraduate course, a higher level research project , or a doctoral thesis.

If you are adventurous enough to pursue the task in *Ex. 3* as a topic for research, here is a hint. In order to construct a rigorous theory in a rigorous fashion, you would need appropriate logical and mathematical formalisms. We have already mentioned the formalism of formal logics for scientific laws formulated as if-then conditions in systems of formal logic, but we also need the representational system of Graph Theory. The one used in *Fig. 7-2* is that of a tree diagram, which is a visual representation in a Directed Acyclic Graph (DAG) in Graph Theory in mathematics. We are going to need these two in the construction of the theory of fitness as well.

In addition, a PhD student would do well to explore representational systems that computer scientists call Attribute-Value model.
(see: https://en.wikipedia.org/wiki/Entity-attribute-value_model)

To think about: At the end of §7.4, we said that §7.2– §7.4 have crucially used the methodological strategies of description, abstraction, generalisation, reasoning, prediction, and representation. Did we miss something?

Having gone through §7.5 and §7.6 as well, what do you think is the methodological strategy that we have used in these sections?

7.8 Levels and Dimensions of the Tree of Life

The Tree of Life in *Fig. 7-2* is that of taxa of animate entities based on anatomical properties. The discussion of the habitats of taxa lead to the possibility of a different system of categorization based on the similarities and differences in their habitats.

Take what the Wikipedia entry on water birds says:

> "A water bird, alternatively waterbird or aquatic bird, is a bird that lives on or around water. In some definitions, the term water bird is especially applied to birds in freshwater ecosystems, although others make no distinction from seabirds that inhabit marine environments. Some water birds (e.g. wading birds) are more terrestrial while others (e.g. waterfowls) are more aquatic, and their adaptations will vary depending on their environment. These adaptations include webbed feet, beaks, and legs adapted to feed in the water, and the ability to dive from the surface or the air to catch prey in water. "

Let us generalise the term 'water bird' to include to 'water life forms' to include all animate entities. If we make this move, the discussion in the previous section points to the following categorization:

17. In the *three dimensional space surrounding them* in their habitat, nimate entities require (a) the presence, and (b) the absence, of:
i. water   ii. air   iii. soil

Suppose we use the symbol '+' for 'require the presence of' and '–' for 'require the absence of'. Using this notation, we may categorise animate entities as:

18. a. Water: [+W] or [–W]
b. Air: [+A] or [–A]
c. Soil: [+S] or [–S]

What we have in (18) is an attribute-value system. (see https://en.wikipedia.org/wiki/Entity-attribute-value_model). In this system, 'water', 'air', and 'soil' are 'attributes': each of them a dimension, an axis, or a parameter of categorisation, with '+' and '–' as the values of an entity along a given parameter.

Given this system, we may specify the values of the aspects of their habitat for fish, bird, mouse, earthworm, and frog as follows:

19. a. Fish $\begin{bmatrix}+W\\-A\\-S\end{bmatrix}$ b. Bird $\begin{bmatrix}-W\\+A\\-S\end{bmatrix}$ c. Mouse $\begin{bmatrix}-W\\+A\\-S\end{bmatrix}$ d. Frog $\begin{bmatrix}+W\\-A\\-S\end{bmatrix}$ **OR** $\begin{bmatrix}-W\\+A\\-S\end{bmatrix}$ e. Earthworm $\begin{bmatrix}-W\\+A\\-S\end{bmatrix}$ OR $\begin{bmatrix}-W\\-A\\+S\end{bmatrix}$

A word about the terms 'water', 'air', and 'soil', and what the notation means in this attribute-value system:

20.	Must be surrounded by	Must not be surrounded by
a.	[+W]: "water"	[–W]: "water"
b.	[+A]: "air"	[–A]: "air"
c.	[+S]: "soil"	[–S]: "soil"

Now, the term 'water' may refer either to:

(i) water molecules (H_20); or

(ii) the substance we call water, with three physical states: gas, liquid, solid; or

(iii) water in the liquid state.

In our notation here, it means 'water as a substance in the liquid state'. This is because, for instance:

21. Fish cannot survive if there is air around them, even if the air has water molecules in it. Nor can they survive if the water is in the solid or gas state.

Aeroponic plants cannot survive without water as an aggregate of H_2O molecules, but they can survive without water as a substance in the liquid form (https://en.wikipedia.org/wiki/Aeroponics)

Do we need to specify every aspect of the habitat for every species/variety? Take water molecules, and the following universal law on animate entities:

Law 1: All life forms, in order to survive, need water molecules as part of their composition.

Let us use the symbol [WM] to denote the dimension of the water molecule. We may now specify the property that comes from Law 1 as:

22. a. Fish b. Bird c. Mouse d. Frog e. Earthworm

$$\text{a.}\begin{pmatrix}+\text{WM}\\+\text{W}\\-\text{A}\\-\text{S}\end{pmatrix}\quad \text{b.}\begin{pmatrix}+\text{WM}\\-\text{W}\\+\text{A}\\-\text{S}\end{pmatrix}\quad \text{c.}\begin{pmatrix}+\text{WM}\\-\text{W}\\+\text{A}\\-\text{S}\end{pmatrix}\quad \text{d.}\begin{pmatrix}+\text{WM}\\+\text{W}\\-\text{A}\\-\text{S}\end{pmatrix}\textbf{OR}\begin{pmatrix}+\text{WM}\\-\text{W}\\+\text{A}\\-\text{S}\end{pmatrix}\quad \text{e.}\begin{pmatrix}+\text{WM}\\-\text{W}\\+\text{A}\\-\text{S}\end{pmatrix}\textbf{OR}\begin{pmatrix}+\text{WM}\\-\text{W}\\-\text{A}\\+\text{S}\end{pmatrix}$$

However, all the organisms classified in (22) are subcategories of animate entities. So, if we assign the property to animates, the mother category, as follows:

23. Animates: [+WM]

there is no need to specify it on each of its subcategories. Given this specification on the category 'animate', its daughters and other subcategories will inherit this assignment of attribute value.

What all this points to is a need to construct a classificatory system for animate entities involving attributes and values of the kind illustrated above. That system would complement the system illustrated in *Fig. 7-2*.

Exercise 4:

TASK: Read the Wikipedia entries on the habitats of as many taxa as you can deal with, and construct a classification of animals along the lines indicated above, including the specification of temperature, nutrients, and so on.

As a learning task or a research project, this one is much like the one in *Ex. 3*.

7.9 Summing up

In a recent article in the journal *Nature*, Nobel Laureate Paul Nurse points out that biology needs to go beyond mere data and descriptions to ***ideas*** that can explain and predict what we find in the data and descriptions.

Paul Nurse, "Biology must Generate Ideas as well as Data."
Downloadable at: https://www.nature.com/articles/d41586-021-02480-z

What we have done in this unit may be viewed as a way of following Nurse's recommendation, by demonstrating what it takes to generate ideas and develop them as theories whose predictions are testable.

With that remark, we will deliberately leave the rest of the "Summing-Up" to you. What we would like you as a reader to do is to go through all the units, and write a summary of the monograph. When you do that, make sure to specify what you learnt from it that is of value to you, and will be of value to you in your future work.

Finally, we have a request. Please send us that summary at tara.mohanan@gmail.com. We would like to find out whether or not our efforts have benefited you, and how we can revise the monograph to benefit more learners, and provide higher value.