Part III

The Embodiment of Infinity

8

The Basic Metaphor of Infinity

Infinity Embodied

One might think that if any concept cannot be embodied, it is the concept of infinity. After all, our bodies are finite, our experiences are finite, and everything we encounter with our bodies is finite. Where, then, does the concept of infinity come from?

A first guess might be that it comes from the notion of what is finite—what has an end or a bound—and the notion of negation: something that is *not* finite. But this does not give us any of the richness of our conceptions of infinity. Most important, it does not characterize infinite *things*: infinite sets, infinite unions, points at infinity, transfinite numbers. To do this, we need not just a negative notion ("not finite") but a positive notion—a notion of infinity as an entity-in-itself.

What is required is a *general* mathematical idea analysis of all the various concepts of infinity in mathematics. Such an analysis must answer certain questions: What do the various forms of infinity have to do with each other? That is, What is the relationship among ideas like infinite sets, points at infinity, and mathematical induction, where a statement is true for an infinity of cases. How are these ideas related to the idea of a limit of an infinite sequence? Or an infinite sum? Or an infinite intersection? And finally, how are infinitely large things conceptually related to infinitely small things?

To begin to see the embodied source of the idea of infinity, we must look to one of the most common of human conceptual systems, what linguists call the *aspectual system*. The aspectual system characterizes the structure of event-concepts—events as we conceptualize them. Some actions, for example, are inherently iterative, like tapping or breathing. Others are inherently continuous, like moving. Some have inherent beginning and ending points, like jumping. Some just have ending points, like arriving. Others just have starting points, like leaving or embarking on a trip. Those that have ending points also have resulting states. In addition, some actions have their completions conceptualized as part of the action (e.g., landing is part of jumping), while others have their completions conceptualized as external to the actions (e.g., we do not normally conceptualize landing as part of flying, as when a comet flies by the earth).

Of course, in life, hardly anything one does goes on forever. Yet we conceptualize breathing, tapping, and moving as *not having completions*. This conceptualization is called *imperfective aspect*. As we saw in Chapter 2, the concept of aspect appears to be embodied in the motor-control system of the brain. Narayanan (1997), in a study of computational neural modeling, showed that the neural computational structure of the aspectual system is the same as that found in the motor-control system.

Given that the aspectual system is embodied in this way, we can see it as the fundamental source of the concept of infinity. Outside mathematics, a process is seen as infinite if it continues (or iterates) indefinitely without stopping. That is, it has imperfective aspect (it continues indefinitely) without an endpoint. This is *the literal concept of infinity* outside mathematics. It is used whenever one thinks of perpetual motion—motion that goes on and on forever.

Continuative Processes Are Iterative Processes

A process conceptualized as not having an end is called an imperfective process—one that is not "perfected," that is, completed. Two of the subtypes of imperfective processes are *continuative* (those that are continuous) and *iterative* (those that repeat and have an intermediate endpoint and an intermediate result). In languages throughout the world, continuous processes are conceptualized as if they were iterative processes. The syntax used is commonly that of conjunction. Consider a sentence like *John jumped and jumped again, and jumped again*. Here we have an iteration of three jumps. But *John jumped and jumped and jumped* is usually interpreted not as three jumps but as an open-ended, indefinite number.

Now, "jump" is an inherently perfective verb: Each jump has an endpoint and a result. But verbs like *swim, fly,* and *roll* are imperfective, with no indicated endpoint. Consider sentences indicating iteration via the syntactic device of conjunction: *John swam and swam and swam. The eagle flew and flew and* *flew.* This sentence structure, which would normally indicate indefinite iteration with perfective verbs, here indicates a continuous process of swimming or flying. The same is true in the case of aspectual particles like *on* and *over*. For example, *John said the sentence over* indicates a single iteration of the sentence. But *John said the sentence over and over and over* indicates ongoing repetition. Similarly, *The barrel rolled over and over* indicates indefinitely continuous rolling, and *The eagle flew on and on* indicates indefinitely continuous flying. In these sentences, the language of iteration for perfectives (e.g., verb *and* verb *and* verb; *over and over and over*) is used with imperfectives to express something quite different—namely, an indefinitely continuous process. In short, the idea of iterated action is being used in various syntactic forms to express the idea of continuous action. This can be characterized in cognitive terms by the metaphor Indefinite Continuous Processes Are Iterative Processes.

There is a cognitive reason why such a metaphor should exist. Processes in general are conceptualized metaphorically in terms of motion via the event structure metaphor, in which processes are extended motions (see Lakoff & Johnson, 1999). Indefinitely continuous motion is hard to visualize, and for extremely long periods it is impossible to visualize. What we do instead is visualize short motions and then repeat them, thus conceptualizing indefinitely continuous motion as repeated motion. Moreover, everyday continuous actions typically require iterated actions. For example, continuous walking requires repeatedly taking steps; continuous swimming requires repeatedly moving the arms and legs; continuous flying by a bird requires repeatedly flapping the wings. This conflation of continuous action and repeated actions gives rise to the metaphor by which continuous actions are conceptualized in terms of repeated actions.

Why is this metaphor important for infinity? The reason is that we commonly apply it to infinitely continuous processes. Continuous processes without end—infinite continuous processes—are conceptualized via this metaphor as if they were infinite iterative processes, processes that iterate without end but in which each iteration has an endpoint and a result. For example, consider infinitely continuous motion, which has no intermediate endpoints and no intermediate locations where the motion stops. Such infinitely continuous motion can be conceptualized metaphorically as iterated motion with intermediate endings to motion and intermediate locations—but with infinitely many iterations.

This metaphor is used in the conceptualization of mathematics to break down continuous processes into infinitely iterating step-by-step processes, in which each step is discrete and minimal. For example, the indefinitely continuous process of reaching a limit is typically conceptualized via this metaphor as an infinite sequence of well-defined steps.

Actual Infinity

The kind of infinity we have just seen—ongoing processes or motions without end—was called *potential infinity* by Aristotle, who distinguished it from *actual infinity*, which is infinity conceptualized as a realized "thing." Potential infinity shows up in mathematics all the time: when you imagine building a series of regular polygons with more and more sides, when you imagine writing down more and more decimals of $\sqrt{2}$, and so on (see Figure 8.1). But the interesting cases of infinity in modern mathematics are cases of actual infinity—cases that go beyond mere continuous or iterative processes with no end. These include, for example, infinite sets (like the set of natural numbers) and points at infinity—mathematical entities characterized by infiniteness.

We hypothesize that the idea of actual infinity in mathematics is metaphorical, that the various instances of actual infinity make use of the ultimate metaphorical *result* of a process without end. Literally, there is no such thing as the result of an endless process: If a process has no end, there can be no "ultimate result." But the mechanism of metaphor allows us to conceptualize the "result" of an infinite process—in the only way we have for conceptualizing the result of a process—that is, in terms of a process that does have an end.

We hypothesize that all cases of actual infinity—infinite sets, points at infinity, limits of infinite series, infinite intersections, least upper bounds—are special cases of a single general conceptual metaphor in which processes that go on indefinitely are conceptualized as having an end and an ultimate result. We call this metaphor the *Basic Metaphor of Infinity*, or the BMI for short. The target domain of the BMI is the domain of processes without end—that is, what linguists call imperfective processes. The effect of the BMI is to add a metaphorical completion to the ongoing process so that it is seen as having a result—an infinite *thing*.

The source domain of the BMI consists of an ordinary iterative process with an indefinite (though finite) number of iterations with a completion and resultant state. The source and target domains are alike in certain ways:

- Both have an initial state.
- Both have an iterative process with an unspecified number of iterations.
- Both have a resultant state after each iteration.

In the metaphor, the initial state, the iterative process, and the result after each iteration are mapped onto the corresponding elements of the target domain. But the crucial effect of the metaphor is *to add to the target domain the completion of the process and its resulting state.* This metaphorical addition is

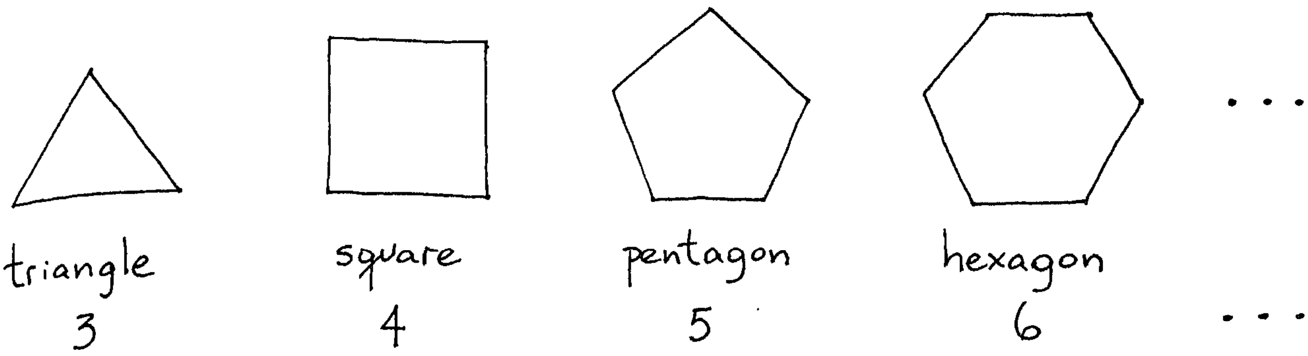


FIGURE 8.1 A case of potential infinity: the sequence of regular polygons with n sides, starting with $n = 3$. This is an unending sequence, with no polygon characterizing an ultimate result.

indicated in boldface in the statement of the metaphor that follows. It is this last part of the metaphor that allows us to conceptualize the ongoing process in terms of a completed process—and so to produce the concept of actual infinity.

THE BASIC METAPHOR OF INFINITY

Source Domain COMPLETED ITERATIVE PROCESSES		*Target Domain* ITERATIVE PROCESSES THAT GO ON AND ON
The beginning state	→	The beginning state
State resulting from the initial stage of the process	→	State resulting from the initial stage of the process
The process: From a given intermediate state, produce the next state.	→	The process: From a given intermediate state, produce the next state.
The intermediate result after that iteration of the process	→	The intermediate result after that iteration of the process
The final resultant state	→	**"The final resultant state" (actual infinity)**
Entailment *E:* The final resultant state is unique and follows every nonfinal state.	→	**Entailment *E:* The final resultant state is unique and follows every nonfinal state.**

Notice that the source domain of the metaphor has something that does not correspond to anything in the literal target domain—namely, a final resultant state.

The conceptual mapping imposes a "final resultant state" on an unending process. The literal unending process is given on the right-hand side of the top three arrows. The metaphorically imposed final resultant state (which characterizes an actual infinity) is indicated in boldface on the right side of the fifth line of the mapping.

In addition, there is a crucial entailment that arises in the source domain and is imposed on the target domain by the metaphor. In any completed process, the final resultant state is *unique*. The fact that it is the *final* state of the process means that:

- There is no earlier final state; that is, there is no distinct previous state within the process that both follows the completion stage of the process yet precedes the final state of the process.
- Similarly, there is no later final state of the process; that is, there is no other state of the process that both results from the completion of the process and follows the final state of the process. Any such putative state would have to be "outside the process" as we conceptualize it.

Thus, the uniqueness of the final state of a complete process is a product of human cognition, not a fact about the external world. That is, it follows from the way we conceptualize completed processes. This entailment is also shown in boldface.

The Basic Metaphor of Infinity maps this uniqueness property for final resultant states of processes onto actual infinity. Actual infinity, as characterized by *any given application of the BMI,* is unique. As we shall see later, the existence of degrees of infinity (as with transfinite numbers) requires multiple applications of the BMI.

What results from the BMI is a metaphorical creation that does not occur literally: a process that goes on and on indefinitely and yet has a unique final resultant state, a state "at infinity." This metaphor allows us to conceptualize "potential" infinity, which has neither end nor result, in terms of a familiar kind of process that has a unique result. Via the BMI, infinity is converted from an open-ended process to a specific, unique entity.

We have formulated this metaphor in terms of a simple iterative step-by-step process—namely: From a given state, produce the next state. From an initial state, the process produces intermediate resultant states. The metaphorical process has an infinity of such intermediate states and a metaphorical final, unique, resultant state.

Notice that the nature of the process is unspecified in the metaphor. The metaphor is general: It covers any kind of process. Therefore, we can form special cases of this general metaphor by specifying what process we have in mind. The process could be a mechanism for forming sets of natural numbers resulting in an infinite set. Or it could be a mechanism for moving further and further along a line until the "point at infinity" is reached. Our formulation of the metaphor is sufficiently precise and sufficiently general so that we can fill in the details of a wide variety of different kinds of infinity in different mathematical domains. Indeed, this will be our strategy, both in this chapter and in chapters to come. We will argue that a considerable number of infinite processes in mathematics are special cases of the BMI that can be arrived at by specifying what the iterative process is in detail. We believe that *all* notions of infinity in mathematics can be seen as special cases of the BMI, and we will discuss what we hope is a wide enough range of examples to make our case.

The Origin of the BMI Outside Mathematics

The Basic Metaphor of Infinity is a general cognitive mechanism. It can occur by itself, as when one speaks of "the infinite" as a thing. But most people don't go around speaking of "the infinite" as a thing; the notion occurs usually only in philosophical or spiritual contexts—or in special cases in which some particular process is under discussion.

There is a long history to the use of this metaphor in special cases, a history that goes back at least to pre-Socratic philosophy. Greek philosophy from its earliest beginnings held that any particular thing is an instance of a higher category. For example, a particular goat is an instance of the category Goat. Moreover, every such category was assumed to be itself a thing in the world. That is, not only were individual goats in the world but the category Goat—a natural kind—was also seen as an entity in the world. This meant that it, too, was an instance of a still higher category, and so on.

Thus there was taken to be an indefinitely long ascending hierarchy of categories. It was further assumed that this indefinitely ascending hierarchy had an end—the category of Being, which encompassed everything (see Lakoff & Johnson, 1999, chs. 16–18). The reasoning implicit in this move from an indefinitely ascending hierarchy to a highest, all-encompassing point in the hierarchy can be seen as an instance of the Basic Metaphor of Infinity, in which the result of an indefinitely iterative process of higher categorization results in a highest category. The following table shows, for each element of the target domain of the BMI, the corresponding element of the special case (the formation of ascending categories). The symbol "⇒" indicates the relation between target domain elements and their corresponding special cases.

THE HIGHEST CATEGORY—BEING

Target Domain ITERATIVE PROCESSES THAT GO ON AND ON		*Special Case of the Target Domain* THE FORMATION OF ASCENDING CATEGORIES
The beginning state	⇒	Specific entities with no categories
State resulting from the initial stage of the process	⇒	Categories of specific entities: The lowest level of categories
The process: From a prior intermediate state, produce the next state.	⇒	From categories at a given level, form categories at the next highest level that encompass lower-level categories.
The intermediate result from that iteration of the process	⇒	Intermediate-level categories formed by each iteration of the process
"The final resultant state" (actual infinity)	⇒	**"The final resultant state": The category of Being**
Entailment *E:* The final resultant state is unique and follows every nonfinal state.	⇒	**Entailment *E:* The category of Being is unique and encompasses all other categories.**

In this case, the process specified in the BMI is the process of forming higher categories. The first step is to form a category from specific cases and make that category a thing. The iterative process is to form higher categories from immediately lower categories. And the final resultant state is the highest category, the category of Being.

What we have just done is consider a special case of the BMI outside mathematics, and we have shown that by considering the category of Being as produced by the BMI, we get all of its requisite properties. Hereafter, we will use the same strategy *within* mathematics to produce special cases of actual infinity, and we will use similar tables to illustrate each special case.

This metaphorical concept of infinity as a unique entity—the highest entity—was extended naturally to religion. In Judaism, the Kabbalistic concept of God is *Ein Sof*—that is, "without end," a single, unique God. The uniqueness entailment applied to deities yields monotheism. In Christianity, God is *all*-powerful: Given an ascending hierarchy of power, that hierarchy is assumed to have a unique highest point, something all-powerful—namely, the power of God. It is no coincidence that Georg Cantor believed that the study of infinity in mathematics had a theological importance. He sought, in his mathematical concept of the infinite, a way to reconcile mathematics and religion (Dauben, 1983).

Processes As Things

Processes are commonly conceptualized as if they were static things—often containers, or paths of motion, or physical objects. Thus we speak of being *in the middle of* a process or at the *tail end* of it. We see a process as having a *length*—*short* or *long*—able to be *extended* or *attenuated, cut short*. We speak of the *parts* of a process, as if it were an object with parts and with a size. This extends to infinite processes as well: Some infinities are *bigger* than other infinities.

As we saw in Chapter 2, one of the most important cognitive mechanisms for linking processes and things is what Talmy (1996, 2000) has called "fictive motion," cases in which an elongated path, object, or shape can be conceptualized metaphorically as a process tracing the length of that path, object, or shape. A classic example is *The road runs through the forest,* where *runs* is used metaphorically as a mental trace of the path of the road.

Processes, as we ordinarily think of them, extend over time. But in mathematics, processes can be conceptualized as atemporal. For example, consider Fibonnacci sequences, in which the $n + 2^{nd}$ term is the sum of the n^{th} term and the $n + 1^{st}$ term. The sequence can be conceptualized either as an ongoing infinite process of producing ever more terms or as a thing, an infinite sequence that is atemporal. This dual conceptualization, as we have seen, is not special to mathematics but part of everyday cognition.

Throughout this chapter, we will speak of infinite processes in mathematics. Sometimes we conceptualize them as extending over time, and sometimes we conceptualize them as static. For our purpose here, the difference will not matter, since we all have conceptual mechanisms (Talmy's fictive-motion schemas) for going between static and dynamic conceptualizations of processes. For convenience, we will be using the dynamic characterization throughout.

What Is Infinity?

Our job in this chapter is to answer a general question, *What is infinity?* Or, put another way, *How do we, mere human beings, conceptualize infinity?* This is not an easy question to answer, because of all the constraints on it. First, the answer must be biologically and cognitively plausible. That is, it must make use of normal cognitive and neural mechanisms. Second, the answer must cover all the cases in mathematics, which on the surface are very different from one another: mathematical induction, points meeting at infinity, the infinite set of natural numbers, transfinite numbers, limits (functions get infinitely close to them), things that are infinitely small (infinitesimal numbers and points), and so on. Third, the answer must be sufficiently precise so that it can be demonstrated that these and other concepts of the infinite in mathematics can indeed be characterized as special cases of a general cognitive mechanism.

Our answer is that that general cognitive mechanism is the Basic Metaphor of Infinity, the BMI. It has both neural and cognitive plausibility. We will devote the remainder of this chapter to showing how it applies in a wide variety of cases. Chapters 9 through 14 extend this analysis to still more cases.

The "Number" ∞

One of the first things that linguists, psychologists, and cognitive scientists learn is that when there are explicit culturally sanctioned warnings not to do something, you can be sure that people are doing it. Otherwise there would be no point to the warnings. A marvelous example of such a warning comes from G. H. Hardy's magnificent classic text, *A Course of Pure Mathematics* (1955; pp. 116–117):

> Suppose that n assumes successively the values 1, 2, 3, ... The word "successively" naturally suggests succession in time, and we may suppose n, if we like, to assume these values at successive moments of time. . . . However large a number we may think of, a time will come when n has become larger than this number. . . . This however is a mere matter of convenience, the variation of n having usually nothing to do with time.
>
> The reader cannot impress upon himself too strongly that when we say that n "tends to ∞" we mean simply that n is supposed to assume a series of values which increases beyond all limit. *There is no number "infinity"*: Such an equation as
>
> $$n = \infty$$
>
> is as it stands *meaningless*: A number n cannot be equal to ∞, because "equal to ∞" means nothing. . . . [T]he symbol ∞ means nothing at all except in the phrase "tends to ∞," the meaning of which we have explained above.

Hardy is going through all this trouble to keep the reader from thinking of infinity as a number just because people do tend to think of infinity as a number, which is what the language "tends to infinity" and "approaches infinity" chosen by mathematicians indicates. Hardy suggests that it is a mistake to think of infinity as a number—a mistake that many people make. If people are, mistakenly or not, conceptualizing infinity as a number, then it is our job as cognitive scientists to characterize the cognitive mechanism by which they are making that "mistake." And if we are correct in suggesting that a single cognitive mechanism—namely, the BMI—is used for all conceptions of infinity, then we have to show how the BMI can be used to conceptualize infinity as a number, whether it is a "mistake" to do so or not.

Cognitive science is, after all, *de*scriptive, not *pre*scriptive. And it must explain, as well, why people think as they do. The "mistake" of thinking of infinity as a number is not random. "∞" is usually used with a precise meaning—as a number in an enumeration, not as a number in a calculation. In "1, 2, 3, . . . , ∞" ∞ is taken as an endpoint in an enumeration, larger than any finite number and beyond all of them. But people do not use ∞ as a number in calculations: We see no cases of "17 times ∞, minus 473," which is of course a meaningless expression, as Hardy correctly points out. The moral here is that there are, cognitively, different uses for numbers—enumeration, comparison, and calculation. As a number, ∞ is used in enumeration and comparison but not in calculation. Even mathematicians use infinity as a number in enumeration, as in the sum of a sequence a_n from $n = 1$ to $n = \infty$:

$$\sum_{n=1}^{\infty} a_n$$

When Hardy warns us not to assume that ∞ is a number, it is because mathematicians have devised notions and ways of thinking, talking, and writing, in which ∞ is a number with respect to enumeration, though not calculation.

Indeed, the idea of ∞ as a number can also be seen as a special case of the BMI. Note that the BMI does not have any numbers in it. Suppose we apply the BMI to the integers used to indicate order of enumeration. The inherent structure of the target domain, independent of the metaphor, has a potential infinity, an unending sequence of ordered integers. The effect of the BMI is to turn this into an actual infinity with a largest "number" ∞.

THE BMI FOR ENUMERATION

Target Domain ITERATIVE PROCESSES THAT GO ON AND ON		*Special Case* THE UNENDING SEQUENCE OF INTEGERS USED FOR ENUMERATION
The beginning state	⇒	No integers
State resulting from the initial stage of the process	⇒	The integer 1

The process: From a prior intermediate state, produce the next state.	⇒	Given integer $n-1$, form the next largest integer n.
The intermediate result after that iteration of the process.	⇒	$n > n-1$
"The final resultant state" (actual infinity)	⇒	**The "integer" ∞**
Entailment *E:* The final resultant state is unique and follows every nonfinal state.	⇒	**Entailment *E:* The "integer" ∞ is unique and larger than every other integer.**

The BMI itself has no numbers. However, the unending sequence of integers used for enumeration (but not calculation) can be a special case of the target domain of the BMI. As such, the BMI produces ∞ as the largest integer used for enumeration. This is the way most people understand ∞ as a number. It cannot be used for calculation. It functions exclusively as an *extremity*. Its operations are largely undefined. Thus, $\infty/0$ is undefined, as is $\infty \cdot 0$, $\infty - \infty$, and ∞/∞. And thus ∞ is not a full-fledged number, which was Hardy's point. For Hardy, an entity either was a number or it wasn't, since he believed that numbers were objectively existing entities. The idea of a "number" that had one of the functions of a number (enumeration) but not other functions (e.g., calculation) was an impossibility for him. But it is not an impossibility from a cognitive perspective, and indeed people do use it. ∞ as *the extreme natural number* is commonly used with the implicit or explicit sequence "1, 2, 3, . . . , ∞" in the characterization of infinite processes. Each such use involves a hidden use of the BMI to conceptualize ∞ as the extreme natural number.

Mathematicians use "1, . . . , *n*, . . . , ∞" to indicate the terms of an infinite sequence. Although the BMI in itself is number-free, it will be notationally convenient from here on to index the elements of the target domain of the BMI with the integers "ending" with metaphorical "∞." For mathematical purists we should note that when an expression like $n-1$ is used, it is taken only as a notation indexing the stage previous to stage n, and not as the result of the operation of subtraction of 1 from n.

Target Domain
ITERATIVE PROCESSES THAT GO ON AND ON

The beginning state (0)
State (1) resulting from the initial stage of the process

The process: From a prior intermediate state $(n-1)$,
produce the next state (n).
The intermediate result after that iteration of the
process (the relation between n and $n-1$)
"The final resultant state" (actual infinity "∞")
Entailment *E*: The final resultant state ("∞") is unique and follows every nonfinal state.

The utility of this notation will become apparent in the next section.

Projective Geometry: Where Parallel Lines Meet at Infinity

In projective geometry, there is an axiom that *all parallel lines meet at infinity.* From a cognitive perspective, this axiom presents the following problems. (1) How can we conceptualize what it means for there to be a point "at infinity?" (2) How can we conceptualize parallel lines as "meeting" at such a point? (3) How can such a conceptualization use the same general mechanism for comprehending infinity that is used for other concepts involving infinity?

We will answer these questions by taking the BMI and filling it in in an appropriate way, with a fully comprehensible process that will produce a final result, notated by "∞," in which parallel lines meet at a point *at* infinity. Our answer must fit an important constraint. The point at infinity must function like any other point; for example, it must be able to function as the intersection of lines and as the vertex of a triangle. Theorems about intersections of lines and vertices of triangles must hold of "points at infinity."

To produce such a special case of the BMI, we have to specify certain parameters:

- A *subject-matter frame,* indicating that the subject matter is geometry—specifically, that it is about lines that intersect at a point.
- The initial step of the process.
- The *iterated process,* or a *step-by-step condition* that links each state resulting from the process to the next state.
- The *resultant state* after each iteration.
- An *entailment* of the uniqueness of the final resultant state.

In the case of projective geometry, we take as the subject-matter frame an isosceles triangle, for reasons that will shortly become clear (see Figure 8.2).

THE ISOSCELES TRIANGLE FRAME

An isosceles triangle, ABC_n with
Sides: AB, AC_n, BC_n
Angles: α_n, β_n
D_n: The distance from A to C_n
Where:
$AC_n = BC_n$
$\alpha_n = \beta_n$
Inference:
If AC_n lies on L_{1n} and BC_n lies on L_{2n},
then L_{1n} intersects L_{2n}.

The iterative process in this case is to move point C_n further and further away from points A and B. As the distance D_n between A and C_n gets larger, the angles α_n and β_n approach 90 degrees more and more closely. As a result, the intersecting lines L_{1n} and L_{2n} get closer and closer to being parallel. This is an unending, infinite process. At each stage n, the lines meet at point C_n. (For details, see Kline, 1962, ch. 11; Maor, 1987.)

The idea of "moving point C_n further and further" is captured by a sequence of moves, with each move extending over an arbitrary distance. By the condition "arbitrary distance" we mean to quantify over *all* the distances for which the condition holds.

Here are the details for filling in the parameters in the BMI in this special case.

PARALLEL LINES MEET AT INFINITY

Target Domain ITERATIVE PROCESSES THAT GO ON AND ON		*Special Case* PROJECTIVE GEOMETRY
The beginning state (0)	$\Rightarrow$	The isosceles-triangle frame, with triangle ABC_0
State (1) resulting from the initial stage of the process	$\Rightarrow$	Triangle ABC_1, where the length of AC_1 is D_1
The process: From a prior intermediate state (n–1), produce the next state (n).	$\Rightarrow$	Form AC_n from AC_{n-1} by making D_n arbitrarily larger than D_{n-1}.
The intermediate result after that iteration of the process (the relation between n and n–1)	$\Rightarrow$	$D_n > D_{n-1}$ and $(90° - \alpha_n) < (90° - \alpha_{n-1})$.

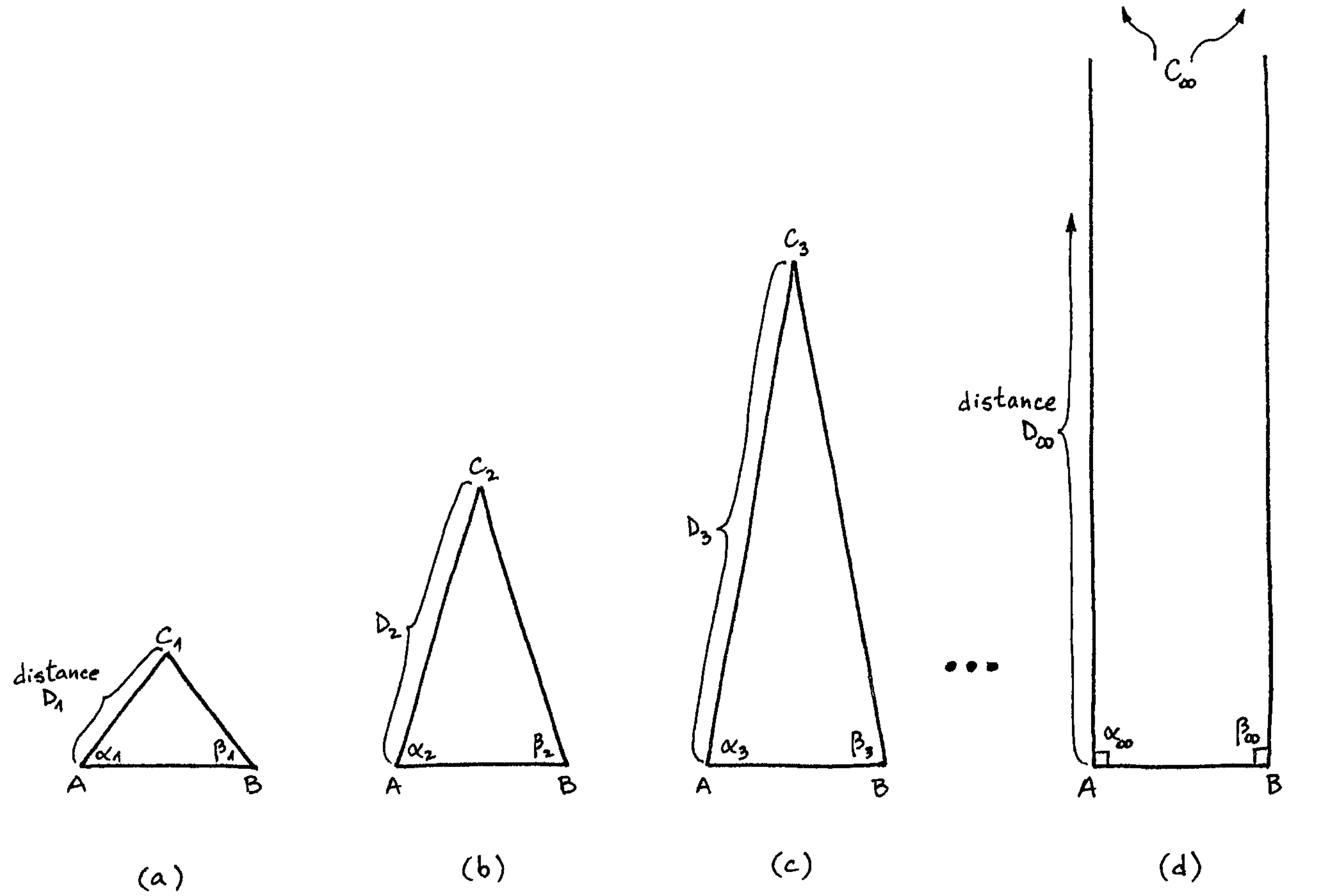


FIGURE 8.2 The application of the Basic Metaphor of Infinity to projective geometry. Drawing (a) shows the isosceles triangle ABC_1, the first member of the BMI sequence. Drawings (b) and (c) show isosceles triangles ABC_2 and ABC_3, in which the equal sides get progressively longer and the equal angles get closer and closer to 90°. Drawing (d) shows the final resultant state of the BMI: ABC_∞, where the angles *are* 90°. The equal sides are infinitely long, and they metaphorically "meet" at infinity—namely the unique point C_∞.

"The final resultant state" (actual infinity "∞")	⇒	**α_∞ = 90°. D_∞ is infinitely long. Sides AC_∞ and BC_∞ are infinitely long, parallel, and meet at C_∞—a point "at infinity."**
Entailment *E:* The final resultant state ("∞") is unique and follows every nonfinal state.	⇒	**Entailment *E:* There is a unique AC_∞ (distance D_∞) that is longer than AC_n (distance D_n) for all finite *n*.**

As a result of the BMI, lines L_1 and L_2 are parallel, meet at infinity, and are separated by the length of line segment AB. Since the length AB was left unspecified, as was the orientation of the triangle, this result will "fit" all lines parallel to L_1 and L_2 in the plane. Thus, this application of the BMI defines the same system of geometry as the basic axiom of projective geometry—namely, that all parallel lines in the plane meet at infinity. Thus, for each orientation there is an infinite family of parallel lines, all meeting "at infinity." Since there is such a family of parallel lines for each orientation in the plane, there is a "point at infinity" in every direction.

Thus, we have seen that there is a special case of BMI that defines the notion "point at infinity" in projective geometry, which is a special case of actual infinity as defined by the BMI. We should remind the reader here that this is a cognitive analysis of the concept "point at infinity" in projective geometry. It is not a mathematical analysis, is not meant to be one, and should not be confused with one. Our claim is a cognitive claim: The concept "point at infinity" in projective geometry is, from a cognitive perspective, a special case of the general notion of actual infinity.

We have at present no experimental evidence to back up this claim. In order to show that the claim is a plausible one, we will have to show that a wide variety of concepts of infinity in mathematics arise as special cases of the BMI. Even then, this will not prove empirically that they are; it will, however, make the claim highly plausible.

The significance of this claim is not only that there is a single general cognitive mechanism underlying all human conceptualizations of infinity in mathematics, but also that this single mechanism makes use of common elements of human cognition—aspect and conceptual metaphor.

The Point at Infinity in Inversive Geometry

Inversive geometry also has a concept of a "point at infinity," but it is a concept very different from the one found in projective geometry. Inversive geometry is defined by a certain transformation on the Cartesian plane. Consider the Cartesian plane described in polar coordinates, in which every point is represented by (r, θ) where θ is an angle and r is the distance from the origin. Consider the transformation that maps r onto $1/r$. This transformation maps the unit circle onto itself, the interior of the unit circle onto its exterior, and its exterior onto its interior. Let us consider what happens to zero under this transformation.

Consider a ray from zero extending outward indefinitely at some angle θ (see Figure 8.3). As r inside the circle gets closer to 0, $1/r$ gets further away. Thus, 1/1,000 is mapped onto 1,000, 1/1,000,000 is mapped onto 1,000,000, and so on. As r approaches 0, $1/r$ approaches ∞. What is the point at 0 mapped onto? It is tempting to map 0, line by line, into a point at ∞ on that line. But there would be many such points, one for each line. What inversive geometry does is define a single "point at infinity" for all lines, and it maps 0, which is unique, onto the unique "point at infinity."

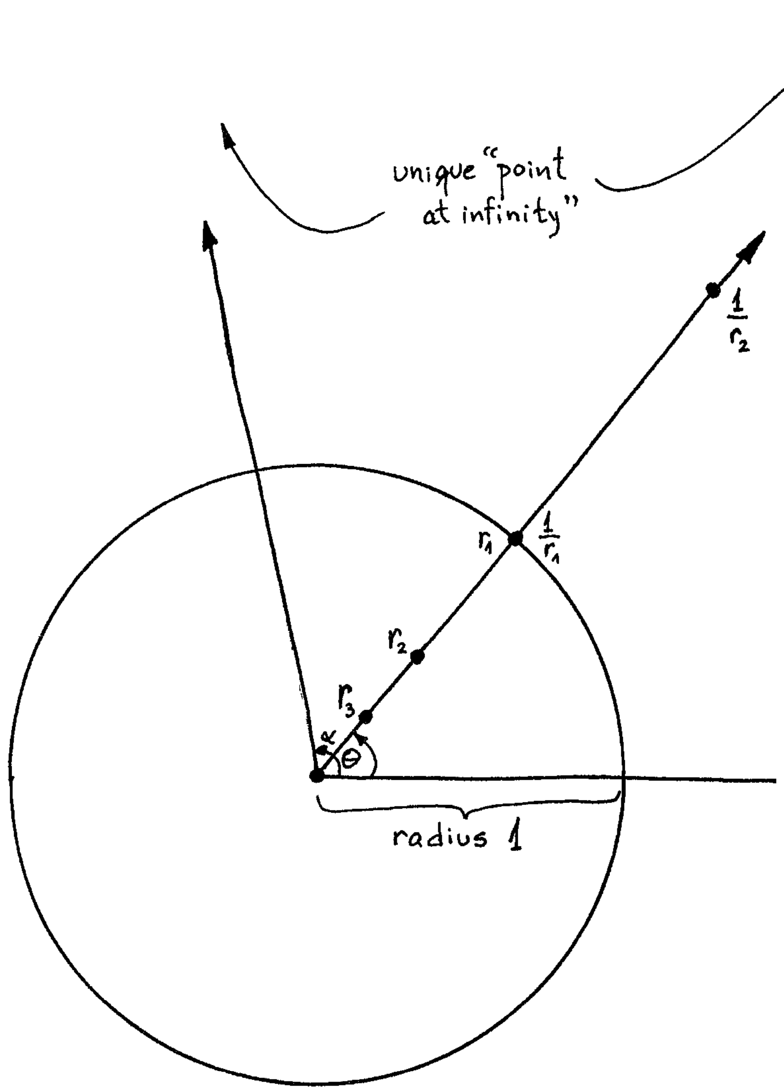


FIGURE 8.3 Inversive geometry: a plane with a circle of radius 1 and rays extending from the center of the circle. Each ray is a nonnegative real-number line, with zero at the center of the circle and 1 on the circle. There is a function $f(x) = 1/x$, mapping points r on each ray inside the circle to corresponding points $1/r$ on that ray outside the circle, and conversely, with zero mapped to ∞. Since there is only one zero, there is only one "point at infinity," which is in all the rays.

If we are correct that the BMI characterizes actual infinity in all of its many forms, then the point at infinity in inversive geometry should also be a special case of the BMI. But, as in most cases, the precise formulation takes some care. Whereas in projective geometry there is an infinity of points at infinity, in inversive geometry there is only one. This must emerge as an entailment of the BMI, given the appropriate parameters: the frame and the iterative process. Here is the frame, which fills in the "subject matter" parameter of the BMI. The frame is straightforward.

THE INVERSIVE GEOMETRY FRAME

The Cartesian plane, with polar coordinates.
The one-to-one function $f(x) = 1/x$, for x on a ray that extends from the origin outward.

This frame picks a special case of the target domain of the BMI. To fill out this special case of the BMI, we need to characterize the iterative process.

The idea behind the process is simple. Pick a ray and pick a point x_1 on that ray smaller than 1. That point has an inverse, which is greater than 1. Keep picking points x_n on the same ray, closer and closer to zero. The inverse points $1/x_n$ get further and further from the origin. This is an infinite, unending process.

We take this process as the iterative process in the BMI. Here are the details:

THE POINT AT INFINITY IN INVERSIVE GEOMETRY

Target Domain ITERATIVE PROCESSES THAT GO ON AND ON		*Special Case* INVERSIVE GEOMETRY
The beginning state (0)	$\Rightarrow$	The origin, with rays projecting outward in every direction, and with an arbitrary ray r at angle θ designated
State (1) resulting from the initial stage of the process	$\Rightarrow$	A point x_1 on r at an arbitrary distance $d_1 < 1$ from the origin. There is an inverse point x_1', at a distance $1/d_1 > 1$ from the origin.
The process: From a prior intermediate state (n–1), produce the next state (n).	$\Rightarrow$	Given point x_{n-1} at distance d_{n-1} from the origin, find point x_n at distance d_n from the origin, where d_n is arbitrarily smaller than d_{n-1}.
The intermediate result after that iteration of the process (the relation between n and n–1)	$\Rightarrow$	There is an inverse point, x_n', at a distance $1/d_n$ from the origin and further from the origin than x_{n-1}' is.
"The final resultant state" (actual infinity "∞")	$\Rightarrow$	**$\mathbf{x}_\infty$ is at the origin; that is, its distance from the origin is zero. $\mathbf{x}_\infty'$ is a unique point infinitely far from the origin.**
Entailment *E:* The final resultant state ("∞") is unique and follows every nonfinal state.	$\Rightarrow$	**Entailment *E:* $\mathbf{x}_\infty$ is unique and is closer to the origin than any other point on ray *r*. There is a unique $\mathbf{x}_\infty'$ that is further from the origin than any other point on ray *r*.**

The BMI then applies to this infinite, unending process and conceptualizes it metaphorically as having a unique final resultant state "∞." At this metaphorical final state, x_∞ is at the origin, and the distance from the origin to x_∞' is infinitely long. In inversive geometry, arithmetic is extended to include ∞ as a number with respect to division: $D_\infty = \infty$, $1/D_\infty = 1/\infty = 0$, and $1/0 = \infty$. This extension of arithmetic says nothing about the addition or subtraction of ∞. It is peculiar to inversive geometry because of the way this special case of the BMI is defined: All points must have inverses, including the point at the origin. Where dividing by zero is normally not possible, this metaphor extends ordinary division to give a metaphorical value to $1/0$ and $1/\infty$. In inversive geometry, ∞ does exist as a number and has limited specified possibilities for calculation.

Moreover, since there is only one zero point shared by all rays, and since $f(x) = 1/x$ is a one-to-one mapping that, via this metaphor, maps 0 onto ∞ and ∞ onto 0, there must therefore be only one ∞ point shared by all rays.

What is interesting about this case is that the same general metaphor, the BMI, produces a concept of the point at infinity very different from that in the case of projective geometry. In projective geometry, there is an infinity of points at infinity (for an image, think of the horizon line), while in inversive geometry there is only one. Moreover, projective geometry has no implicit associated arithmetic, while inversive geometry has an implicit arithmetic (differing from normal arithmetic in its treatment of zero and infinity).

Finally, though we have characterized inversive geometry in terms of a cognitive mechanism, the BMI, mathematicians of course do not. They simply define inversive geometry in normal mathematical terms. Our goal is to show how the same concept of infinity is involved in inversive geometry and other forms of mathematics, while respecting the differences in the concept of infinity across branches of mathematics.

The Infinite Set of Natural Numbers

Within formal arithmetic, the natural numbers are usually characterized by the successor operation: Start with 1. Add 1 to yield a result. Add 1 to the result to yield a new result. And so on. This is an unending, infinite process. It yields the natural numbers, one at a time. Not the *infinite set* containing *all* the natural numbers. Just the natural numbers themselves, each of which is finite. Since it is incapable of being used in calculation, ∞ is not a full-fledged member of the infinite set of natural numbers.

Here is the problem of characterizing the set of natural numbers. The set must be infinite since it contains *all* of the infinitely many numbers, but it cannot contain ∞ as a number.

To get the *set* of natural numbers, you have to collect up each number as it is formed. The set keeps growing without end. To get the *entire* set of natural numbers—all of them, even though the set never stops growing—you need something more. In axiomatic set theory, you add an axiom that simply stipulates that the set exists. From a cognitive perspective, that set can be constructed conceptually via a version of the Basic Metaphor of Infinity. The BMI imposes a metaphorical completion to the unending process of natural-number collection. The result is the *entire collection*, the set of *all* natural numbers!

THE SET OF ALL NATURAL NUMBERS

Target Domain ITERATIVE PROCESSES THAT GO ON AND ON		*Special Case* THE SET OF NATURAL NUMBERS
The beginning state (0)	⇒	The natural number frame, with a set of existing numbers and a successor operation that adds 1 to the last number and forms a new set
State (1) resulting from the initial stage of the process	⇒	The empty set, the set of natural numbers smaller than 1.
The process: From a prior intermediate state (n–1), produce the next state (n).	⇒	Given S_{n-1}, the set of natural numbers smaller than n–1, form $S_{n-1} \cup \{n-1\} = S_n$.
The intermediate result after that iteration of the process (the relation between n and n–1)	⇒	At state n, we have S_n, the set of natural numbers smaller than n.
"The final resultant state" (actual infinity "∞")	⇒	**S_∞, the set of *all* natural numbers smaller than ∞—that is, the set of *all* natural numbers (which does not include ∞ as a number).**
Entailment *E:* The final resultant state ("∞") is unique and follows every nonfinal state.	⇒	**Entailment *E:* The set of all natural numbers is unique and includes every natural number (no more, no less).**

This special case of the BMI does the same work from a cognitive perspective as the axiom of Infinity in set theory; that is, it ensures the existence of an infinite set whose members are all the natural numbers. This is an important point. The BMI, as we shall see, is often the conceptual equivalent of some axiom that guarantees the existence of some kind of infinite entity (e.g., a least upper bound). And just as axioms do, the special cases of the BMI determines the right set of inferences required.

This example teaches us something to be borne in mind throughout the remainder of this book. The meaning of "all" or "entire" is far from obvious when they are predicated of infinite sets. In general, the meaning of "all" involves completeness. One of the first uses of "all" learned in English is the child's use of "all gone" to indicate the completion of a process of consumption, removal, or destruction. *All twelve of the paintings were stolen* indicates completeness of the theft—the *entire* collection was stolen. But in the case of infinity, there is no such thing as the literal completeness of an infinite process. The BMI is necessary in some form—implicit or explicit—to characterize any "all" that ranges over an infinite process. Wherever there is infinite totality, the BMI is in use.

This BMI analysis shows us exactly how the same notion of infinity used to comprehend points at infinity can also be used to conceptualize the infinite set of natural numbers. In this case, we can see a somewhat different version of the concept of infinity. Points at infinity give us a concept of an *infinite extremity*. But the set of natural numbers is not an infinite extremity; rather, it is an *infinite totality*. Totalities always involve sets. Extremities only involve linear orderings.

Mathematical Induction

The process of mathematical induction is crucial in mathematical proofs. It works as follows: Prove that

1. The statement S is true for 1;
2. If the statement S is true for n–1, then it is true for n.

Literally, all this does is provide an unending, infinitely unfolding sequence of natural numbers for which the statement S is true. It does not prove that the statement is true for *all* natural numbers. Something is needed to go from the unending sequence of individual natural numbers, for which S is a single general truth, to *all* natural numbers.

In axiomatic arithmetic, there is a special axiom of Mathematical Induction needed to bridge the gap between the unending sequence of *specific* truths for each number, one at a time, and the single generalization to the infinite totality of *all* natural numbers. The axiom simply postulates that if (1) and (2) are proved, then the statement is true for every member of the set of *all* natural numbers.

The axiom of Mathematical Induction is the equivalent of the Basic Metaphor of Infinity applied to the subject matter of inductive proof. To see why, we can conceptualize the content of the axiom in terms of the BMI. Note that this instance of the BMI involves an infinite totality—the infinite set of all finite natural numbers—not an infinite extremity such as the "number" ∞. In this version of the BMI, a set is built up of all the finite natural numbers that the given statement is true of. At the final resultant state, there is an infinite set of finite numbers.

THE BMI FOR MATHEMATICAL INDUCTION

Target Domain ITERATIVE PROCESSES THAT GO ON AND ON		*Special Case* MATHEMATICAL INDUCTION
The beginning state (0)	⇒	A statement $S(x)$, where x varies over the set of natural numbers
State (1) resulting from the initial stage of the process	⇒	$S(1)$ is true for the members of {1}—the set containing the number 1.
The process: From a prior intermediate state (n–1), produce the next state (n).	⇒	Given the truth of $S(n-1)$ for the members of the set $\{1, \ldots, n-1\}$, establish the truth of $S(n)$ for the members of $\{1, \ldots, n\}$.
The intermediate result after that iteration of the process (the relation between n and n–1).	⇒	$S(n)$ is true for the members of the set $\{1, \ldots, n\}$.
"The final resultant state" (actual infinity "∞")	⇒	**$S(\infty)$ is true for the members of the set of *all* natural numbers.**
Entailment *E*: The final resultant state ("∞") is unique and follows every nonfinal state.	⇒	**Entailment *E*: The set of natural numbers for which $S(\infty)$ is true is unique and includes *all* finite natural numbers.**

Thus, mathematical induction can also be seen as a special case of the BMI.

Generative Closure

The idea of generative closure for operations that generate infinite sets also makes implicit use of the BMI. For example, suppose we start with the set containing the integer 1 and the operation of addition. By adding 1 to itself at the first stage, we get 2, which is not in the original set. That means that the original set was not "closed" under the operation of addition. To move toward closure, we can then extend that set to include 2. At the next stage, we perform the binary operation of addition on 1 and 2, and on 2 and 2, to get new elements 3 and 4. We then extend the previous set by including 3 and 4. And so on. This defines an infinite sequence of set extensions. If we apply the BMI, we get "closure" under addition—the set of *all* resulting extensions.

This will work starting with any finite set of elements and any finite number of binary operations on those elements. At least in this simple case, the concept of closure can be seen as a special case of the BMI. We can also see from this case how to state the general case of closure in terms of the BMI. We let C_0 be any set of elements and O be any finite set of operations, each applying to some element or pair of elements in C_0. Here is the iterative process:

- At stage 1, apply every operation once in every way it can apply to the elements of C_0. Collect the results in a set S_0 and form the union of S_0 with C_0 to yield C_1.
- At stage 2, apply every operation once in every way it can apply to the elements of C_1. Collect the results in a set S_1 and form the union of S_1 with C_1 to yield C_2.
- And so on.

If this process fails to yield new elements at any stage—that is, if $C_{n-1} = C_n$—then the closure is finite and the process stops. Otherwise it goes on. Note that for each n, C_n contains as a subset all of the C_i, for $i < n$.

Letting this be the iterative process, we can characterize infinite closure using the Basic Metaphor of Infinity in the following way. At ∞, C_∞ contains every C_n, for $n < \infty$. That is, it contains every finite combination of operations.

THE BMI FOR GENERATIVE CLOSURE

Target Domain ITERATIVE PROCESSES THAT GO ON AND ON		*Special Case* GENERATIVE CLOSURE
The beginning state (0)	⇒	A finite set of elements C_0 and a finite collection O of binary operations on those elements
State (1) resulting from the initial stage of the process	⇒	C_1 = The union of C_0 and the set S_0 of elements resulting from a single application of each operator in O to each pair of elements of C_0
The process: From a prior intermediate state (n–1), produce the next state (n).	⇒	Given C_{n-1}, form S_{n-1}, the set of elements resulting from a single application of each operator in O to each pair of elements of C_{n-1}. $C_n = C_{n-1} \cup S_{n-1}$.

The intermediate result after that iteration of the process (the relation between n and $n\text{–}1$).	$\Rightarrow$	At state n, we have C_n = the union of C_0 and the union of all S_k, for $k < n$. If $C_{n-1} = C_n$, then the closure is finite and the process ends. Otherwise it continues.
"The final resultant state" (actual infinity "∞")	$\Rightarrow$	**C_∞ = the infinite closure of C_0 and its extensions under the operations in O.**
Entailment *E*: The final resultant state ("∞") is unique and follows every nonfinal state.	$\Rightarrow$	**Entailment *E*: C_∞ is unique and includes all possible finite iterations of applications of the operations in O to the elements of C_0 and elements resulting from those operations.**

Notice that although the closure is infinite, there are no infinite sequences of operations. Each sequence of operations is finite, but there is no bound to their length.

"All"

It is commonplace in formal logic to use the symbol for the universal quantifier "$\forall$" in statements or axioms concerning such entities as the set of *all* natural numbers. This symbol is just a symbol and requires an interpretation to be meaningful. In formal logic, interpretations of such quantifiers are given in two ways: (1) via a metalanguage and (2) via a mapping onto a mathematical structure—for instance, the generative closure produced from the set {1} under the operation "+", written "Closure [{1}, +]".

1) "($\forall x$: $x \in N$) Integer (x)" is true if and only if x is an integer for *all* members of the set of natural numbers, N.
2) "($\forall x$: $x \in N$) Integer (x)" is true if and only if x is a member of Closure [{1}, +].

In statement (1) above, the word "all" occurs in the metalanguage expression on the right of "if and only if." In statement (2), the closure of {1} and its extensions under the operation of addition occurs on the right of "if and only if."

Now, in statement (1), the symbol $\forall$ is defined in terms of a prior understanding of the word "all." From a cognitive point of view, this just begs the question of what ∀ means, since it requires a cognitive account of what "all" means when applied to infinite sets. Cognitively, "all" is understood in terms of a linear scale of inclusion, with "none" at one endpoint, "all" at the other, and values like "some," "most," and "almost all" in between. To give a cognitive account of the meaning of "all" applied to an infinite set, we need both the scale of inclusion and a cognitive account of infinite sets, which is given by the BMI.

Statement (2) is an attempt to characterize the meaning of ∀ for a given infinite set without using the word "all" and instead using the concept of generative closure. From a cognitive perspective, this requires a cognitive account of the meaning of generative closure for infinite sets, which we have just given in terms of the BMI.

The moral is that, in either case, one version of the BMI or another is needed to give a cognitive characterization of the meaning of ∀ when it is applied to infinite sets. The use of the symbol ∀ in symbolic logic does not get us out of the problem of giving a cognitive characterization of infinite sets. As far as we can tell, the BMI does suffice in all cases, as we shall see in the chapters to come.

Up to this point, we have shown six diverse mathematical uses of actual infinity which can each be conceptualized as a special case of the Basic Metaphor of Infinity. Each case involves either *infinite extremity* (points at infinity and the number ∞) or *infinite totality* (the set of natural numbers and mathematical induction).

The Basic Conclusions

By comparing the special cases, one can see precisely how a wide variety of superficially different mathematical concepts have a similar structure from a cognitive perspective. Since much of mathematics is concerned with comparisons of kinds of mathematical structures, this sort of analysis falls within the tradition of structural comparison in mathematics. These analyses can be seen as being very much in the tradition of algebra or category theory, in that they seek to reveal deep similarities in cases that superficially look dissimilar.

The difference, of course, is that we are discussing conceptual structure from a *cognitive* perspective, taking into account cognitive constraints, rather than from a purely mathematical point of view, which has no cognitive constraints whatsoever. That is, mathematicians are under no obligation to try to understand how mathematical understanding is embodied and how it makes use of normal cognitive mechanisms, like image schemas, aspectual structure, conceptual metaphors, and so on.

We are hypothesizing that the analyses given above are correct from a cognitive perspective. This hypothesis is testable, at least in principle. (See Gibbs, 1994, for an overview of literature for testing claims about metaphorical thought.) The hypothesis involves certain subhypotheses.

- The BMI is part of our *unconscious* conceptual system. You are generally not aware of conceptual mechanisms while you are using them. This should also hold true of the BMI. It is correspondingly *not* claimed that the use of the BMI is conscious.
- The concept of actual infinity, as characterized by the BMI, makes use of a metaphor based on the concept of aspect—imperfective aspect (in ongoing events) and perfective aspect (in completive events).
- As with any cognitive model, it is hypothesized that the entailments of the model are true of human cognition.

What we have provided in this chapter is a mathematical idea analysis of various concepts of infinity. Mathematics itself does not characterize what is common among them. The common mathematical notion for infinity—". . . ," as in the sequence "1 + 1/2 +1/4 + . . . "—does not even distinguish between potential and actual infinity. If it is potential infinity, the sum only gives an endless sequence of partial sums always less than 2; if it is actual infinity, the sum is exactly 2. Our idea analysis in terms of the BMI makes explicit what is implicit.

In addition, our mathematical idea analysis provides a clear and explicit answer to the question of how finite embodied beings can have a concept of infinity—namely, via conceptual metaphor, specifically the BMI! And it shows the relationships among the various ideas of infinity found in mathematics.

9

Real Numbers and Limits

IT IS NO SURPRISE THAT THE CONCEPT OF INFINITY is central to the concepts of real numbers and limits. The idea of a limit arose in response to the fact that there are infinite series whose sums are finite. Real numbers are fundamentally conceptualized using the concept of infinity: in terms of infinite decimals, infinite sums, infinite sequences, and infinite intersections. Since we are hypothesizing that there is a single general concept of actual infinity characterized via the BMI, we will be arguing that the conceptualization of the real numbers inherently uses the BMI.

There is no small irony here. The real numbers are taken as "real." Yet no one has ever seen a real number. The physical world, limited at the quantum level to sizes no smaller than the Planck length, has nothing as fine-grained as a real number. Computers, which do most of the mathematical calculations in the world, can use only "floating-point numbers" (see Chapter 15), which are limited in information to a small number of bits, usually 32. No computers use—or could use—real numbers in their infinite detail.

We will be hypothesizing that the real numbers, which require the concept of infinity, are conceptualized using the Basic Metaphor of Infinity. The irony is that what are called "real" numbers are fundamentally metaphorical in their conceptualization. This has important consequences for an understanding of what mathematics is (see Chapters 15 and 16). For now, we will simply show how the BMI can conceptualize the mathematical ideas required to characterize the real numbers: infinite decimals, infinite polynomials, limits of infinite sequences, least upper bounds, and infinite intersections of intervals. We will put off discussing Dedekind cuts until Chapter 13, where we will argue that they, too, are instances of the BMI.

Numerals for the Natural Numbers

In building up to the infinite decimals, we have to start with the numerals for the natural numbers in decimal notation—that is, using the digits 0, 1, . . . , 9. To get the set of *all* numerals for natural numbers, just stringing digits one after another without end is not enough. That will give us a perpetually growing class (a potential infinity) but not *all* the numerals for *all* the natural numbers (an actual infinity). To get the infinite set with *all* the numerals, we need to use a special case of the BMI.

Here's how the BMI process works to build up an infinity of indefinitely long finite decimal representations for the infinite set of indefinitely large finite natural numbers. We will start at the beginning state with N_1, the empty set. In the first stage of the process, we will add the one-place decimal representations to form the set $N_2 = \{1, \ldots, 9\}$. In the second stage, we will form N_3, the set of 2-place strings of numerals from the members of N_2 as follows: If s is a member of N_2, it is a member of N_3. Also, s followed by one of the numerals 0, 1, . . . , 9 is also a member of N_3. Thus, for example, let s be 5. Then the strings of digits 50, 51, . . . , 59 are all members of N_3. In this way, we get $N_3 = \{1, \ldots, 9, 10, 11, \ldots, 99\}$, the set of all one- or two-place numerals. We then generalize this process and use it as the iterative process in the BMI, as follows:

NUMERALS FOR NATURAL NUMBERS

Target Domain ITERATIVE PROCESSES THAT GO ON AND ON		*Special Case* THE INFINITE SET OF NUMERALS FOR THE NATURAL NUMBERS
The beginning state (1)	⇒	N_1: The empty set
State (2) resulting from the initial stage of the process	⇒	The set of numerals $N_2 = \{1, \ldots, 9\}$
The process: From a prior intermediate state (n–1), produce the next state (n).	⇒	Let N_{n-1} be the set of numerals whose number of digits is less than n–1. Let s be a string of digits in N_{n-1}. The members of N_n take one of the following forms: s, or s0, or s1, . . . , or s9.
The intermediate result after that iteration of the process	⇒	The set of numerals N_n, each member of which has less than n digits
"The final resultant state" (actual infinity "∞")	⇒	**N_∞: the set containing all the numerals for all the natural numbers**

Entailment *E:* The final resultant state ("∞") is unique and follows every nonfinal state.	⇒	**Entailment *E:* N_∞ is unique and contains all the numerals for the natural numbers.**

At each stage n, we get all the numerals with *less than n* digits. At the "final state," where n is the metaphorical "number" ∞, we get the *infinite totality* of all the finitely long numerals whose number of digits is less than ∞—namely, all the finitely long numerals.

Although our cognitive model of this use of the BMI is complicated, what it describes is something we all find quite simple. This is normal with cognitive models, whose job is to characterize every minute detail of what we understand intuitively as "simple."

Infinite Decimals

Consider $\pi = 3.14159265\ldots$ π is a precise number characterized by an infinitely long string of particular digits to the right of the decimal point. This is not just a sequence that gets longer and longer but an *infinitely long* fixed sequence—a thing. It is a particular infinite sequence, not an ongoing process. This is a case of actual infinity, not potential infinity.

The same is true of any infinite decimal:

$$147963.674039839275\ldots$$

Though we cannot write it down, the ". . . " is taken as representing a fixed infinitely long sequence of particular digits, again a case of actual infinity.

Thus, each infinite decimal can be seen as composed of two parts:

a. A finite sequence of digits to the left of the decimal point, which is a numeral indicating a natural number.
b. An infinitely long sequence of digits to the right of the decimal point, indicating a real number between zero and one.

We have already shown how the BMI characterizes (a). We now need to show how a further use of a special case of the BMI characterizes (a) with (b) added to the right.

We start out with the numerals for the natural numbers—those to the left of the decimal point. That is the set N_∞, which was characterized in the previous section.

INFINITE DECIMALS

Target Domain ITERATIVE PROCESSES THAT GO ON AND ON		*Special Case* NUMERALS FOR REAL NUMBERS
The beginning state (0)	⇒	R_0: the set consisting of members of N_∞, each followed by a decimal point
State (1) resulting from the initial stage of the process	⇒	R_1: For each string of digits s in R_0, $s0$, $s1$, $s2$, . . . , $s9$ is in R_1.
The process: From a prior intermediate state (n–1), produce the next state (n).	⇒	The members of R_n are of one of the following forms: either $s0$, or $s1$, . . . , or $s9$, where the string of digits s is a member of R_{n-1}.
The intermediate result after that iteration of the process	⇒	R_n: the set of numerals whose members have n digits after the decimal point
"The final resultant state" (actual infinity "∞")	⇒	**R_∞: the infinite set of numerals whose members have an infinite number of digits after the decimal point**
Entailment *E:* The final resultant state ("∞") is unique and follows every nonfinal state.	⇒	**Entailment *E:* R_∞ is unique and its members all have more digits after the decimal point than any members of any other R_n.**

Given the numerals for the natural numbers at the beginning, we get the resultant set R_1 by adding one digit (either 0, 1, . . . , or 9) after the decimal point. From this, we get the next set R_2 by the specified process—namely, adding another digit to the members of R_1. This gives us the numbers to two decimal places by the end of state 2. By the end of state n, we have the set of all the numbers to n decimal places. The final state given by the BMI is a metaphorical ∞^{th} resultant state in which the number of digits after the decimal point is infinite. The result of this special case of the BMI is the set of all infinite decimals.

This special case of the BMI creates two kinds of infinities at once: an *infinite totality* (the set consisting of the infinity of infinitely long decimals), and within each set, each member is an *infinite extremity* (a single numeral of infinite length).

Note, incidentally, that the number 4 in this notation is represented by the infinite decimal 4.0000. . . . Thus, all numerals characterized by this version of the BMI are infinite decimals, even numerals for natural numbers like 4.

These are the numerals for all the real numbers. They are not the real numbers themselves but only names for them. Moreover, we have not yet specified how to associate each name with a corresponding real number. This is a nontrivial job, since each name is infinitely long and there is an uncountable infinity of real numbers. The way this is accomplished by mathematicians is to associate each of these names with a unique infinite polynomial and to represent each real number as a unique infinite polynomial. This, too, requires a use of the BMI.

Infinite Polynomials

A polynomial is a sum—a sum of products of a certain form. A simple polynomial is: $(4 \cdot 10^3) + (7 \cdot 10^2) + (2 \cdot 10^1) + (9 \cdot 10^0)$. This polynomial corresponds to the natural number written in decimal notation as 4,729. Such simple polynomials are said to have a "base" 10, since they are sums of multiples of powers of 10. The preceding polynomial has the form:

$$(a_3 \cdot 10^3) + (a_2 \cdot 10^2) + (a_1 \cdot 10^1) + (a_0 \cdot 10^0),$$

where $a_3 = 4$, $a_2 = 7$, $a_1 = 2$, and $a_0 = 9$. Polynomials are numbers; decimals are numerals—names for polynomials. 83.76 is the numeral naming the polynomial: $(8 \cdot 10^1) + (3 \cdot 10^0) + (7 \cdot 10^{-1}) + (6 \cdot 10^{-2})$.

In general, a polynomial has the form:

$$a_n \cdot 10^n + \ldots + a_0 \cdot 10^0 + a_{-1} \cdot 10^{-1} + a_{-2} \cdot 10^{-2} + \ldots + a_{-k} \cdot 10^{-k},$$

where n is a non-negative integer and k is a natural number and a_n and a_{-k} are integers between 0 and 9. For example, in 459.6702, $a_2 = 4$ and $a_{-4} = 2$.

Since we have already shown how the infinite decimals can be characterized using special cases of the BMI, we can now map the infinite decimals conceptualized in this way onto corresponding infinite polynomials. Each infinite decimal has the form:

$$a_n \ldots a_0 \, . \, a_{-1} \ldots a_{-k} \ldots$$

It corresponds to an infinite polynomial:

$$a_n \cdot 10^n + \ldots + a_0 \cdot 10^0 + a_{-1} \cdot 10^{-1} + \ldots + a_{-k} \cdot 10^{-k} + \ldots$$

In other words, there is a symbolic map characterizing the one-to-one relationship between infinite decimals and infinite polynomials.

HOW NUMERALS SYMBOLIZE POLYNOMIALS

Conceptual Domain INFINITE POLYNOMIALS		*Symbolic Domain* INFINITE DECIMALS
$a_n \cdot 10^n + \ldots + a_0 \cdot 10^0 + a_{-1} \cdot 10^{-1} + \ldots + a_{-k} \cdot 10^{-k} + \ldots$	$\leftrightarrow$	$a_n \ldots a_0 \,.\, a_{-1} \ldots\ a_{-k} \ldots$

By virtue of this one-to-one mapping, the BMI structure of infinite decimals is mapped onto infinite polynomials.

Infinite polynomials are real numbers. Indeed, one of the *definitions* of real numbers is that they are infinite polynomials. The relationship between real numbers and infinite polynomials will become clearer as we proceed.

The irony should not be lost in the details: It takes a metaphor—the BMI—to conceptualize the "real" numbers.

Limits of Infinite Sequences

An infinite sequence of real numbers is commonly conceptualized as a function from the natural numbers to the real numbers. The natural numbers constitute an infinite set conceptualized via the BMI. Thus, the BMI is used implicitly in conceptualizing an infinite sequence. Each infinite sequence is usually conceptualized as an infinite set of ordered pairs (i, r_i), where the i's are the members of the set of natural numbers and each appears exactly once in the first place of some pair, while the second place in each pair is a real number. The use of the BMI in characterizing what is infinite about an infinite sequence is therefore straightforward: It is like the use of the BMI to characterize the infinite set of natural numbers, since there is one term of a sequence for each natural number.

The notion of a limit of an infinite sequence is more complex. A limit is a real number that the values of the sequence "get infinitely close to" as the number of terms increases "to infinity." We normally conceptualize the "convergence" of an infinite sequence to a limit by means of the concept of "approaching": The sequence "approaches" the limit as the number of terms "approaches infinity." That is, the value of x_n gets progressively closer to L (for limit) as n gets progressively "closer to infinity."

Lurking in the background here is the spatial metaphor that numbers are points on a line. Metaphorically, the limit is a fixed point L on the number line. As the number of terms n gets progressively "closer to infinity," the values x_n of the sequence get progressively closer to point L. As "n approaches ∞," the sequence comes "infinitely close to" L. It is this conception of the limit of a sequence that we shall attempt to characterize via the BMI. The difficulty will be to show just what it means for natural numbers to "approach infinity" and for the terms of the sequence to come "infinitely close" to the limit.

The metaphors involved here are fairly obvious. Nothing is literally moving or "approaching" anything. The value of each term of a sequence is fixed; since the values don't change, the terms can't literally "approach." And infinity is not something you can literally get "close to."

There are three reasons for undertaking this task. First, we want to show exactly how the BMI enters into the notion of a limit. Second, we want to show how limits are usually conceptualized. And third, we want to show that there is a precise arithmetic way to characterize limits using the BMI.

Before we begin, however, we should contrast the concept of a limit as we will describe it with the formal definition usually used. The formal definition of a limit of a sequence does not capture the idea of "approaching" a limit. Here is the formal definition.

> The sequence $\{x_n\}$ has L as a limit if, for each positive number ε, there is a positive integer n_0 with the property that $|x_n - L| < \varepsilon$ for all $n \geq n_0$.

This is simply a static condition using the quantifiers "for each" and "there is." This static condition happens to cover the case where $\{x_n\}$ "approaches L" as n "approaches infinity." It also happens to cover all sorts of other irrelevant cases where there are combinations of epsilons that have nothing to do with "approaching" anything.

Since such irrelevant cases are usually not discussed in textbooks, it is worthwhile for our purposes to give an example here. Consider the sequence: $\{x_n\} = \frac{n}{n+1}$. The terms are: 1/2, 2/3, 3/4, 4/5, ... The limit is 1. In the formal definition, epsilon ranges over all positive numbers. Take the positive number 43. No matter what n_0 one picks, the condition will be met. For example, let $n = 999$. Then $|x_n - L| = |999/1000 - 1| = 0.001$. This is certainly less than 43. So what? This choice of values is irrelevant to the question of whether the sequence converges to a limit. Our normal, everyday understanding of converging to a limit has a concept of a limit in which such irrelevant values don't occur. The point is not that the irrelevant values are mathematically harmful; they aren't. The definition works perfectly well from a mathematical perspective. And the fact that irrelevant cases fit the definition does not matter from a mathematical perspective.

But it does matter from a cognitive perspective. We normally conceptualize limits using an idea of a limit without such irrelevant cases. We hypothesize that a special case of the BMI is used, and that it can be formulated precisely.

Our formulation uses the idea of the "process" in the BMI to characterize the process of "approaching" in the prototypical case. The process is one of getting "progressively closer," so that at each stage n the "distance" between the limit L and x_n becomes smaller. "Distance" in the geometric metaphor Numbers Are Points on a Line (see Chapter 12) is characterized metaphorically in terms of arithmetic difference: the absolute value $|x_n - L|$, which must become smaller as n becomes larger. That "distance" should "approach zero" as n "approaches infinity."

This is achieved using the BMI in the following way.

- The concept "n gets progressively larger" is characterized by making the iterative process in the BMI the addition of 1 to n, which is the way we understand "n getting progressively larger" in a sequence.
- The concept "n approaches infinity" is characterized via the BMI, which creates a metaphorical final, ∞^{th} resultant state of the process.
- The concept "approach" implicitly uses the metaphor Numbers Are Points on a Line, where the *distance* between points is metaphorically the *difference* between numbers. The magnitude of metaphorical "distance" between a term of the sequence x_n and the limit L is thus a positive real number r_n specifying the difference $|x_n - L|$.
- To go into greater detail, the "distance" between x_n and the limit L is the interval on the number line between x_n and L. According to the metaphor A Line Is a Set of Points, that interval is metaphorically a set of points. Since each such "point" is metaphorically a real number, the set of such points is, again metaphorically, the set R_n of real numbers r greater than zero and less than $|x_n - L|$.
- The "approach" is characterized by specifying the iterative process in the BMI to be the process by which the remaining metaphorical distance to the limit—the arithmetical difference $|x_n - L|$—gets smaller, and hence the set R_n excludes more and more real numbers. In other words, $R_n \subset R_{n-1}$.
- As the process continues, more and more terms of the sequence are generated. We collect them step by step in sets, so that at stage n we have set S_n, which contains the first n terms of the sequence. Thus, the set formed at step n contains all the terms of the sequence generated so far.
- By the final stage of the BMI, the ∞^{th} stage, all of the terms of the sequence have been generated and collected in the set S_∞. Moreover, all the sets R_n of real numbers r, such that $0 < r < |x_n - L|$, have been generated. As x_n gets progressively closer to L, the set R_n comes to exclude more and more real numbers, until, at the ∞th stage of the process, R_∞ is completely empty. That is, x_n has gotten so close to L that there is no positive real number r such that for every finite n, $0 < r < |x_n - L|$.

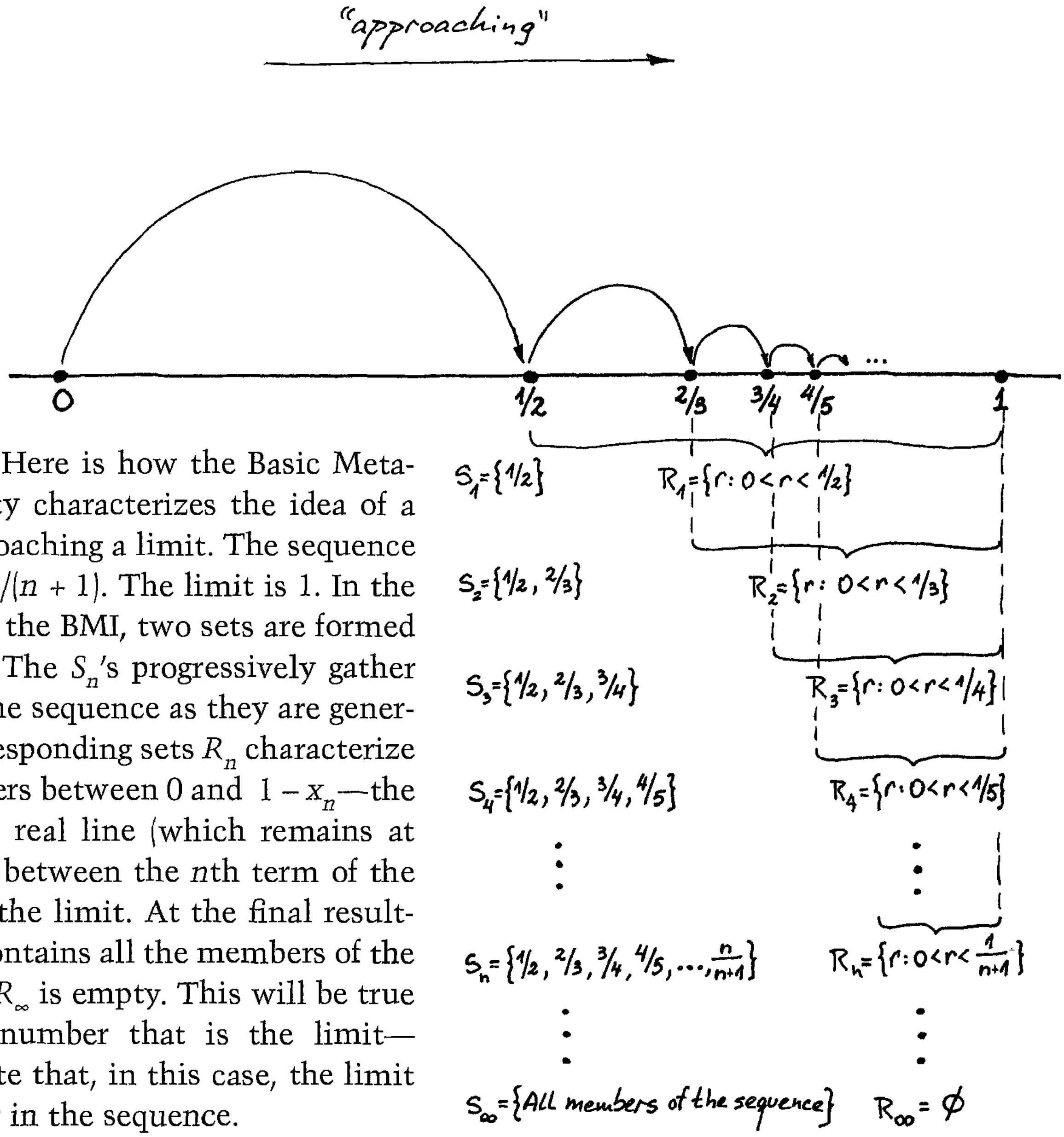


FIGURE 9.1 Here is how the Basic Metaphor of Infinity characterizes the idea of a sequence approaching a limit. The sequence here is $\{x_n\} = n/(n+1)$. The limit is 1. In the special case of the BMI, two sets are formed at each stage. The S_n's progressively gather the terms of the sequence as they are generated. The corresponding sets R_n characterize the real numbers between 0 and $1 - x_n$—the portion of the real line (which remains at the nth stage) between the nth term of the sequence and the limit. At the final resultant state, S_∞ contains all the members of the sequence and R_∞ is empty. This will be true only for the number that is the limit—namely, 1. Note that, in this case, the limit does not occur in the sequence.

- This is what is implicitly meant when we say that the infinite sequence x_n "approaches L as a limit." Note, incidentally, that L can be entirely outside the sequence and still have terms of the sequence infinitely close to it.

Here is what this looks like as a special case of the BMI. First we characterize the "Sequence and Limit frame," and then we use it in the special case of the BMI (see Figure 9.1). Note that n is used to name the stages of the BMI and also to name the indexes of the terms of the sequence $\{x_n\}$.

THE SEQUENCE AND LIMIT FRAME (PROTOTYPICAL VERSION)

A statement defining a sequence $\{x_n\}$.
A set S_n containing the first n terms of $\{x_n\}$.
A finite number L.
A set R_n of real numbers r such that $0 < r < |x_n - L|$.

THE BMI FOR INFINITE SEQUENCES (PROTOTYPICAL VERSION)

Target Domain ITERATIVE PROCESSES THAT GO ON AND ON		*Special Case* INFINITE SEQUENCES WITH A LIMIT L
The beginning state (0)	$\Rightarrow$	The Sequence and Limit frame
State (1) resulting from the initial stage of the process	$\Rightarrow$	S_1 = the set containing the first term of the sequence
The process: From a prior intermediate state ($n-1$), produce the next state (n).	$\Rightarrow$	From S_{n-1} containing the first $n-1$ terms of the sequence, form S_n containing the first n terms of the sequence.
The intermediate result after that iteration of the process	$\Rightarrow$	The set S_n. The set R_n containing all positive real numbers r such that $0 < r < \lvert x_n - L \rvert$. $R_n \subset R_{n-1}$.
"The final resultant state" (actual infinity "∞")	$\Rightarrow$	**The set S_∞ containing all the terms of the sequence** **There is no positive real number r such that $0 < r < \lvert x_i - L \rvert$ for all x_i in S_∞. Hence, $R_\infty = \varnothing$.** ***L* is the limit of the sequence.**
Entailment *E*: The final resultant state ("∞") is unique and follows every nonfinal state.	$\Rightarrow$	**Entailment *E*: *L* is the unique limit of the sequence.**

To clarify this further, here are all the stages and the sets S_n at each stage.

Stage $n = 1$	**Stage $n = 2$**	**Stage $n = 3$**	**Stage n**	**Stage $n = \infty$**
S_1	S_2	S_3	S_n	S_∞
The set consisting of the first term of the sequence $\{x_n\}$:	The set consisting of the first two terms of the sequence $\{x_n\}$:	The set consisting of the first three terms of the sequence $\{x_n\}$:	The set consisting of the first n terms of the sequence $\{x_n\}$:	The set consisting of all the terms of the sequence $\{x_n\}$:
$\{x_1\}$	$\{x_1, x_2\}$	$\{x_1, x_2, x_3\}$	$\{x_1, x_2, \ldots, x_n\}$	$\{x_1, x_2, \ldots\}$

A similar table could be built for the R_n's. The bottom row of such a table would show sets that exclude more and more elements. The set in the lower right cell of such a table would be the empty set.

Let us look again at our example of how this works.

- Consider the sequence $\{x_n\} = \frac{n}{n+1}$.
- The terms are: $x_1 = 1/2$, $x_2 = 2/3$, $x_3 = 3/4$, $x_4 = 4/5$, ...
- The set $S_1 = \{1/2\}$; $S_2 = \{1/2, 2/3\}$; $S_3 = \{1/2, 2/3, 3/4\}$; $S_4 = \{1/2, 2/3, 3/4, 4/5\}$; . . .
- The limit $L = 1$.
- The set $R_1 = \{r: 0 < r < 1/2\}$. $R_2 = \{r: 0 < r < 1/3\}$. $R_3 = \{r: 0 < r < 1/4\}$. $R_4 = \{r: 0 < r < 1/5\}$.

As n gets larger, the sets S_n come to have progressively more terms that get closer and closer to 1. The sets R_n come to exclude more and more positive real numbers. At stage $n = \infty$, S_∞ contains all of the infinite number of terms of the sequence $\{x_n\} = \frac{n}{n+1}$. It does not contain the number 1. At stage $n = \infty$, there is no positive real number r such that $r < |\frac{n}{n+1} - 1|$ for every finite n. Thus, the set R_∞ is empty. In this special case of the BMI, this is what it means for the infinite sequence $\frac{n}{n+1}$ to "approach" 1 as a limit as n "approaches ∞."

Notice that there are no epsilons and no quantifier statement "For every number epsilon there is an integer n_0." The irrelevant cases are not here at all. If we put this special case of the BMI together with the Numbers Are Points on a Line metaphor and Talmy's fictive-motion schema (see Chapter 2), we get the metaphor of "approaching" a limit as "n approaches infinity."

The values of the sequence are metaphorically conceptualized as locations along a line. Visualizing the process via these metaphors, there are two coordinated trajectors in motion: As the first moves from integer to integer starting with 1, the second moves correspondingly from point-location to point-location on the number line, starting with 1/2. As the first trajector moves from 1 to 2, the second moves from 1/2 to 2/3. Via the BMI, ∞ is the endpoint of the line, the final point-location that the first trajector can metaphorically "approach" infinitely close to. When the first trajector "reaches infinity," the second trajector "approaches the limit"; that is, it gets to a sequence of point-locations infinitely close to the limit, so close that there is no positive real number that can measure any distance between such point-locations and the limit.

This is how the prototypical idea of "approaching a limit as n approaches infinity" can be precisely conceptualized via the BMI and other conceptual metaphors.

The General Notion of a Limit Using the BMI

So far we have analyzed the prototypical version of a sequence that approaches a limit—namely, the case in which a sequence converges directly toward the limit. But many sequences converge indirectly—winding around and going back and forth as they ultimately converge to a limit. Because such cases exist, we will have to construct a fully general version of the BMI for convergent sequences.

To get an idea of the problem, let us consider what we will call "teaser sequences." Here is an example:

$$3/6,\ 4/6,\ 5/6,\ 9/12,\ 10/12,\ 11/12,\ 15/18,\ 16/18,\ 17/18,\ 21/24,\ \ldots$$

The sequence grows from 3/6, to 5/6, then drops back to 9/12, which is between 4/6 and 5/6 in size. Then it grows again to 11/12, which is more than 5/6. Then it drops back again—not all the way to 9/12, but only to 15/18, which is equivalent to 5/6 in size. Then it grows to 17/18, which is more than 11/12. Then it drops back to 21/24, which is between 15/18 and 16/18 in size. Then it grows again and repeats the pattern (see Figure 9.2).

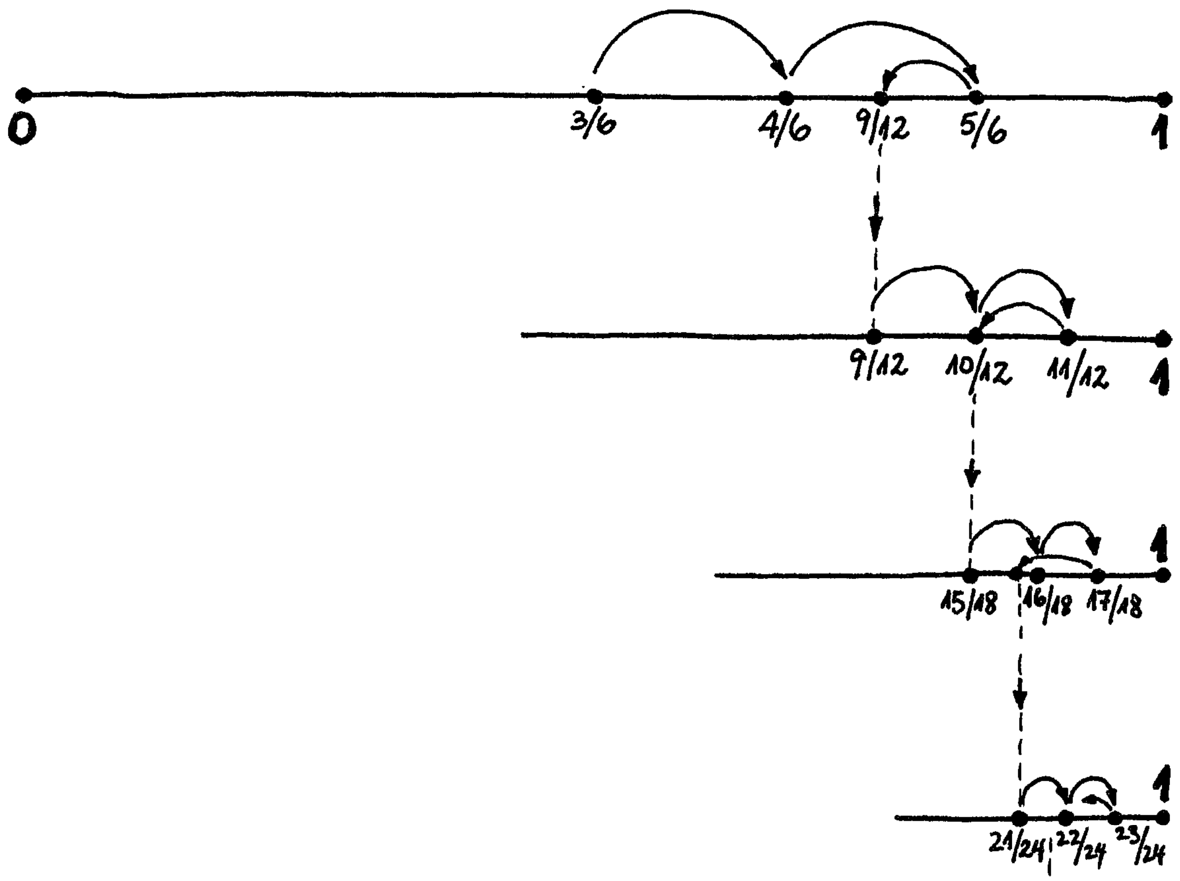
From the figure we can see that the *teaser elements*—those that drop back in size—are 9/12, 15/18, 21/24, ... The endpoints of each subsequence are 5/6, 11/12, 17/18, 23/24, ... The teaser sequence has two important properties:

1. The teaser elements form a directly convergent sequence.
2. Each teaser element is smaller than all the elements of the sequence that follow it.

From these properties it follows that the sequence as a whole converges.

How can we reformulate the special case of the BMI for limits of sequences to accommodate all such cases of indirect convergence? To answer this question, we have to see where the limitations arise in our previous formulation.

In the case of directly convergent sequences, the n in the definition of the sequence $\{x_n\}$ corresponds to the n characterizing the stages of the BMI. This does not work for teaser sequences. To generalize the notion of the limit, we need to distinguish between the positive integers characterizing the stages of the BMI and the positive integers indexing the elements of the sequence. We will continue to use the variable n for the integers characterizing the stages of the BMI, but we shall use the variable m for the positive integers characterizing the teaser sequence, which we will now call $\{x_m\}$. The following table illustrates the



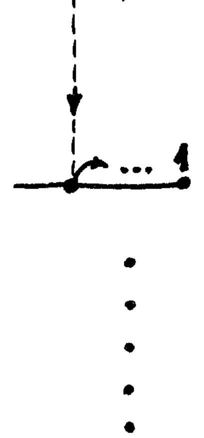


FIGURE 9.2 The diagram shows a "teaser sequence"—a sequence that converges indirectly towards a limit. It moves forward from 3/6 to 4/6 to 5/6, then back to 9/12, which is between 4/6 and 5/6. Then it moves forward to 10/12 and to 11/12. Then it moves back again to 15/18 (which is equal to 10/12), then forward to 16/18 and to 17/18 (more than 11/12). Then back to 21/24 (more than 15/18, but less than 16/18) and forward to 22/24 and to 23/24 (more than 17/18). And so on. The sequence, endlessly moving back and forth, ultimately converges to the limit 1.

kind of relationship between n and m that would be needed to handle the convergence of teaser sequences.

CONVERGENT SEQUENCE OF TEASER ELEMENTS

BMI stage n	Sequence term m	Value of the nth term of $\{x_t\}$, the sequence of teaser elements of $\{x_m\}$
1	1	3/6
2	4	9/12
3	7	15/18
4	10	21/24
5	13	27/30
...	...	...
n	$3n-2$	$3(2n-1)/6n$
...	...	...

We can characterize the relationship between the stage n of the BMI with the term m of the sequence $\{x_m\}$ with a statement. In this case, the statement is $m = 3n - 2$. Thus the second stage of the BMI maps onto the fourth term of the sequence, 9/12, which is the second teaser term. The third stage of the BMI maps onto the seventh term of the sequence, 15/18, which is the third teaser term; and so on. This generates a new sequence, the *sequence of teaser terms*, which we denote as $\{x_t\}$, for teasers.

Such statements allow us to adapt the earlier version of the BMI to characterize the direct convergence of this sequence of teaser terms $\{x_t\}$. Since the teaser elements x_t converge directly and since each is less than all the following elements in the original sequence $\{x_m\}$, the convergence of the teaser elements guarantees the convergence of the sequence $\{x_m\}$ as a whole. What we need to do now is characterize convergence for $\{x_t\}$, the sequence of teaser elements. To do so, we need to look at the differences between the limit and each teaser element.

Given that the limit L is equal to 1 in this case, we can form the sequence of differences between the limit and each of the teaser terms: $1 - 3/6 = 3/6$, $1 - 9/12 = 3/12$, $1 - 15/18 = 3/18$, and so on. This yields the following sequence of differences: 3/6, 3/12, 3/18, ... We will refer to the tth difference as ε_t. Thus, $\varepsilon_1 = 3/6$, $\varepsilon_2 = 3/12$, and so on, as we can see in the following table.

BMI stage n	$1 - x_t$	ε_t
1	3/6	ε_1
2	3/12	ε_2
3	3/18	ε_3
4	3/24	ε_4
5	3/30	ε_5
. . .	. . .	. . .
n	$3/6n$	ε_n
. . .	. . .	. . .

Thus, corresponding to each finite stage n of the BMI, there is an ε_t. As the BMI proceeds, a decreasing sequence of epsilons is produced, which characterizes the "approach of the sequence to the limit." As n approaches infinity, the epsilons approach zero, and the sequence approaches the limit.

When $n = m$, we have the prototypical version of approaching limits described earlier. When $n \leq m$, the BMI will map (via the statement $m = 3n - 2$) the nth stage onto the teaser elements of $\{x_m\}$. At the final stage where $n = \infty$, all the teaser elements have been chosen for every $n < \infty$. In other words, the entire infinite sequence $\{x_t\}$ of teaser terms has been chosen from the sequence $\{x_m\}$ as a whole.

To state this special case of the BMI, we need to revise the Sequence and Limit frame as follows. In describing this frame, we use the phrase "critical elements" for those terms of the sequence that must converge in order for the sequence as a whole to converge. Teaser elements are thus special cases of critical elements.

THE SEQUENCE AND LIMIT FRAME (GENERAL VERSION)

A statement defining a sequence $\{x_m\}$, where m is a positive integer.
A statement defining a sequence $\{x_t\}$ of *critical elements*, of $\{x_m\}$.
A set S_n containing the first n elements of $\{x_t\}$.
A finite number L.
A set R_n of real numbers r such that $0 < r < |L - x_n|$, where x_n is the nth term of $\{x_t\}$.
A constraint: Let x_n be the nth term of $\{x_t\}$; Let x_n' be the critical element in $\{x_m\}$ corresponding to the x_n; and let x_i be a term in $\{x_m\}$ with $i > n'$. $|L - x_n'| > |L - x_i|$.

The constraint in the frame sets a bound for each teaser element: no subsequent term of $\{x_m\}$ can be further from the limit L than a given critical element.

The special case of the BMI about to be stated will characterize two infinite sequences of sets, the S_n's and the R_n's. As n increases, the S_n's will come to contain more and more of the critical elements of the sequence and the R_n's will come to contain fewer and fewer real numbers between the nth term of $\{x_t\}$ and the limit. At ∞, S_∞ will contain all the elements of the sequence $\{x_t\}$ and R_∞ will be empty, which means that there will be no fixed real number that characterizes the difference between the limit and the terms of the sequence $\{x_t\}$, as n approaches ∞. The constraint in the frame guarantees that the sequence $\{x_m\}$ will converge to the limit L as n approaches ∞. Here is that special case of the BMI.

THE BMI FOR INFINITE SEQUENCES (GENERAL VERSION)

Target Domain ITERATIVE PROCESSES THAT GO ON AND ON		*Special Case* INFINITE SEQUENCES WITH A LIMIT L
The beginning state (0)	$\Rightarrow$	The Sequence and Limit frame (general version).
State (1) resulting from the initial stage of the process	$\Rightarrow$	S_1 = the set containing the first term of the sequence $\{x_t\}$.

The process: From a prior intermediate state $(n-1)$, produce the next state (n).	$\Rightarrow$	From S_{n-1} containing the first $n-1$ terms of the sequence $\{x_t\}$, form S_n containing the first n terms of the sequence $\{x_t\}$.
The intermediate result after that iteration of the process	$\Rightarrow$	The set S_n. The set R_n containing all positive real numbers r such that $0 < r < \|x_n - L\|$. $R_n \subset R_{n-1}$.
"The final resultant state" (actual infinity "∞")	$\Rightarrow$	**The set S_∞ contains all the terms of the sequence $\{x_t\}$. There is no positive real number r such that $0 < r < \|x_n - L\|$ for all x_n in S_∞. Hence, $R_\infty = \varnothing$. L is the limit of the sequence $\{x_t\}$ of critical elements of $\{x_m\}$. L is also the limit of the entire sequence $\{x_m\}$.**
Entailment *E*: The final resultant state ("∞") is unique and follows every nonfinal state.	$\Rightarrow$	**Entailment *E*: L is the unique limit of the sequences $\{x_t\}$ and $\{x_m\}$.**

To clarify this further, here are all the stages and the sets S_n at each stage of the BMI for the sequence of teaser elements given earlier.

BMI stage $n = 1$	**BMI stage $n = 2$**	**BMI stage $n = 3$**	**BMI stage n**	**BMI stage $n = \infty$**
S_1	S_2	S_3	S_n	S_∞
The set consisting of the first term of the sequence $\{x_t\}$:	The set consisting of the first two terms of the sequence $\{x_t\}$:	The set consisting of the first three terms of the sequence $\{x_t\}$:	The set consisting of the first n terms of the sequence $\{x_t\}$:	The set consisting of all the terms of the sequence $\{x_t\}$:
$\{3/6\}$	$\{3/6, 9/12\}$	$\{3/6, 9/12, 15/18\}$	$\{3/6, \ldots, 3(2n-1)/6n\}$	$\{3/6, 9/12, \ldots\}$

Here we see the gradual formation of the sequence of teaser terms. The bottom right cell has the set consisting of *all* the terms of the sequence $\{x_t\}$. A similar table could be built for the R_n's. The bottom row of such a table would show sets that exclude more and more elements. The set in the lower right cell of such a table would be the empty set.

We have now shown how the BMI can characterize the general case of the convergence of a sequence to a limit.

Infinite Sums

Before we proceed, there is an additional metaphor that should be discussed. Consider the sequence of partial sums given by $\sum_{k=1}^{n} \frac{9}{10^k}$. The first three terms are 0.9, 0.99, and 0.999.

When k grows indefinitely, the result is an infinite sum $\sum_{k=1}^{\infty} \frac{9}{10^k}$. Mathematicians take this infinite sum as being *equal* to 1. This requires an extra metaphor:

INFINITE SUMS ARE LIMITS OF INFINITE SEQUENCES OF PARTIAL SUMS

Source Domain LIMITS OF INFINITE SEQUENCES		*Target Domain* INFINITE SUMS
The limit of an infinite sequence of partial sums	→	An infinite sum
$\lim_{n\to\infty} \sum_{k=1}^{n} a_k$	→	$\sum_{k=1}^{\infty} a_k$

Via this metaphor, one can "define" an infinite sum as being the limit of an infinite sequence of partial sums. This metaphor piggybacks on the version of the BMI defining the limit.

We can see how this works in the classic example of the sum:

$$0.9 + 0.09 + 0.009 + 0.0009 + \ldots.$$

In the special case of the BMI for infinite sequences, we plug in the following values:

- $s_n = \sum_{k=1}^{n} \frac{9}{10^k}$
- $L = 1$
- $s_1 = 0.9$; $s_2 = 0.99$; $s_3 = 0.999$; . . .
- $\varepsilon_1 = |1 - 0.9| = 0.1$; $\varepsilon_2 = |1 - 0.99| = 0.01$; $\varepsilon_3 = |1 - 0.999| = 0.001$; . . .
- s_∞ is the infinite sum, 0.99999. . .
- Since there is no positive real number r that is less than $|L - s_\infty|$ for every n, it follows that $|L - s_\infty|$, which equals $|1 - 0.9999\ldots| = 0$. Thus, $1 = 0.999\ldots.$

In other words, the sequence of partial sums—0.9, 0.99, 0.999, . . . —has 1 as its limit. Via the infinite sum metaphor, An Infinite Sum Is the Limit of An Infinite Sequence of Partial Sums, the infinite sum 0.9999. . . equals 1. This is an entailment of two metaphors: the special case of the BMI for limits and the infinite sum metaphor.

Limits of Functions

One of the most basic ideas in calculus is that of the limit of a function. We can now straightforwardly extend the account we have just given of limits of infinite sequences using the BMI to the notion of the limit L of a function $f(x)$ as x "approaches" a real value a.

As we saw in the previous section, the notion of a function "approaching" a value can be characterized in terms of limits of sequences plus the metaphors:

- Numbers Are Points on a Line,
- Talmy's fictive-motion schema, and
- the Change of a Function Is the Coordinated Motion of Two Trajectors, one in the domain and one in the range of the function.

Given these conceptual metaphors, we can now extend the BMI account of the limit of a sequence to the limit of a function $f(x)$ as x approaches a.

LIMIT OF A FUNCTION $f(x)$ AS x APPROACHES a
(USING THE BMI DEFINED FOR SEQUENCES)

Let $f(x)$ be a function.
Suppose that for every infinite sequence

$$\{r_i\} = r_1, \ldots, r_i, \ldots$$

such that $\{r_i\}$ converges to a,
there is a corresponding infinite sequence

$$\{f(r_i)\} = (f(r_1), \ldots, f(r_i), \ldots)$$

that converges to L.
We define L to be the limit of $f(x)$ as x approaches a.
Moreover, if $f(x)$ is defined at a, then $f(a) = L$ (which assumes the continuity implicit in the above fictive-motion schema).

Applying the metaphors mentioned above, this will yield the concept of the limit L of a function $f(x)$ as x approaches a.

It is straightforward to show that whenever this condition holds, the traditional epsilon-delta condition holds:

LIMIT OF A FUNCTION (USING EPSILONS AND DELTAS)

$\lim_{x \to a} f(x) = L$ iff
for all $\varepsilon > 0$, there is a $\delta > 0$,
such that if $0 < |a - x| < \delta$,
then $|L - f(x)| < \varepsilon$.

If we let each choice of epsilon be some $f(r_i)$ in a sequence $\{f(r_i)\}$, then r_i will be a suitable delta, and the condition will be met.

The point of this is not to eliminate the epsilon-delta condition from mathematics but, rather, to comprehend how we understand it. We understand it first in geometric terms using the notion of "approaching a limit." And we understand "approaching a limit" in arithmetic terms via the BMI, as described in the previous section. Given that, we can understand the idea of "approaching a limit" for a function in arithmetic terms using the characterization above, which presupposes the BMI. In that characterization, each sequence $\{f(r_i)\}$ metaphorically defines a sequence of "steps" in some "movement" toward L, while each corresponding sequence $\{r_i\}$ defines a sequence of "steps" in some "movement" toward a.

So far, so good. We have depended upon the use of the BMI in the previous section, which characterized limits of sequences. This works fine for cases where $f(x)$ is defined at a. Then it is clear that $f(a) = L$. But what if $f(x)$ is not defined at a? Does our characterization of the limit still hold? And if it does, can we meaningfully add the value $f(a) = L$ without contradiction?

Let us take an example.

$$f(x) = \frac{x^2 - 1}{x - 1}$$

This function is not defined for $x = 1$, since that would make the denominator equal to zero. For values other than $x = 1$, the function acts the same as $f(x) = x + 1$, since the numerator $x^2 - 1 = (x + 1)(x - 1)$. Now consider the values $x = 0.9$, 0.99, 0.999, and so on. The corresponding values of $f(x)$ are 1.9, 1.99, 1.999, and so on. On our characterization of the limit of $f(x)$ as x approaches 1, there would be a limit—namely, 2. Even though strictly speaking it is not the case that $f(1) = 2$!

There are several morals to this discussion:

- The BMI as used for limits of sequences can provide a mechanism for characterizing limits of functions in arithmetic terms.
- With the appropriate additional metaphors, the analysis given also characterizes in precise terms the intuitive notion of "approaching a limit." It is thus false that such a mathematical idea cannot be made precise in intuitive terms. Indeed, the fact that conceptual metaphors can be precisely formulated allows us to make such intuitive ideas precise.
- The epsilon-delta definition of a limit does not characterize the mathematical idea of approaching a limit, since it does not involve any "approach" to anything. It is just a different idea that happens to cover the same cases, as well as countless irrelevant but mathematically harmless cases.

We can now see why students have problems learning the epsilon-delta definition of a limit. They are taught an intuitive idea of what a "limit" is in terms of a motion metaphor, and then told, incorrectly, that the epsilon-delta condition expresses the same idea (see Núñez, Edwards, & Matos, 1999).

Least Upper Bounds

We have seen an example of how the BMI can characterize real numbers—both the symbols for real numbers (infinite decimals) and the real numbers themselves (infinite polynomials). However, the common axiomatization of the real numbers does not use this characterization directly. The real numbers are taken to be whatever objects fit the axioms. There are ten axioms. The first six characterize fields, and the next three add ordering constraints, so that the first nine axioms characterize ordered fields. And the tenth, the *Least Upper Bound axiom*, characterizes complete ordered fields. The real numbers are the only complete ordered field. Here are the axioms.

1. Commutative laws for addition and multiplication.
2. Associative laws for addition and multiplication.
3. The distributive law.
4. The existence of identity elements for both addition and multiplication.
5. The existence of additive inverses (i.e., negatives).
6. The existence of multiplicative inverses (i.e., reciprocals).
7. Total ordering.
8. If x and y are positive, so is $x + y$.
9. If x and y are positive, so is $x \cdot y$.

10. The Least Upper Bound axiom: Every nonempty set that has an upper bound has a least upper bound.

Upper bounds and least upper bounds are defined in the following way:

Upper Bound
b is an *upper bound* for S if
$x \leq b$, for every x in S.

Least Upper Bound
b_0 is a *least upper bound* for S if
- b_0 is an upper bound for S, and
- $b_0 \leq b$ for every upper bound b of S.

Axioms 1 through 9 are not sufficient to distinguish the reals from the rationals, since the rational numbers also fit them. Add axiom 10 and you get the reals.

The question to be asked is, Why? What do infinite decimals and infinite polynomial sums have to do with least upper bounds? What does the BMI have to do with least upper bounds? To make the question clear, let us take an example. Consider $\pi = 3.14159265358979323846264 3 \ldots$. The irrational number π has an infinite number of decimal places. We cannot write all of them down. We can write down only approximations to some number of decimal places. But each such approximation is a rational number, not an irrational number. For any approximation of π to n decimal places, we can write down a smallest upper bound for that approximation. The following table gives a sequence of rational approximations to π to a given number of decimal places, the corresponding smallest upper bound to that number of decimal places, and the difference between the two.

π to n Decimal Places	Least Upper Bounds of π to n Decimal Places	The Difference to n Decimal Places	The Number of Decimal Places
3.1	3.2	0.1	1
3.14	3.15	0.01	2
3.141	3.142	0.001	3
3.1415	3.1416	0.0001	4
3.14159	3.14160	0.00001	5
3.141592	3.141593	0.000001	6
3.1415926	3.1415927	0.0000001	7
3.14159265	3.14159266	0.00000001	8
3.141592653	3.141592654	0.000000001	9

Each of the first three columns provides the first nine terms of an infinite sequence. That sequence can be thought of from the point of view of potential infinity as ongoing, or from the perspective of actual infinity as *being* infinite. In the case of actual infinity, we can conceptualize these infinite sequences via the BMI. The first two sequences—the approximation of π to n places and its least upper bound to n places—become identical in more and more places as n increases. Moreover, the differences between them, as column three shows, become smaller and smaller as n increases. If we apply the BMI to the sequence of differences, then "at" $n = \infty$, the value of the difference "is" zero. That means that, "at" $n = \infty$, the value of the real number π is identical to the least upper bound of all rational approximations to π. π is also its *own* least upper bound!

This is extremely important in understanding the nature of a least upper bound. Recall the definition:

Upper Bound
b is an *upper bound* for S if
$x \leq b$, for every x in S.

Least Upper Bound
b_0 is a *least upper bound* for S if
- b_0 is an upper bound for S, and
- $b_0 \leq b$ for every upper bound b of S.

Each of the finite terms in the second column—a least upper bound to n places—is an upper bound for π. This is true for every finite term in the sequence defined by the second column. Thus, there is an infinite number of upper bounds for π descending sequentially in value; that is, the number of finite upper bounds getting smaller and smaller is endless. From the perspective of potential infinity, there *is no least upper bound!* It is only from the perspective of actual infinity, characterized via the BMI, that a least upper bound comes into existence. In other words, that least upper bound is "created," from a cognitive point of view, by the Basic Metaphor of Infinity. Moreover, by virtue of the BMI, that least upper bound is unique!

Here we can see clearly that the existence of least upper bounds is not guaranteed by the first nine axioms, which characterize the rational numbers. That is why axiomatically oriented mathematicians have added the Least Upper Bound axiom. It was brought into mathematics for the purpose of "creating" the real numbers from the rationals. It doesn't follow from anything else, and *it is not a special case of anything else!*

What is the difference between using the BMI to characterize real numbers and using the Least Upper Bound axiom? The question is a bit strange, something like comparing apples and oranges. The BMI is an unconscious cognitive mechanism that makes use of ordinary elements of human cognition: aspect, conceptual metaphor, and so on. The BMI has arisen spontaneously, outside mathematics, and has been applied within mathematics to characterize cases of actual infinity. The Least Upper Bound axiom is a product of formal mathematics, consciously and intentionally constructed to characterize the real numbers, given the first nine axioms. It is also used in other branches of mathematics, especially in set theory.

But the Least Upper Bound axiom is anything but intuitively clear on its own terms. Indeed, to understand it, most students of mathematics have to conceptualize it in terms of the BMI. From a cognitive perspective, the BMI is conceptually more basic.

Here is a characterization of the concept of a least upper bound from the perspective of the Basic Metaphor of Infinity.

THE BMI VERSION OF LEAST UPPER BOUNDS

Target Domain ITERATIVE PROCESSES THAT GO ON AND ON		*Special Case* LEAST UPPER BOUNDS
The beginning state (0)	⇒	S is a set of real numbers. A real number b is an upper bound for S if $b \geq$ every member of S. b_0 is a least upper bound of S if it is smaller than every upper bound b. Let B be an infinite set of upper bounds b_n for S.
State (1) resulting from the initial stage of the process.	⇒	Choose an upper bound b_1.
The process: From a prior intermediate state (n–1), produce the next state (n).	⇒	Given an upper bound b_{n-1}, choose an upper bound b_n arbitrarily less than b_{n-1}.
The intermediate result after that iteration of the process (the relation between n and n–1)	⇒	A finite sequence $\{b_i\}$ of upper bounds, with $b_n < b_{n-1}$
"The final resultant state" (actual infinity "∞")	⇒	**A least upper bound, b_∞**
Entailment *E*: The final resultant state ("∞") is unique and follows every nonfinal state.	⇒	**Entailment *E*: The least upper bound, b_∞, is unique and less than all other upper bounds.**

Of course, we can correspondingly characterize greatest lower bounds in a symmetrical fashion, substituting lower bounds for upper bounds and interchanging "<" and ">."

Is 0.9999. . . = 1.000. . . ?

Consider 0.9999999. . . . Its least upper bound is 1.0000000. . . . Within the real number system, these two infinite decimals will symbolize the same number (their corresponding polynomials would give the same infinite sum). The reasoning is the following: Within the real numbers, two numbers are identical if they are not distinct. Two numbers are distinct if there is a nonzero difference between them; that is, x and y are distinct if and only if there is a positive number $d > 0$, such that $|x - y| = d$.

In the case of 0.999999. . . and 1.000000. . . , there can be no such d. No matter how small a number d you pick, you can always take 0.999999. . . and 1.00000. . . out to enough decimal places so that the difference between them is less that than d.

We can see this clearly using the BMI. Suppose again we consider the least upper bounds to n places.

0.9. . . 9 to n places	1.0. . . 0 to n places	The difference to n places
0.9	1.0	0.1
0.99	1.00	0.01
0.999	1.000	0.001
0.9999	1.0000	0.0001
0.99999	1.00000	0.00001
0.999999	1.000000	0.000001

As in the case of π, the sequence of differences gets smaller and smaller. Applying the BMI to this sequence, the difference is zero "at infinity." Therefore, "at" metaphorical infinity, 0.9999. . . = 1.0000. . . . That is, there are no real numbers between the infinite decimals 0.9999. . . and 1.0000. . . . They are, then, just different symbols for the same number.

As we shall see in Chapter 11, this result does *not* hold for every number system. There are number systems beyond the reals (namely, the hyperreals) in which 0.9999. . . and 1.0000. . . are *distinct* numbers—not just distinct numerals but actually different numbers, with a hyperreal number characterizing the difference between 0.9999. . . and 1.000. . . . But we will get to that later. In the system of real numbers, 0.9999. . . does equal 1.000. . . .

At this point, we can see exactly why the Least Upper Bound axiom, when added to the first nine, characterizes the real numbers and not the rationals. The sequence of finite rational approximations of π to n places consists of rational numbers alone. π, the infinite sequence, is not in the set of finite rational sequences. As an irrational number, it stands outside the set of rational numbers altogether. But π is the least upper bound of that sequence. Thus, insisting that an ordered field (like the rationals) contain all its least upper bounds adds the irrational numbers—the numbers written with *infinite decimals*—to the rationals.

One of the beauties of the BMI is that it allows us to see clearly just *why* adding the Least Upper Bound axiom to the axioms for ordered fields takes us beyond the rationals to the reals.

Infinite Intersections of Nested Intervals

We have just seen how least upper bounds take us from an axiom system that can fit the rationals without the reals to one that necessarily includes the reals. That is, if we add least upper bounds to all convergent sequences of rationals, we get the reals. The BMI, as we have just seen, can be used to show why this is so.

We are now in a position to use the BMI to shed light on another way in which mathematicians commonly characterize the real numbers—as infinite intersections of nested closed intervals with rational endpoints on the number line. In the process, we can show why the infinite-intersection characterization of the reals is equivalent to the least-upper-bound characterization.

Let us once more take as an example $\pi = 3.14159265358979323846264 3 \ldots$. Let us take the number line containing the rational numbers. Consider the following set of nested intervals, as given in the following table. The first interval is [3.1, 3.2]. The second is [3.14, 3.15]. The third is [3.141, 3.142]. And so on.

Left-Hand Side of the Closed Interval	Right-Hand Side of the Closed Interval
3.1	3.2
3.14	3.15
3.141	3.142
3.1415	3.1416
3.14159	3.14160
3.141592	3.141593
3.1415926	3.1415927
3.14159265	3.14159266
3.141592653	3.141592654

This is exactly the same table that we used in the last section, but with approximations to n places conceptualized as left ends of the intervals, while upper bounds to n places are conceptualized as right ends of the intervals. We are now in a position to characterize infinite sequences of nested intervals as special cases of the BMI.

NESTED INTERVALS

Target Domain ITERATIVE PROCESSES THAT GO ON AND ON		*Special Case* INFINITE SEQUENCES OF NESTED INTERVALS
The beginning state (0)	$\Rightarrow$	An unending sequence of nested closed intervals, each defined by the pair of rational numbers $[l_n, u_n]$, where $l_n > l_{n-1}$ and $u_n < u_{n-1}$
State (1) resulting from the initial stage of the process	$\Rightarrow$	The first term of the sequence: the interval $[l_1, u_1]$
The process: From a prior intermediate state ($n-1$), produce the next state (n).	$\Rightarrow$	Form the intersection of $[l_{n-1}, u_{n-1}]$ and $[l_n, u_n]$.
The intermediate result after the n^{th} finite iteration of the process	$\Rightarrow$	The interval $[l_n, u_n]$
"The final resultant state" (actual infinity "∞")	$\Rightarrow$	**The "interval" $[l_\infty, u_\infty]$, where $l_\infty > l_n$ and $u_\infty < u_n$ for every finite n. The distance between l_∞ and u_∞ is zero. That is, the "interval" is a point, $p = l_\infty = u_\infty$.**
Entailment *E*: The final resultant state ("∞") is unique and follows every nonfinal state.	$\Rightarrow$	**Entailment *E*: p is a unique point and is greater than every l_n and less than every u_n.**

This is a general way of characterizing the reals, given the rationals: p need not be a rational number, depending on how the intervals are chosen at each of the infinitely many finite stages. Now we can see the equivalence of various ways of using the BMI to get the real numbers from the rationals: infinite decimals (which symbolize infinite polynomials), least upper bounds, and infinite intersections of nested intervals.

Conclusion

The concepts discussed in this section—infinite decimals, infinite polynomials, limits of infinite sequences (and of functions), infinite sums, least upper bounds, and infinite intersections of nested intervals—all involve the notion of actual infinity. We have argued that actual infinity is conceptualized metaphorically using the Basic Metaphor of Infinity. What we have shown in this chapter is how a single cognitive mechanism—the BMI—with different special cases, can characterize each of these concepts.

Since real numbers are characterized only via such concepts (and the Dedekind cut; see Chapter 13), we have effectively shown how the real numbers are conceptualized via the Basic Metaphor of Infinity. In short, the "real" numbers are, from a cognitive point of view, a metaphorical construct.

The various "definitions" of the real numbers are implicitly metaphorical definitions, making implicit conceptual use of the Basic Metaphor of Infinity. These metaphorical definitions are equivalent; that is, they have the same (or equivalent) entailments. Given the equivalence of their entailments, we can think of them as characterizing the same "entities" with the same properties; that is, the entities and the properties these different special cases define can be put in one-to-one correspondence. For this reason, many mathematicians speak of *the* real numbers, as if they were objectively existing entities in the universe. Because of their metaphorical equivalence, this creates no practical problems: If you happen to think of these metaphorical objects as if they exist objectively, you will not get into any mathematical difficulties.

It is by no means obvious that the real numbers are necessarily conceptualized only via one version or another of the Basic Metaphor of Infinity. It requires *mathematical idea analysis* to show this. One of the benefits of mathematical idea analysis is that it shows explicitly how the various characterizations of the real numbers are related to one another. It also makes explicit the distinction between the symbolization of real numbers (using infinite decimal representations) and the numbers themselves.

10

Transfinite Numbers

GEORG CANTOR IS PERHAPS MOST FAMOUS for having created a mathematical system for precisely characterizing numbers of infinite size with a precise arithmetic. In Cantor's system, the "transfinite" numbers are not merely infinite but have different degrees of infinity, with one infinite number being "larger" than another.

The concepts Same Size As and Larger Than for infinite numbers came out of Cantor's metaphor, which conceptualized two sets as having the "same size"—the same *number* of members—if they could be put in one-to-one correspondence (see Chapter 7). We will designate the size of a set determined via Cantor's metaphor as the *Cantor size,* or cardinality, of the set.

As we saw in Chapter 8, the BMI is used implicitly to conceptualize the infinite set of all natural numbers. Cantor, by his famous rational array diagram, proved that the set of rational numbers (fractions) has the same Cantor size as the set of natural numbers. He achieved this by constructing a sequence of the rationals, thus implicitly showing that the rationals could be put in one-to-one correspondence with the natural numbers. Figure 10.1 shows Cantor's display, indicating how the sequence can be constructed.

As we trace along the line in the diagram, we form a series of the rational numbers we encounter:

$$1, 2, 1/2, 1/3, 3, 4, 3/2, 2/3, 1/4, \ldots$$

Note that the diagram is not, and cannot be, complete, since the line in itself is intended to cover infinitely many steps. What is implicit in how the diagram is intended to be understood is the BMI. The BMI is implicitly and unconsciously used in the following ways in comprehending this diagram:

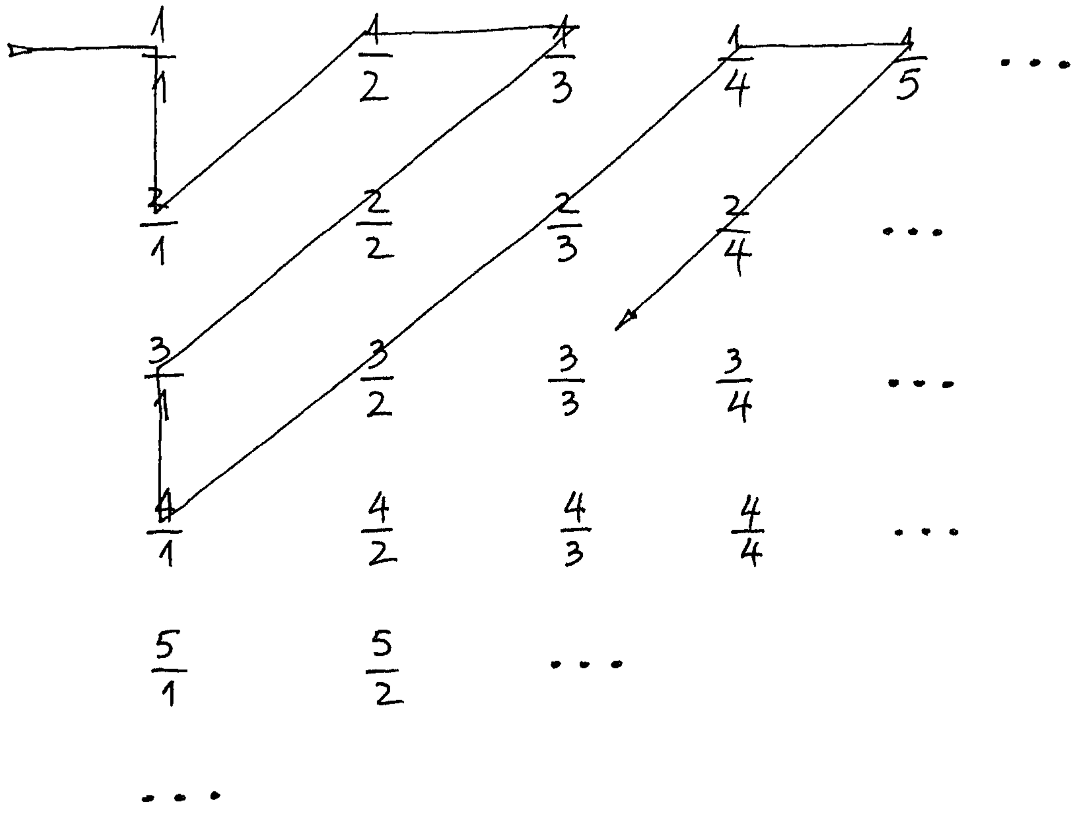


FIGURE 10.1 Cantor's one-to-one correspondence between the natural numbers and the rational numbers. This is an array in which all possible fractions (rational numbers) occur. All the fractions with 1 as numerator are in the first row, all those with 2 as numerator are in the second row, and so on. All those with 1 as denominator are in the first column, all those with 2 as denominator are in the second column, and so on. The zigzag line systematically goes through all the rational numbers. Cantor mapped the first rational number the line hits (1/1) to 1, the second (2/1) to 2, the third (1/2) to 3, and so on. This established a one-to-one correspondence between the natural numbers and the rationals.

- First, the special case of the BMI for the infinite set of natural numbers is used to guarantee that the first column of the matrix contains *all* the natural numbers—or equivalently, all the natural numbers as numerators with 1 as a denominator.
- Next, a similar version of the BMI is used to guarantee that the second column contains *all* the fractions with denominator 2.
- So far we have the beginning of a sequence of sequences: (the sequence of all fractions with denominator 1, the sequence of all the fractions with denominator 2, . . .). This sequence of sequences (extending downward) is also unending, and another use of the BMI is implicit to guarantee that this infinite sequence of sequences is complete. This gives us an understanding of the infinite array of rational numbers, *without any path through it.*

- Finally, there is the linear path that systematically covers every member of the array, creating a single sequence 1/1, 2/1, 1/2, 1/3, 2/2, 3/1, 4/1, 3/2, 2/3, 1/4, . . . which exhausts *all* the rational numbers. Here, too, a version of the BMI is used to guarantee that *all* rational numbers are included in the infinite sequence.
- Since a sequence, technically, is a mapping from the natural numbers to the terms of the sequence, the sequence itself constitutes a one-to-one correspondence between the natural numbers and the rational numbers. By Cantor's metaphor, the set of rationals is thereby assigned the same Cantor size as the set of natural numbers.

In short, the BMI is used over and over, implicitly and unconsciously, in comprehending this diagram. The diagram, via Cantor's metaphor, is taken as a *proof* that the natural numbers and the rational numbers have the same Cantor size—that is, the same cardinality.

Of course, according to our normal, everyday notion of size, there are more rational numbers than natural numbers, since all the natural numbers are contained in the rationals, and if they are taken away there are lots of rationals left over. Moreover, between any two natural numbers, there are an infinite number of rationals that are not natural numbers. In short, according to our normal non-Cantorian notions of number and size, there are zillions more rational numbers than there are natural numbers.

However, this chapter is not about our normal notion of number and size but about Cantor's. If one accepts Cantor's metaphor and the BMI, remarkable things follow.

Cantor's Diagonalization Proof

Cantor's most celebrated proof is his diagonalization argument, which demonstrated that the real numbers cannot be put in a one-to-one correspondence with the rational numbers. The proof works by contradiction. If the reals can be put in one-to-one correspondence with the rationals, then they can also be put into one-to-one correspondence with the natural numbers, which means they can be put in a list. Assume that each real number between zero and one is represented by an infinite decimal. Then there should be a list of the sort given in Figure 10.2, where line 1 would contain the first infinite decimal, line 2 the second infinite decimal, and so on.

Cantor showed that no matter what such a list looks like, there must be at least one real number that is not on the list. He drew a diagonal line that picks

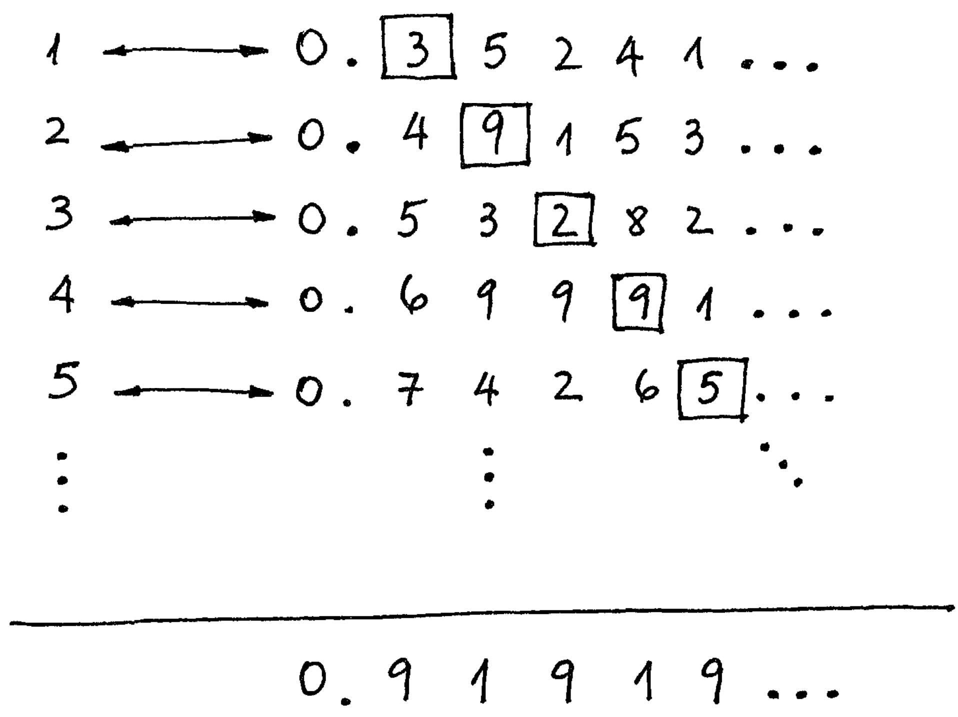


FIGURE 10.2 Cantor's diagonal argument shows that there can be no one-to-one correspondence between the natural numbers and the real numbers. It suffices to prove the result for the reals between 0 and 1, since if *they* cannot be put in one-to-one correspondence with the natural numbers, all the reals certainly cannot. Suppose there were such a correspondence, as given in the diagram, where the natural numbers are on the left and the corresponding real numbers on the right. Cantor showed how to construct a real number that, contrary to assumption, cannot be on the list. He constructed an infinite decimal as follows: For the nth number on the list, he replaced the digit in the nth decimal place by another digit; for instance, if that digit was 9, he replaced it by 1, and if it was not 9, he replaced it by 9. The resulting number differed from every infinite decimal on the list in at least one decimal place. Therefore, any assumed one-to-one correspondence between natural numbers and real numbers cannot exist.

out the first digit of the first number, the second digit of the second number, and so on. Then he constructed a new infinite decimal by using a rule for replacing the digits of the diagonal (e.g., where there is a 9 he puts a 1, and where there is a digit other than 9 he puts a 9). He replaced the first digit, then the second, and so on infinitely. The result is an infinite decimal that is not on the list, since it must differ from *every* infinite decimal on the list in at least one decimal place.

There is one minor technicality in the proof. Each number must be represented uniquely. Since the two numerals .499999. . . and .50000. . . represent the same number, one representation must be chosen. Cantor chose the first in this case and all similar cases.

As before, this proof of Cantor's uses an array. Each place in the array is indexed by two integers (n, m): the row of the array (the n^{th} infinite decimal) and the column of the array (the m^{th} digit). The diagonal line goes through all the digits where $n = m$.

There are many uses of the BMI implicit in Cantor's proof.

- First, there is the use of the special case of the BMI for infinite decimals. Each line is unending, yet complete.
- Then there is the use of the special case of the BMI for the set of all natural numbers. Each row corresponds to a natural number, and *all of them* must be there.
- Third, there is the sequence along the diagonal ($a_{11}, a_{22}, a_{33}, \ldots$). It, too, is assumed to include *all* the digits on the diagonal. This is another implicit use of the BMI.
- And finally, there is the process of replacing each digit on the diagonal with another digit. The process is unending, but must cover the *whole* diagonal. Another implicit special case of the BMI.

These implicit uses of the BMI in Cantor's proof are of great theoretical importance. The proof is usually taken to be a formal proof, and the result is taken to be objective—independent of any minds. But the proof has not been, and cannot be, written down formally using actual infinity, with an array that is *actually* infinitely long and wide. The actual infinities involved in the proof are all mental entities, conceptualized via the BMI. In other words, the proof is not literal; it is inherently metaphorical, using the Basic Metaphor of Infinity in four places. Since it uses metaphor, it makes use of a cognitive process that exists not in the external, objective world but only in minds. Thus, the result that there are more real numbers than rational numbers is an *inherently* metaphorical result, not a result that transcends human minds.

This, of course, does not make it a less valid mathematical result, any more than the fact that the real numbers require the BMI makes them any less "real."

Cantor's proof that the reals cannot be put into one-to-one correspondence with the natural numbers and, hence, with the rationals had a dramatic import. It showed that the set of reals has a larger Cantor size than the rationals, that there are "more" reals than rationals (if one accepts Cantor's metaphor). Since the set of natural numbers is infinite, this meant that the set of real numbers is "larger than" that infinity, that there are *degrees of infinity*, that some infinities are "larger than" others (in a mathematics accepting Cantor's metaphor). Cantor called the number of elements in the set of natural numbers "$\aleph_0$" and the number of elements in the set of real numbers "C" for "continuum" (see Chapter 12).

Cantor further proved a remarkable result about power sets, the set of all subsets of a given set. He proved that there can be no one-to-one correspondence between the members of a set and its power set—even for infinite sets! Recall (from Chapter 7) that he had previously shown that an infinite set (the natural numbers) could be put in one-to-one correspondence with one of its subsets (the even positive integers). The power-set proof showed that there was at least one general principle about the impossibility of one-to-one correspondences between infinite sets:

Number of Elements in a Set	Number of Elements in the Corresponding Power Set
1	2
2	4
3	8
.	.
.	.
.	.
.	.
.	.
n	2^n
.	.
.	.
.	.
.	.
.	.
$\aleph_0$	$2^{\aleph_0}$

From this table we can see that for finite sets with n members, the number of elements in a power set is 2^n. Cantor called the number of elements in the power set of the natural numbers "$2^{\aleph_0}$". Whereas 2^3 equals 2 times 2 times 2, $2^{\aleph_0}$ is not $2 \cdot 2 \cdot 2 \cdot \ldots \cdot 2$ an $\aleph_0$ number of times. Such a concept is not defined. Rather, the symbol $2^{\aleph_0}$ is a name for the number of elements in the power set of natural numbers—whatever that number is. What Cantor proved was that that number is "greater than" $\aleph_0$, assuming Cantor's metaphor. Of course, the symbol $2^{\aleph_0}$ was not chosen randomly. It was a choice based on the application of the BMI to *numerals*, not numbers. For each finite set, the numeral naming the number of elements in the power set of a set with n elements is 2^n. Applying the BMI to the numerals, we get a new symbol—a new numeral naming the number of elements in the power set of the natural numbers.

Cantor was able to put his two remarkable proofs together. He had proved, using his central metaphor, that

- The number of real numbers is "greater than" the number of natural numbers.
- The number of elements in a power set is "greater than" the number of elements in the set it is based on.

He was able to prove a further remarkable result:

- The number of elements in the set of real numbers is the "same" as the number of elements in the power set of the natural numbers. Or, as written in symbols: $C = 2^{\aleph_0}$.

Transfinite Arithmetic

Cantor's metaphor, together with the BMI, does something that the BMI alone does not do. It creates for infinite sets unique, precise cardinal numbers, which have a well-defined arithmetic. Recall that without Cantor's metaphor, the BMI alone can create the "number" ∞, the largest natural number. But ∞, not being one of the natural numbers, does not combine in a meaningful way with other natural numbers. As we saw, $0 \cdot \infty$, $\infty - \infty$, ∞/∞, and ∞/n are not well defined as operations on natural numbers.

Cantor's metaphor, together with the Basic Metaphor of Infinity, defines the transfinite cardinal number $\aleph_0$ differently from the way the BMI alone characterizes ∞ as the endpoint of the natural-number sequence. Transfinite cardinals were carefully characterized by Cantor so that they would have a meaningful arithmetic.

Consider $\aleph_0 + 1$. This is the cardinality of all sets in one-to-one correspondence with a set formed as follows: Take the set of natural numbers and include in it one additional element. The resulting set can be put in one-to-one correspondence with the set of natural numbers in the following way:

Start with the set of natural numbers: $\{1, 2, 3, \ldots\}$. Call the additional element "*." Adding *, we get the set $\{*, 1, 2, 3, \ldots\}$. We can now map the set of natural numbers one-to-one onto this set in the following way: Map 1 to *, 2 to 1, 3 to 2, and so on. Thus the set with cardinal number $\aleph_0 + 1$ is in one-to-one correspondence with the set of natural numbers which has the cardinal number $\aleph_0$. Thus, $\aleph_0 + 1 = \aleph_0$.

Note, incidentally, that including an additional element * in the set of natural numbers yields the same set as that including all the natural numbers in a set with the element *. For this reason, addition is commutative for the transfinite cardinal numbers: $1 + \aleph_0 = \aleph_0 + 1$.

By arguments of the above form, we can arrive at some basic principles of cardinal arithmetic.

$k + \aleph_0 = \aleph_0 + k = \aleph_0$, for any natural number k.
$\aleph_0 + \aleph_0 = \aleph_0$
$2 \cdot \aleph_0 = \aleph_0$
$k \cdot \aleph_0 = \aleph_0$, for any natural number k.
$\aleph_0 + \aleph_0 + \aleph_0 + \ldots = \aleph_0$
$\aleph_0 \cdot \aleph_0 = \aleph_0$
$(\aleph_0)^2 = \aleph_0$
$(\aleph_0)^k = \aleph_0$, for any natural number k.

Beyond $\aleph_0$

Cantor, as we have seen, proved that there are sets with a number of elements greater than $\aleph_0$. He used the symbol $\aleph_1$ to indicate the next largest cardinal number greater than $\aleph_0$. He hypothesized that that number was the number of members of the set of real numbers, which he had called "C" and had proved equal to $2^{\aleph_0}$. This is called the *Continuum hypothesis*, written as: $2^{\aleph_0} = \aleph_1$. It says that if "larger" and "smaller" are defined by Cantor's metaphor, then the smallest set larger than $\aleph_0$ has the cardinal number (i.e., the Cantor size) $2^{\aleph_0}$.

Cantor stated the Continuum hypothesis as a conjecture, which he was never able to prove. Later, Kurt Gödel (in 1938) and Paul Cohen (in 1963) proved theorems that together showed that the conjecture is neither absolutely true nor absolutely false relative to generally accepted axioms for set theory. Its truth depends on what further axioms one takes "set" to be defined by. From the perspective of our mathematical idea analysis, this means that whether or not the Continuum hypothesis is "true" depends on the underlying conceptual metaphors characterizing the concept "set."

The Hierarchy of Transfinite Cardinals

A power set is always larger than the set it was based on. Moreover, given a power set, we can always form the power set of that power set, and so on. This is true for infinite as well as finite sets within Cantor's set theory.

Given a set of cardinality $\aleph_0$, we can form the power set of the power set of . . . of the power set of the natural numbers. This is a hierarchy of sets of ever-increasing Cantor size:

$$\aleph_0 < 2^{\aleph_0} < 2^{2^{\aleph_0}} < \ldots$$

To summarize, the Basic Metaphor of Infinity allows us to form the entire set of natural numbers. Cantor's metaphor, applied with the BMI, allows us to derive the cardinal number, $\aleph_0$, for the entire infinite set of natural numbers. Cantor's metaphor and the BMI jointly allow us to "prove" that there are "more" reals than natural numbers and rational numbers. Cantor's metaphor allowed Cantor to prove that there are always more elements in a power set than in its base set, even for infinite sets. And this result provides a hierarchy of transfinite cardinal numbers in Cantor's set theory.

Ordinal Numbers

Take an ordinary natural number—say, 17. This number can be used in at least two ways. It can have a *cardinal* use; that is, it can be used to indicate how many elements there are in some collection. For example, there might be 17 pens in your desk drawer. It can also have an *ordinal* use; that is, it can be used to indicate a position in a sequence. For example, today might be the 17th day since it last rained. These are two very different uses of numbers.

Natural numbers are neutral between these two uses. They have exactly the same properties, independently of how they are being used. For example, 17 is prime, whether you use it to estimate size or position in a sequence. $17 + 1 = 1 + 17$, no matter which of those two uses you are making of numbers.

Most important, there is a simple relation between the cardinal and ordinal uses of natural numbers. Suppose you are forming a collection. You lay out 17 items in sequence to go into the collection. The last item is the 17th item. This is an ordinal use of 17. For convenience we will say that such a sequence is 17 elements long—or equivalently, that it has a length of 17. When we put this sequence of elements into a collection, the number of the items in the collection is 17. This is a cardinal use of 17. In this case, the (ordinal) number for the last item counted corresponds to the (cardinal) number for the size of the collection. Ordinality and cardinality always yield the same number, no matter how you count. For this reason, the arithmetic of the natural numbers is the same for cardinal and ordinal uses.

But this is not true for transfinite numbers. Cantor's metaphor determines "size" for an infinite collection by pairing, not counting in a sequence. Cantor's metaphor, therefore, is only about cardinality (i.e., "size") not about ordinality (i.e., sequence). When Cantor's metaphor is used to form transfinite numbers, those metaphorically constituted "numbers" are defined in terms of sets, with no sequences. Such sets have no internal ordering and so can have only cardinal uses, not ordinal uses.

What this means is that in the realm of the infinite there are no transfinite numbers that can have both cardinal and ordinal uses. Rather, two different types of numbers are needed, each with its own properties and its own arithmetic. And as we will see shortly, there is another strange difference between the cardinal and the ordinal in the domain of the infinite: You can get different results by "counting" the members of a fixed collection in different orders! The reason for this has to do with the different metaphors needed to extend the concept of number to the transfinite domain for ordinal as opposed to cardinal uses.

The Transfinite Ordinals

In the Basic Metaphor of Infinity, let the unending process be the process of generating the natural numbers so as to form a sequence, with number 1 first, number 2 second, and so on. The sequence—not the set!—formed is (1, 2, 3, . . .). This is the sequence of the natural numbers *in their natural order*. Here 2 is the second ordinal number, 3 is the third ordinal number, and so on.

Given this process of sequence formation, a special case of the BMI will create a final resultant state of the process: the infinite sequence of all the natural numbers in their natural order. Such a special case of the BMI will also assign to this entire infinite sequence an ordinal number, which Cantor called "ω." This is the first transfinite ordinal number. Every sequence of length ω is an infinite sequence that has been "completed" by virtue of the BMI. It has an endpoint—the number ω. The last position in this sequence is thus called the ω position.

TRANSFINITE ORDINALS

Target Domain ITERATIVE PROCESSES THAT GO ON AND ON		*Special Case* THE SEQUENCE OF THE NATURAL NUMBERS IN THEIR NATURAL ORDER
The beginning state (0)	⇒	No sequence
State (1) resulting from the initial stage of the process	⇒	The 1-place sequence (1), whose length is the ordinal number 1
The process: From a prior intermediate state (n–1), produce the next state (n).	⇒	Given sequence (1, . . . , n–1), form the next longest sequence (1, . . . , n).
The intermediate result after that iteration of the process	⇒	The sequence (1, . . . , n) whose length is the ordinal number n

"The final resultant state" (actual infinity "∞")	$\Rightarrow$	**The infinite sequence (1, . . . n, . . .), whose length is the transfinite ordinal number ω.**
Entailment *E:* The final resultant state ("∞") is unique and follows every nonfinal state.	$\Rightarrow$	**Entailment *E:* ω is the unique smallest transfinite ordinal number, and it is larger than any finite ordinal number.**

What distinguishes this special case of the BMI from others—for example, the set of natural numbers, the "number" ∞, and so on? Here we are building up *sequences* as *results* of stages of processes. In the case of ∞, the results were numbers. In the case of the set of all natural numbers, the results were sets of natural numbers. Infinite sequences, counted by ordinal numbers, just work differently than individual numbers or sets of numbers—as Cantor recognized clearly.

Other sequential arrangements of the same numbers in different orders will also have a length given by the ordinal number ω. For example, the following sequences all have length ω:

2, 1, 3, 4, 5, . . .
3, 2, 1, 4, 5, . . .
4, 1, 2, 3, 5, 6, . . .

Cantor's Ordinal Metaphor

To make arithmetic sense of infinite ordinal numbers for infinite sequences of all sorts, Cantor assumed another metaphor:

CANTOR'S ORDINAL METAPHOR

Source Domain ONE-TO-ONE MAPPINGS FOR SEQUENCES		*Target Domain* ORDINAL NUMBERS
Sequences A and B can be put in a one-to-one correspondence	$\rightarrow$	Sequences A and B have the same ordinal Number

The problem here is: What is to count as a one-to-one correspondence for an infinite sequence? Cantor assumed an account of one-to-one correspondences for sequences that made implicit use of the BMI. He separated the finite terms of sequences using the comma symbol, as in (1, 2, 3, . . .). However, he used the semicolon to indicate the end of an implicit use of the BMI.

Consider the following sequences:

$$1, 3, 5, 7, \ldots$$
$$2, 4, 6, 8, \ldots$$

Just as the set of odd numbers can be put in one-to-one correspondence with the natural numbers, so the *sequence* of odd numbers can be put in one-to-one correspondence with the sequence of the natural numbers. Therefore, they both have the same ordinal number, ω. The same is true of the even numbers. Each of these has one use of ". . . " at the end, indicating one implicit use of the BMI to complete the sequence.

Thus, Cantor implicitly used the BMI as part of his understanding of what a "one-to-one correspondence for an infinite sequence" is to mean. The ". . . " indicates an implicit use of the BMI, and it counts as part of the definition of one-to-one correspondence for a sequence.

Now consider the sequence of all the odd numbers "followed by" the sequence of all the even numbers:

$$1, 3, 5, 7, \ldots; 2, 4, 6, 8, \ldots$$

The sequence of odd numbers (1, 3, 5, 7, . . .) contains a ". . . ," indicating the implicit use of the BMI to complete the sequence. The sequence of all the odd numbers "followed by" the sequence of all the even numbers contains two uses of ". . . ," indicating two implicit uses of the BMI to complete the sequences. Since the complete sequence of the odd numbers has length ω, and the complete sequence of even numbers following it also has length ω, the entire sequence has length $\omega + \omega = 2\omega$.

Defining one-to-one correspondence for an infinite sequence in this fashion has important consequences for an arithmetic of transfinite ordinal numbers. It is quite different from the arithmetic of transfinite cardinal numbers. Recall that in transfinite cardinal arithmetic, $\aleph_0 + \aleph_0 = \aleph_0$, whereas in transfinite ordinal arithmetic, $\omega + \omega = 2\omega$.

Recall that "addition" for transfinite cardinals is defined in terms of set union. Recall also that the set of positive odd integers can be put in one-to-one correspondence with the natural numbers, and so it has the cardinal number $\aleph_0$. The same is true of the set of positive even integers. Thus, the union of the set of positive odd integers and the set of positive even integers has the cardinality $\aleph_0 + \aleph_0$. Since this set union is just the set of all natural numbers $\aleph_0$, it follows that $\aleph_0 + \aleph_0 = \aleph_0$. Similarly, $\aleph_0 + 1 = \aleph_0$, since the union of a set the size of the natural numbers with a set of one element can be put in one-to-one correspondence with the set of natural numbers, as we have seen.

But the situation is very different with the ordinal numbers. Once we have established the ω position in an infinite sequence—the "last" position in that sequence—we can go on forming longer sequences by appending a further sequence "after" the ω position. If we add another element to the sequence already "counted" up to ω, the element added will be the $(\omega + 1)$ member of the sequence. Add another element and we have the $(\omega + 2)$ member of the sequence. Add ω more elements (via another application of the BMI) and we have the $(\omega + \omega)$, or (2ω), member of the sequence. If we keep adding ω elements to the sequence and you do it ω times, then the BMI will apply ω times, and the resulting sequence will be ω^ω elements long. We can keep on going, generating an unending hierarchy of transfinite ordinal numbers—namely:

$$\omega^{\omega^{\omega^{\cdots}}}$$

In this way, an ordinal arithmetic is built up.

$$\omega \neq \omega + 1$$
$$\omega + \omega = 2\omega$$
$$\omega \cdot \omega = \omega^2$$
$$\omega^a \cdot \omega^b = \omega^{a+b}$$

Incidentally, not all the basic laws of arithmetic apply to transfinite ordinals. For example, the commutative law does not hold: $1 + \omega \neq \omega + 1$. As we saw, $\omega \neq \omega + 1$. However, $1 + \omega = \omega$. Here is the reason: $1 + \omega$ is the length of a sequence. That sequence is made up of a sequence of length 1 followed by a sequence of length ω. For example, consider the sequence of the number 1 followed by the infinite sequence of the numbers 2, 3, 4, ... This is a sequence of length 1 followed by a sequence of length ω. But this sequence is the same as the sequence 1, 2, 3, 4, . . . , which is of length ω. For this reason, $1 + \omega = \omega$. But $\omega \neq \omega + 1$. Therefore, $1 + \omega \neq \omega + 1$, and the commutative law does not hold.

The general reason for this is that "+" is metaphorically defined in sequential terms in ordinal arithmetic, where the BMI applies to define infinite sequences. "$A + B$" in ordinal arithmetic refers to the length of a sequence. That sequence is made up of a sequence of length A followed by a sequence of length B. Addition is therefore asymmetrical when A is transfinite and B is not. If A is ω, which is the smallest transfinite ordinal number, adding a sequence with a finite length B after A will therefore always yield a sequence whose length is greater than ω. The reason is that the BMI produces the infinite sequence A with a "last" term. Putting a finite sequence before that "last" term does not result in a higher order of infinity. But putting it after that "last" infinite term produces a "longer" infinite sequence.

Suppose the sequence of finite length B comes first, followed by a sequence of length ω. We will then get a sequence: $1, \ldots, B$ followed by a sequence of length ω, the shortest infinite sequence. Adding a finite number of terms *at the beginning* of the shortest infinite sequence does not make that infinite sequence any "longer."

Summary

What is the difference between $\aleph_0$ and ω? The first is a transfinite cardinal that tells us "how many" elements are in an infinite set. The second is a transfinite ordinal that indicates a position in an infinite sequence.

The transfinite cardinal number $\aleph_0$ is, via the cardinal number metaphor, the set of all sets that are in one-to-one correspondence with the set of all natural numbers. The set of all natural numbers is generated by the Basic Metaphor of Infinity. $\aleph_0$ indicates the "size" of an infinite set by virtue of Cantor's metaphor (Cantor size). In this way, $\aleph_0$ is characterized by these three metaphors.

The transfinite ordinal number ω is the ordinal number corresponding to the well-ordered set of all natural numbers. It is generated by the Basic Metaphor of Infinity and Cantor's ordinal metaphor. This difference between transfinite cardinals and transfinite ordinals is the root of the difference between their arithmetic.

$\aleph_0$ and ω are both infinite numbers of very different sorts, with different arithmetics. Both are distinct from the infinite number ∞. As we saw in Chapter 8, ∞ is generated by the Basic Metaphor of Infinity when applied to the process of forming natural numbers via the Larger Than relation. Via this application of the BMI, ∞ is the largest of the natural numbers, which, as we saw, is the only numerical property it has. Otherwise, it is arithmetically defective. $\infty + 1$, for example, is not defined, while $\aleph_0 + 1$ and $\omega + 1$ are well defined and part of precise arithmetics—different precise arithmetics.

We should point out that the symbol + in "$\aleph_0 + 1$" and "$\omega + 1$" stands for different concepts than the + we find in ordinary arithmetic (e.g., in "$2 + 7$"). Types of transfinite addition are not mere generalizations over the usual forms of addition for finite numbers; rather, they are inherently metaphorical. They operate on metaphorical entities like $\aleph_0$ and ω, which are conceptualized via the BMI. Transfinite cardinal and ordinal addition not only work by different laws (different from ordinary addition and from each other) but are defined by infinite operations on sets and sequences, unlike ordinary addition.

What does mathematical idea analysis tell us in these cases? First, it shows explicitly how the BMI is involved in characterizing transfinite cardinals and ordinals. Second, it tells us explicitly what metaphors distinguish the transfinite cardinals from the transfinite ordinals. And, third, it shows not only how transfinite cardinal and ordinal arithmetics arise as consequences of these metaphor combinations but exactly *why* and *how* their arithmetics are different.

Finally, such an analysis shows how transfinite numbers are embodied. They are not transcendental abstractions but arise via such general, *nonmathematical* embodied mechanisms as aspect and conceptual metaphor.

11

Infinitesimals

Specks

When we look into the distance, normal-size objects that are very far away appear to us as specks, barely discernible entities so small as to have no internal structure. A large object in your field of vision—say, a mountain or a stadium—will have a discernible width. But the width of the stadium plus the width of the speck is not discernible at all. Nor is the width of one speck discernible from the widths of three specks or five specks. But you know that if you could zoom in on a speck sufficiently closely, it would look like a normal-size object.

Our conception of a speck, if arithmetized, might go like this: The widths of discernible objects correspond to real numbers. A speck-number ∂ is greater than zero but less than any real number. A real number plus a speck number is not discernibly different from the real number alone: $r + \partial \approx r$. A small number of adjacent specks are not discernibly different from one speck: $r \cdot \partial \approx \partial$, where r is small.

Infinitesimals

Our everyday experience with specks corresponds in these basic ways to mathematical entities called *infinitesimals*. Within mathematics, the idea of an infinitesimal became popular with Leibniz's development of calculus. Both Leibniz and Newton, in developing independently the calculus of instantaneous change, used the metaphor that Instantaneous Change Is Average Change over an Infinitely Small Interval. Thus, for a function $f(x)$ and an interval of length Δx, instantaneous change is formulated as $\frac{f(x + \Delta x) - f(x)}{\Delta x}$. The instantaneous change in

$f(x)$ at x is arrived at when Δx is "infinitely small." The concept of "infinitely small," of course, has to be arithmetized to get a numerical result. This requires an arithmetization metaphor.

Leibniz and Newton had different metaphors for arithmetizing "an infinitely small interval." For Newton, the derivative of a function for a given number was (metaphorically) the tangent of the curve of the function at the point corresponding to that number. But, as we have seen, that tangent cannot be used with the metaphor that Instantaneous Change Is Average Change over an Infinitely Small Interval, because *at* the tangent the infinitely small interval is zero and you can't divide by zero. So Newton came up with the limit metaphor, a version of the Basic Metaphor of Infinity (Chapters 8 and 9). The tangent line was conceptualized metaphorically as the limit of a sequence of secant lines, with the secant lines becoming progressively smaller but always having a real length. As Δx approached zero (without reaching it), the secant lines approached the tangent (see Figure 11.1). When Δx got arbitrarily small (but remained a real number), the difference between the secant line and the tangent was some arithmetic function of the Δx's. If that arithmetic function of the Δx's also got arbitrarily small—insignificant, for any practical purpose—*one could ignore it*. By ignoring this secant-tangent difference, the calculation based on a sequence of secants could be used for determining the tangent closely enough. Newton's solution is the one that has come down to us today as the "definition" of the derivative.

Leibniz solved the problem of arithmetizing "an infinitely small interval" in a very different way. He hypothesized infinitely small numbers—infinitesimals—to designate the size of infinitely small intervals. For Leibniz, *dx* was an infinitesimal number, a number greater than zero but less than all real numbers. For Leibniz, "$df(x)/dx$" was a ratio of two infinitesimal numbers (see Figure 11.2). Since neither number was zero, division was possible. The quotient—the tangent to the curve—was a real-valued function plus some arithmetic function of infinitesimals. Any arithmetic function of infinitesimals was so small it could be ignored.

As far as calculation was concerned, there was no difference between the two approaches. Both yielded the same results. For two hundred years after Leibniz, mathematicians thought in terms of infinitesimals, used them in calculations, and always got accurate results. But ontologically there is a world of difference between what Newton did and what Leibniz did. Newton did not need infinitesimal *numbers as existing mathematical entities*. His secants always had a real length, while Leibniz's had an infinitesimal length. Leibniz hypothesized a new kind of number. Newton did not.

On the other hand, Newtonians had to use geometry, with tangents and secants, to do calculus. Leibnizians could do calculus using arithmetic without

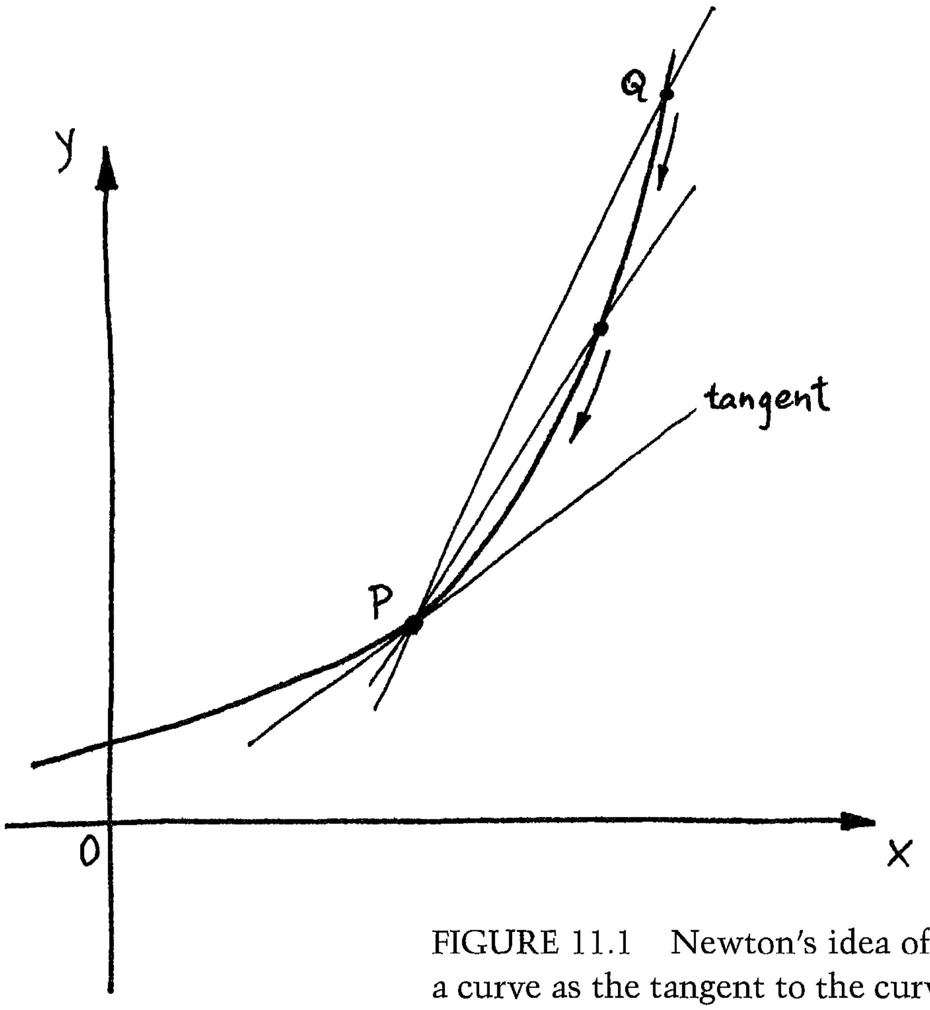


FIGURE 11.1 Newton's idea of the derivative at a point P on a curve as the tangent to the curve at P. He saw this tangent as the limit of a sequence of secants, generated by the movement of point Q on the curve toward point P.

geometry—by using infinitesimal *numbers*. Until the work of Karl Weierstrass in the late nineteenth century, the Newtonian approach to calculus relied on geometry and lacked a pure nongeometric arithmetization. Many mathematicians saw geometry as relying on "intuition," whereas arithmetic was seen as more precise and "rigorous." This perceived imprecision was considered a disadvantage of the Newtonian approach to calculus; its advantage was that it did not rely on infinitesimal numbers.

The Archimedean Principle

As we saw at the beginning of the chapter, infinitesimals require a somewhat different arithmetic than real numbers. Real numbers obey the *Archimedean principle*. Given any real number s, however small, and any arbitrarily large real number L, you can always add the small number to itself enough times to make it larger than the large number: $s + s + \ldots + s > L$, for some sufficiently large number of additions. Equivalently, there is a real number r, such that $r \cdot s > L$.

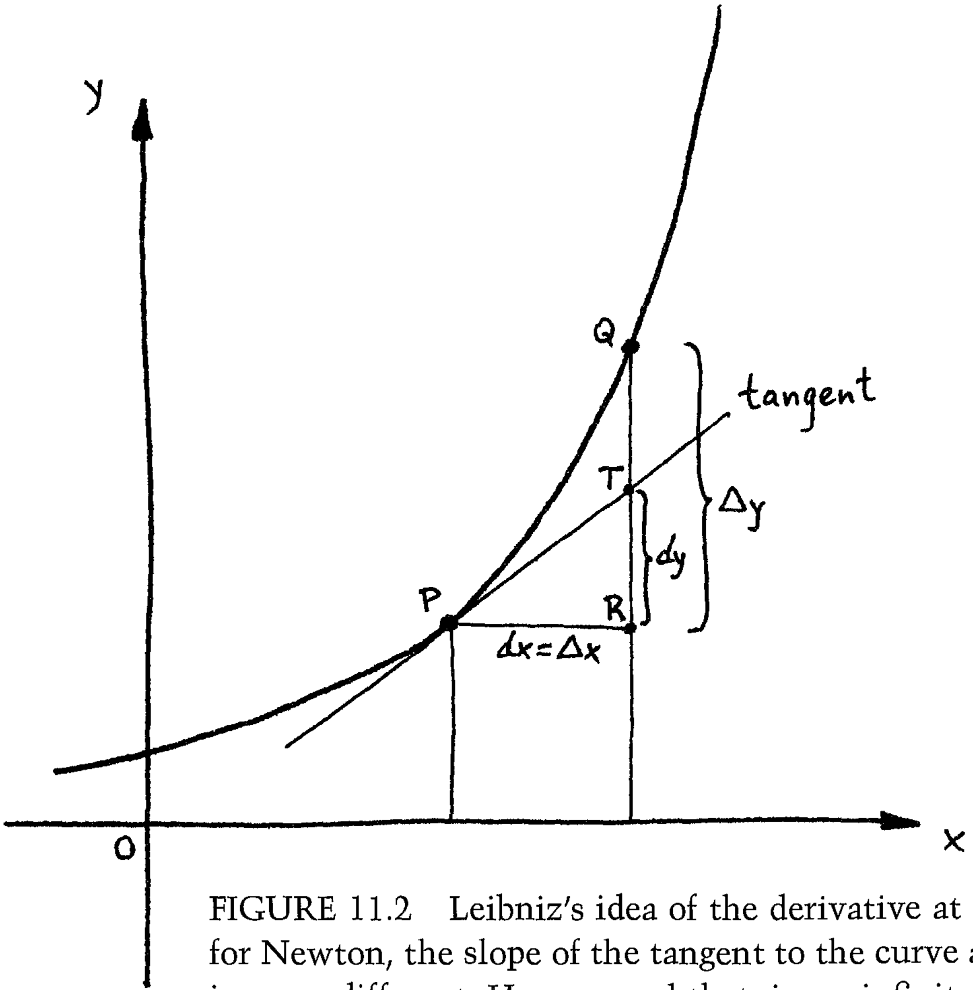


FIGURE 11.2 Leibniz's idea of the derivative at a point P on a curve was, as for Newton, the slope of the tangent to the curve at that point. But his reasoning was different. He reasoned that, in an infinitesimally small region around P, PT and PQ would be virtually identical. Thus, if the change in x (i.e., Δx) were infinitesimally small, the change in y (i.e., Δy) would be infinitely close to the rise of the tangent (dy). Thus, $\Delta y/\Delta x$ would be infinitesimally close to dy/dx—the slope of the tangent.

But this is not true of infinitesimals. Infinitesimals are so small that there is no real number large enough so that $r \cdot \partial > L$. You can't add up enough infinitesimals to get any real number, no matter how small that real number is and no matter how many infinitesimals you use.

Another way of thinking of this has to do with the concept of commensurability—having a common measure. Take two real numbers r_1 and r_2. There is always a way of performing arithmetic operations on r_1 with other real numbers to yield the number r_2. For example, given π and $\sqrt{2}$, you can multiply π by the real number $\sqrt{2}/\pi$ to get $\sqrt{2}$. But, given a real number r and an infinitesimal ∂, there are no arithmetic operations on r using the real numbers alone that will allow you to get ∂. The infinitesimals and the reals are therefore incommensurable.

The fact that the infinitesimals are not Archimedean and that they are incommensurable with the real numbers has made many mathematicians wary of them—so wary that they will not grant their existence. After all, all of classical mathematics can be done in the Newtonian way, without infinitesimals. If they are not needed, why even bother to consider them? Citing Occam's Razor, the a priori philosophical principle that you should not postulate the existence of an entity if you can get by without it, many mathematicians shun the infinitesimal numbers.

Why Monads Have No Least Upper Bounds

Infinitesimals interact with real numbers: They can be added to, subtracted from, and multiplied and divided by real numbers. Given any real number r, you can add to it and subtract from it any number of infinitesimals without reaching any other real number greater or smaller than r. This means that around each real number r there is a huge cluster of infinitesimals. That cluster was referred to by Leibniz as a *monad*.

Can a monad around a real number reach a monad around a neighboring real number? The answer is no. Here is the reason. Suppose the monads around two real numbers, r_1 and r_2, such that $r_1 < r_2$, were to overlap. This means that there would be two infinitesimal numbers ∂_1 and ∂_2 such that $r_1 + \partial_1 > r_2 - \partial_2$ even though $r_1 < r_2$. This would mean that by adding $\partial_1 + \partial_2$ to r_1 one could go beyond r_2. But this, in turn, would mean that the Archimedean principle worked for infinitesimals, which it cannot. Therefore, the monads around two real numbers can never overlap. No monad can ever reach any other monad.

Let us now ask: Can the monad around a real number r—call it $M(r)$—have a least upper bound? Suppose it did. Let the least upper bound of $M(r)$ be L. If L were real, L could have a monad of infinitesimals around it. In that monad would be an infinitesimal copy of the negative half of the real line, with each number on it less than L but still greater than all the members of $M(r)$. This means that each number in the negative half of the monad around L is an upper bound of $M(r)$ and less than L. Thus, L cannot be a least upper bound of $M(r)$. Now suppose that the least upper bound, L, of $M(r)$ were not a real number but a member of a monad around some real number r. The same argument would apply but in a smaller scale. L would have a tiny monad around it, and each number in the negative half of the monad around L would be less than L and greater than every member of $M(r)$. Thus each would be an upper bound of $M(r)$ but less than L, and L could not be a least upper bound.

Therefore, when the infinitesimals are combined with the real numbers, the least upper bound property cannot hold in general. However, because the least upper bound property holds of sets of real numbers, it will also hold of sets of numbers on infinitesimal copies of the real line.

The other nine axioms of the real numbers can apply to infinitesimals without contradiction.

An Implicit Use of the BMI

Though there is no *arithmetic* way to get the infinitesimals from the reals, there is a metaphorical mechanism that will do just that: the Basic Metaphor of Infinity. Newton explicitly used the notion of the limit, which implicitly uses the BMI, to characterize calculus. A different use of the BMI is implicit in Leibniz's development of calculus using infinitesimals.

We can get the infinitesimals from the reals using the BMI as follows: Form the following sequence of sets starting with the real numbers, as characterized by the first nine axioms for the real numbers in Chapter 9. (Recall that the Archimedean principle is *not* one of these axioms.) In the rest of the chapter, we will speak of a set of numbers, "satisfying" certain axioms. This is a shorthand meaning: the set of those numbers, such that they, together with their additive and multiplicative inverses and the identity elements 1 and 0, form a larger set that satisfies the specified axioms.

The set of all numbers bigger than zero and less than 1/1, satisfying the first nine axioms for the real numbers.
The set of all numbers bigger than zero and less than 1/2, satisfying the first nine axioms for the real numbers.
The set of all numbers bigger than zero and less than 1/3, satisfying the first nine axioms for the real numbers.

And so on.
Let this be the unending process in the BMI. Using the BMI, we get:

The set of all numbers greater than zero and smaller than *all* real numbers, satisfying the first nine axioms for the real numbers.

This is the set of infinitesimals.
Here is the special case of the BMI that gives rise to the infinitesimals:

INFINITESIMALS

Target Domain ITERATIVE PROCESSES THAT GO ON AND ON		*Special Case* INFINITESIMALS
The beginning state (0)	⇒	Consider sets of numbers that are greater than zero and less than $1/n$, satisfying the first nine axioms for the real numbers.

State (1) resulting from the initial stage of the process	$\Rightarrow$	$S(1)$: the set of all numbers greater than zero and less than $1/1$, satisfying the first nine axioms for the real numbers
The process: From a prior intermediate state (n-1), produce the next state (n).	$\Rightarrow$	From $S(n\text{-}1)$, form $S(n)$.
The intermediate result after that iteration of the process (the relation between n and n–1)	$\Rightarrow$	$S(n)$: the set of all numbers greater than zero and less than $1/n$, satisfying the first nine axioms for real numbers.
"The final resultant state" (actual infinity "∞")	$\Rightarrow$	**$S(\infty)$: the set of all numbers greater than zero and less than *all* real numbers, satisfying the first nine axioms for the real numbers**
Entailment *E:* The final resultant state ("∞") is unique and follows every nonfinal state.	$\Rightarrow$	**Entailment *E:* $S(\infty)$ is unique.**

These infinitesimal numbers are metaphorical creations of the BMI. In the iterations, the numbers get smaller and smaller but are always real numbers at every finite stage. It is only at the final stage, characterized by the BMI, that we get a set of numbers greater than zero and smaller than *all* real numbers.

Because infinitesimals satisfy the first nine axioms for the real numbers, arithmetic operations apply to them. Thus you can multiply real numbers by infinitesimals—for example, $3 \cdot \partial = \partial \cdot 3 = 3\partial$. There is a multiplicative inverse:

$$1/\partial, \text{ such that } \partial \cdot 1/\partial = 1.$$

You can also take ratios of infinitesimals:

$$3\partial/4\partial = (3{\cdot}\partial)/(4{\cdot}\partial) = 3/4 \cdot \partial/\partial = 3/4 \cdot 1 = 3/4.$$

Thus, ratios of infinitesimals can be real numbers, as required by Leibnizian calculus.

Since the concept of the limit and the concept of the infinitesimals are special cases of the BMI, it is no surprise that Newton and Leibniz got equivalent results. They *both* implicitly used the Basic Metaphor of Infinity to extend the real numbers and the operations on them, but they used it in different ways. Newton used the BMI to define a new *operation*—taking the limit—but not new metaphorical entities. Leibniz used the BMI to define new *entities*—infinitesimals—but not a new metaphorical operation.

As for the principle of Occam's Razor, one might as easily apply it to kinds of operations as to kinds of entities. Thus, one might argue, against the Newtonian method, that it required a new kind of operation—limits that could be avoided if you had infinitesimals. Occam's Razor cuts both ways. A pro-infinitesimal anti-limit argument is just as much an Occam's Razor argument as a pro-limit anti-infinitesimal argument. Occam's Razor (if one chooses to accept it) in itself thus does not decide which use of the Basic Metaphor of Infinity, Newton's or Leibniz's, should be accepted for characterizing calculus. For historical reasons, Newton's use of the BMI was ultimately accepted, but both versions were used for a long time.

Robinson and the Hyperreals

In the late nineteenth century, the German mathematician Karl Weierstrass developed a pure nongeometric arithmetization for Newtonian calculus (see Chapter 14). This work effectively wiped out the use of infinitesimal arithmetic in doing calculus for over half a century. The revival of the use of infinitesimals began in the 1940s, with research by Leon Henkin and Anatolii Malcev. There is a theorem in mathematical logic called the Compactness theorem, which says:

> In a logical system L that can be satisfied by (i.e., mapped consistently onto) a set-theoretical model M, if every finite subset of a collection of sentences of L is satisfied in M, then the entire collection is satisfied in some model M^*.

Here is the reason that this theorem is true.

- Every finite subset of sentences that is satisfied in M can be consistently mapped onto M.
- If a finite subset of sentences can consistently be mapped onto a model M, then that subset of sentences is consistent; that is, it does not contain contradictions.
- If every finite subset of sentences is satisfied in M, then every finite subset is consistent.
- If every finite subset is consistent, then the entire collection is consistent and therefore has some model M^*, which may not be the same as M.

From a conceptual perspective, we can see the Compactness theorem as another special case of the BMI. Here are the BMI parameters that allow us to see the Compactness theorem as a form of the BMI:

THE LOGICAL COMPACTNESS FRAME

The sentences of a formal logic are denumerable; that is, they can be placed in a one-to-one correspondence with the natural numbers. If every finite subset of a set of sentences S has a model, then the subset $S(n)$ consisting of the first n sentences of S has a model $M(n)$.

THE BMI FOR LOGICAL COMPACTNESS

Target Domain ITERATIVE PROCESSES THAT GO ON AND ON		*Special Case* LOGICAL COMPACTNESS
The beginning state (0)	⇒	The logical compactness frame, with models for all finite sets of sentences
State (1) resulting from the initial stage of the process	⇒	There is an ordered pair ($S(1)$, $M(1)$), consisting of the first sentence and a model that satisfies it.
The process: From a prior intermediate state ($n - 1$), produce the next state (n).	⇒	Given ($S(n - 1)$, $M(n - 1)$), there is a model $M(n)$ satisfying $S(n)$.
The intermediate result after that iteration of the process (the relation between n and $n - 1$)	⇒	There is an ordered pair ($S(n)$, $M(n)$), consisting of the set $S(n)$ of the first n sentences and the model $M(n)$ that satisfies it.
"The final resultant state" (actual infinity "∞")	⇒	**There is an ordered pair ($S(\infty)$, $M(\infty)$), consisting of the set $S(\infty)$ of the *entire* infinite sequence of sentences and a model $M(\infty)$ that satisfies $S(\infty)$.**

Since the subject matter of this special case is the existence of some model or other, not a particular model, the uniqueness entailment applicable to other special cases of the BMI does not arise here.

This version of the BMI works as follows: Suppose we know that every finite set of sentences of S has a model. Since the sentences are countable, they can be ordered. Form the sequence of the first i subsets of sentences and the model paired with them. At each stage, there is one more sentence and a simple extension of the previous model to satisfy that sentence, given the others. The process of forming larger and larger subsets of sentences and larger and larger extensions of the previous models to satisfy those sentences goes on indefinitely. The BMI adds an endpoint at which all the sentences are considered together and an ultimate model is constructed that satisfies not just all finite subsets of sentences but the infinite set of all the sentences.

Of course, this is a cognitive characterization, not a mathematical characterization of the Compactness theorem. The mathematical proof uses the notion of a least upper bound/greatest lower bound (Bell & Slomson, 1969), which, from a cognitive viewpoint, is another version of the BMI (see Chapter 9). The mathematical version of the Compactness theorem can now be used to generate a model in which there is at least one infinitesimal. The result is due to Henkin.

Henkin considered the following set of sentences for some constant, C:

C is a number bigger than zero and less than 1/2.
C is a number bigger than zero and less than 1/3.
C is a number bigger than zero and less than 1/4.
And so on.

For any finite subset of such sentences, there is some number C such that the finite subset of sentences is satisfied in any model of the axioms for the real numbers. According to the Compactness theorem, there is a model M^* in which all such sentences are satisfied. In that model, C must be greater than zero and smaller than any real number. As such, C is an infinitesimal. Hence, model theory shows that there are set-theoretical models of the real numbers in which at least one infinitesimal exists.

Since this is a result in classical mathematics, the conclusion came as something of a shock to those who thought that classical mathematics could get along without the existence of infinitesimals. The Compactness theorem requires that infinitesimals exist, even within the bounds of classical mathematics. In other words, Occam's Razor or no Occam's Razor, if there are least upper bounds, there are infinitesimals.

Of course, once you have a set-theoretical structure with any infinitesimals in it, the axioms of Infinity (another special case of the BMI) and Union (Chapter 7), taken together, allow for the formation of infinite unions of sets, so that there can then be set-theoretical models containing hugely more infinitesimals.

Once infinitesimals were shown to "exist" relative to the assumptions of classical mathematics and formal model theory, an entire full-blown formal theory of infinitesimal mathematics and calculus using them was constructed, mostly due to the work of Abraham Robinson. A branch of mathematics called *nonstandard analysis* has come out of Robinson's (1966, 1979) work.

It is not our intention here to review the field of nonstandard analysis. What we want to do instead is give the reader a conceptual feel of what infinitesimal numbers are, how they work, and what deep philosophical consequences follow from their "existence." We will use the Basic Metaphor of Infinity to show what the structure of infinitesimal arithmetic is like and how infinitesimals operate. Then we will show the consequences.

The remainder of this chapter will take the following form:

- First, we will use the BMI to generate "the first infinitesimal," which we will call ∂.
- Second, we will show that ∂ together with the reals forms a number system we will call the *granular numbers.*
- Third, we will form a system of numerals for the granular numbers, using "∂" as a numeral for the first infinitesimal number.
- Fourth, we will show that an understanding of this number system can be grounded in an understanding of the concept of a *speck.*
- Fifth, we will show a truly remarkable result. In the granular numbers, the sum of an infinite series *is not equal to its limit!* Moreover, the difference between the limit and the infinite sum can be precisely calculated in the granular numbers.
- Sixth, we will indicate how all of calculus can be done with the granular numbers. Indeed, there is a one-to-one translation of classical proofs in calculus without infinitesimals to a corresponding proof using the granular numbers.
- Seventh, we will then use the BMI again to show that starting from the "first infinitesimal," we can generate a countable infinity of "layers" of infinitesimals. We then gather these together, using the BMI to form an infinite union to produce an infinite set of layers of infinitesimals. Whereas there is a numeral system for the number system generated by the first infinitesimal, and numeral systems at any finite layer of infinitesimals, there is no system of numerals that can describe the full set of infinitesimals. This gives us the unsurprising result: *There are number systems with no possible system of numerals.*

The Granular Numbers

We use the term "granularity" as a metaphorical extension of its use in photography. In a photograph, background details that look like specks may be sufficiently important that you want to zoom in on them and get a more "fine-grained" print that allows you to see those details with greater resolution. The degree to which you can zoom in or out is the degree of "granularity" of the print. Since our understanding of infinitesimals is grounded in our understanding of specks, it seems natural to use the term "granular numbers" for those infinitesimals that we want to zoom in on and examine in detail. What links our *experiential grounding* for infinitesimals to an *arithmetic* of infinitesimals is the metaphor that Multiplication by an Infinitesimal Is Zooming In. For example, multiplying the numbers in the real line by an infinitesimal allows us to zoom in on an infinitesimal copy of the real line.

We will approach the infinitesimals conceptually, rather than through the model-theoretic methods used by Robinson, and we will build them up gradually so as to get a better look at their structure. We have already discussed the entire set of infinitesimals via the Basic Metaphor of Infinity, using the unending process of constructing "the set of all numbers bigger then zero and less than $1/n$, satisfying the first nine axioms for the real numbers." We then used the BMI to add a metaphorical end to the process. This time we proceed in a different way, in order to give the reader both a sense of the structure of the infinitesimals and an idea of how slightly different special cases of the BMI can yield radically different results. Where our previous use of the BMI iteratively formed sets of numbers and finally formed the set of infinitesimal numbers, we will use a version of the BMI to iteratively form individual numbers, not sets. The result will be *a single infinitesimal number*, which we call ∂.

In this use of the BMI, we will use the following unending process of number formation, where n is a natural number:

- $1/1$ is a number greater than zero, satisfying the first nine axioms for real numbers, and of the form $1/n$ where n is an integer.
- $1/2$ is a number greater than zero, satisfying the first nine axioms for real numbers, and of the form $1/n$ where n is an integer.
- $1/3$ is a number greater than zero, satisfying the first nine axioms for real numbers, and of the form $1/n$ where n is an integer.
- $1/4$ is a number greater than zero, satisfying the first nine axioms for real numbers, and of the form $1/n$ where n is an integer.

As the integers n get larger, the numbers $1/n$ get smaller endlessly. The Basic Metaphor of Infinity conceptualizes this unending process as having an endpoint with a unique result: A number, ∂, of the form $1/n$, which is greater than zero, satisfies the first nine axioms for real numbers—and, via the BMI, is less than all real numbers. We will pronounce "∂" as "delta" and refer to it as "the first infinitesimal number produced by the BMI." We will include the symbol ∂ in the list of numerals 0, 1, 2, 3, and so on.

Since ∂ satisfies the first nine axioms for the real numbers, it combines freely with them. And since ∂ is also a numeral, we can write down all the combinations that define "granular arithmetic." Moreover, ∂ is of the form $1/k$, where k is an integer. We will call that integer H for "huge number," a number bigger than all real numbers. Note that $\mathrm{H} = 1/\partial$. Also note that the symbol H is not in italics, since H is a fixed number, not a variable.

Incidentally, the real numbers are contained among the granulars; expressions with no ∂'s are just ordinary real numbers. Note also that since the granulars satisfy the first nine axioms of the real numbers, they are linearly ordered.

For the sake of completeness, here is the special case of the BMI for forming ∂.

THE FIRST INFINITESIMAL

Target Domain ITERATIVE PROCESSES THAT GO ON AND ON		*Special Case* THE GRANULAR NUMBERS
The beginning state (0)	$\Rightarrow$	Individual numbers of the form $1/k$, each greater than zero and meeting the first nine axioms of the real numbers (where k is a positive integer)
State (1) resulting from the initial stage of the process	$\Rightarrow$	1/1: It is of the form $1/k$, is greater than zero, and meets the first nine axioms of the real numbers.
The process: From a prior intermediate state $(n - 1)$, produce the next state (n).	$\Rightarrow$	Given $1/(n - 1)$, form $1/n$. It is of the form $1/k$, which is greater than zero and meets the first nine axioms of the real numbers.
The intermediate result after that iteration of the process (the relation between n and $n - 1$)	$\Rightarrow$	$1/n$, with $1/n < 1/(n - 1)$
"The final resultant state" (actual infinity "∞")	$\Rightarrow$	**The infinitesimal number $1/\mathrm{H} = \partial$, where ∂ is greater than zero and meets the first nine axioms of the real numbers. H is an integer greater than all the real numbers.**
Entailment *E:* The final resultant state ("∞") is unique and follows every nonfinal state.	$\Rightarrow$	**Entailment *E:* $\partial = 1/\mathrm{H}$ is unique and is smaller than all $1/k$, where k is a natural number.**

Granular Arithmetic

By using normal arithmetic operations on ∂ and the reals, we arrive at an arithmetic of granular numbers. Since ∂ satisfies the first nine axioms for the real numbers, it has the normal properties that real numbers have.

$$0 + \partial = \partial + 0 = \partial \qquad \partial + (-\partial) = 0$$
$$1 \cdot \partial = \partial \cdot 1 = \partial \qquad \partial/\partial = 1$$
$$0 \cdot \partial = \partial \cdot 0 = 0 \qquad \partial^1 = \partial$$
$$\partial^0 = 1 \qquad \partial^{-1} = 1/\partial$$
$$\partial^n/\partial^m = \partial^{n-m}$$

Here, ∂ is acting like any nonzero real number.

The Special Property

What makes ∂ an infinitesimal is the special property it acquired by being created via the BMI: *Any multiple of ∂ and a real number is greater than zero and smaller than any positive real number.*

$$0 < \partial \cdot r < s, \text{ for any positive real numbers } r \text{ and } s.$$

Since $\partial/\partial = 1$, the ratio of two real multiples of ∂ is a real number. For example, $2\partial/5\partial = 2/5$, since $\partial/\partial = 1$.

Since ∂ can be multiplied by itself, we can have $\partial \cdot \partial = \partial^2$. Since multiplying anything by ∂ results in a quantity that is infinitely smaller, ∂^2 is infinitely smaller than ∂. As one continues to multiply by ∂, one gets numbers that are progressively infinitely smaller: ∂^{100}, $\partial^{1{,}000{,}000}$ and so on, without limit. As a consequence, all real multiples of ∂^2 are less than all real multiples of ∂.

The granulars also form infinite polynomials based on ∂. Thus, there are granulars like: $3\partial + 7\partial^2 + 64\partial^3 + \ldots$, where the sequence continues indefinitely.

The Monads of Real Numbers

∂ and its multiples can be added to or subtracted from any real number. Thus, we have granular numbers like $47 + 7\partial - 57\partial^2$ or $53 - 5\partial + 87\partial^5$. Thus any multiple of a real number times ∂ or of ∂ times itself occurs added to and subtracted from the real numbers. This has two remarkable consequences.

Suppose we extend the metaphor that Numbers Are Points on a Line to the granular numbers. We can do this, since the granular numbers are linearly ordered (because they satisfy the first nine axioms for the real numbers). That means that every granular number can be put in correspondence with a point on a line. The result is the *granular-number line,* which contains within it the points corresponding to all the real numbers. The first remarkable consequence is that on the granular-number line, the real numbers are rather sparse. Most of the numbers are granulars.

The fact that multiples of ∂ are less than any positive real number has a second remarkable consequence: the formation of monads, as described earlier. For convenience, we repeat the principal points of that discussion here.

Pick a real number and add to it and subtract from it any sum of multiples of ∂^n, where n is a natural number. By doing this you will not reach any other real number. Around each real number on the granular-number line, there is an infinity of granular numbers extending in both directions. This is the monad around that real number.

Let us look at the structure of each such monad around a real number. We can get a sense of that structure by performing the following construction:

- Take the real-number line and multiply every number in it by ∂. Call this the ∂^1 number line.
- Take the ∂^1 number line and multiply every number in it by ∂. The result consists of each real number multiplied by ∂^2. Call this the ∂^2 number line.
- And so on for ∂^n.
- Now pick a real number r and add every number in the ∂^1 number line to it. This will give you an infinitely small copy of the real-number line around r. Let us call this the ∂^1 monad around r.
- Now, to every number in the ∂^1 monad around r, add every number in the ∂^2 number line. This will give you an infinitely smaller copy of the real-number line around r and around each multiple of ∂ times a real number; that is, in the monad around r there will be a real number of real-number lines. Now keep doing this indefinitely for ∂^n. You will get more and more.
- Around every real number r there will be a monad of infinitely small copies of the real number line, with infinitely smaller and smaller copies of the real-number line embedded in it.
- The monads around any two distinct real numbers do not overlap.

In other words, on the granular-number line, each real number is separated from every other real number by a vast structure of granulars stretching indefinitely in both directions but never reaching a member of another monad. By adding infinitesimals to a real number, you can at best get within an infinitesimal distance of another real number. That's how small infinitesimals are!

The Huge Numbers

Recall that ∂, as we constructed it, has the form $1/n$, where n is an integer. Recall that we are calling this integer H, for "huge number." H is the reciprocal of ∂. Thus,

$$\partial = 1/\mathrm{H},\ \partial \cdot \mathrm{H} = 1,\ \mathrm{H} = 1/\partial,\ \partial = \mathrm{H}^{-1},\ \text{and}\ \mathrm{H} = \partial^{-1}.$$

Since ∂ is less than all real numbers, H is greater than all real numbers. Just as ∂ is infinitesimal, H is infinite. Since H satisfies the first nine axioms for real numbers, it can enter into all arithmetic operations. Thus, there are numbers H + 1, H − 1, H + 2, H − 2, and so on. There are also 2H, 3H, and so on, as well as $\mathrm{H}^2, \mathrm{H}^3, \ldots, \mathrm{H}^\mathrm{H}$, and even

$$\mathrm{H}^{\mathrm{H}^{\mathrm{H}^{\cdots}}}.$$

Thus there is no bound to the hugeness of the huge numbers. What we get are higher and higher levels of infinity. Similarly, we can have

$$\partial^{\mathrm{H}^{\mathrm{H}^{\cdots}}},$$

which defines infinitely smaller and smaller levels of infinitesimals.

As an integer, H has successors. There is H + 1, H + 2, and so on indefinitely. It also has predecessors: H − 1, H − 2, and so on indefinitely. However, all the granular numbers of the form H − n, where n is a real integer, are still huge. You can't get back down to the real numbers by subtracting a real number from H. Similarly, by performing the successor operation—the successive addition of 1 to a finite integer—you can never count all the way up to H or even H − r, for any real r. That will give you some idea of just how "huge" H is. H and the natural numbers are not connectable via the successor operation. And yet, via the operation of the Basic Metaphor of Infinity, H is a metaphorical integer.

So far we have defined ∂^n in terms of self-multiplication. But the exponential function, which is not self-multiplication (see Case Study 2), naturally extends to the granulars, so that expressions like ∂^π, e^∂, and e^H are well defined. As a result, we can raise ∂ and H to the power of any real or granular number. This gives us a *continuum* of ∂ levels. This is a granular continuum, not just a real continuum; that is, there is a ∂ level—a level of granularity—not just for every real number on the real-number line but also for every granular number on the granular-number line!

Why Infinite Sums Are Not *Equal to Their Limits in Granular Arithmetic*

Within the real numbers alone—without granulars or infinitesimals—infinite sums are *defined* in terms of limits. For example, consider the infinite sum of $9/10^n$—that is, 0.9999. . . . As we saw in Chapter 9, this infinite sum is usually defined—via the Infinite Sums Are Limits metaphor—as *being* its limit:

$$\lim_{n\to\infty} \sum_{i=1}^{n} 9/10^i = 1.$$

Since the limit is conceptualized via the special case of the BMI for limits, this is a metaphorical sum. What the expression above says is that 0.999999. . . = 1.

Now, for any finite series of the form

$$\sum_{i=1}^{n} 9/10^i$$

where there is a last term n, there is a difference between the last term and the limit. That difference is

$$1 - \sum_{i=1}^{n} 9/10^i = 1/10^n.$$

For example, if $n = 5$, then the sum is 0.99999 and the difference between the limit 1 and the sum is 1 – 0.99999, which = 0.00001, or $1/10^5$. If the sum over the real positive integers is taken to infinity, the difference at stage n is of the form $1/10^n$. Taking the limit as n approaches ∞, $1/10^n$ converges to 0. At the limit—the only place where the sum is infinite—there is no difference between the infinite sum and the limit. The infinite sum *equals* the limit.

The granular numbers give us a very different perspective on what an infinite sum is. In the granular numbers, H is an integer larger than all real numbers. Relative to the real numbers, therefore, H is infinite. Indeed, since H is beyond all the real numbers, a sum up to H includes additional terms beyond a mere real-number sum approaching ∞.

To see this, let us look closely at H. As an integer, H occurs in a sequence of integers: . . . , H – 3, H – 2, H – 1, H, H + 1, H + 2, ... H is so huge that subtracting any finite real number from H still yields a number incommensurably larger than any real number. Now consider the unending sequence of natural numbers: 1, 2, 3, ... No matter how big this sequence gets, every member of this sequence must not only be smaller than H but *infinitely* smaller than H. We can see this rather easily. Suppose we simultaneously start counting up from 1 and down from H. We get the following two sequences, each of which is open-ended:

$$1, 2, 3, \ldots, 10^{100}, \ldots \qquad \ldots . H - 10^{100}, \ldots, H - 3, H - 2, H - 1, H$$

No matter how far we go in both directions, no contact can be made. Thus, H is infinitely beyond the infinity of natural numbers.

Since H is a specific granular number, since it is infinite, and since it is an integer, we can create an infinite sum whose *last term* is the specific granular integer H. This is an unusual feature of the granular numbers. Series can have an infinite number of terms, yet also have a *last* term. We could, in principle, have an infinite series with more than H terms—say, H + 1 or H^3—but for our purposes H will suffice to make the series not just infinite but longer than any infinite series whose terms correspond to natural numbers alone. In other words, a sum up to H is not just an infinite sum but it has infinitely more terms than a sum over all the natural numbers!

Now consider the series

$$\sum_{i=1}^{H} 9/10^i.$$

This is the series $9/10 + 9/100 + 9/1000 + \ldots + 9/10^H$. The last term of this series is $9/10^H$. Therefore, the difference between the limit 1 and the sum to H terms is $1/10^H$. This is a specific granular number. It is an infinitesimal. With the granular numbers, then, this infinite series is not only less than its limit but *measurably* less than its limit. It is *exactly* $1/10^H$ less than its limit.

Now consider another series with a limit of 1—namely,

$$\sum_{i=1}^{n} 1/2^i.$$

This is the sum $1/2 + 1/4 + 1/8 + \ldots$, whose limit is 1. At any term n, the difference between the limit and the nth term is $1/2^n$. For example, the sum for $n = 5$ is $1/2 + 1/4 + 1/8 + 1/16 + 1/32$, which equals $1 - 1/32$. Now take the sum of this series up to H.

$$\sum_{i=1}^{H} 1/2^i.$$

The precise difference between the limit 1 and the sum to H terms is $1/2^H$.

Note that $1/10^H$ is smaller than $1/2^H$. Thus the infinite sum (up to H terms) of $9/10^n$ is larger than the infinite sum of $1/2^n$, even though they both have the same limit. The precise difference between the two numbers is the granular number $1/2^H - 1/10^H$.

What limits do is to ignore precise infinitesimal differences—differences that can be calculated with precision in the granular numbers.

The Real Part

Since the real numbers are embedded in the granulars, we can define an operator R that maps each granular number onto its "closest" real number. R maps all the infinitesimals—the granulars whose absolute value is less than any positive real number—onto 0. R maps all the positive huge numbers onto ∞ and all the negative huge numbers onto $-\infty$. We will call R the *real-approximation operator*. From the perspective of the granular numbers, each real number is an extremely gross approximation to an infinity of granular numbers.

In addition, R is defined so that for granulars a and b, $R(a + b) = R(a) + R(b)$ and $R(a \cdot b) = R(a) \cdot R(b)$. In other words, the real approximation to the sum is the sum of the real approximations, and the real approximation to the product is the product of the real approximations. Here is an example of the effect of R. Consider the granular number $37 - 14\partial + 19\partial^3$.

$$\begin{aligned} R(37 - 14\partial + 19\partial^3) &= R(37) - (R(14) \cdot R(\partial)) + (R(19) \cdot R(\partial^3)) \\ &= 37 - (14 \cdot 0) + (19 \cdot 0) \\ &= 37 \end{aligned}$$

Or take H − 19.

$$\begin{aligned} R(\mathrm{H} - 19) &= R(\mathrm{H}) - R(19) \\ &= \infty - 19 \\ &= \text{undefined} \end{aligned}$$

Suppose we apply R to the sums of infinite series up to H. Consider the two sums we just discussed:

- $1 - 1/10^H$ and
- $1 - 1/2^H$

Their real parts are

$$R(1 - 1/10^{\mathrm{H}}) \text{ and } R(1 - 1/2^{\mathrm{H}}),$$

which equal

$$R(1) - R(1/10^{\mathrm{H}}) \text{ and } R(1) - R(1/2^{\mathrm{H}}),$$

which equal

$$1 - 0 \text{ and } 1 - 0,$$

which equal

$$1 \text{ and } 1.$$

The two sums differ from 1 by an infinitesimal. The real-approximation operator ignores that infinitesimal and makes both sums equal to their limits. In other words, the real-approximation operator has the same effect as the notion of a limit. That effect is to ignore precise differences that are visible in the granulars. It also has the same effect as the notion of a limit for huge granulars. $R(\mathrm{H} - 19)$ has the same effect as $\lim_{n \to \infty} (n - 19)$. The former is undefined, as is the latter, which technically has no limit since $n - 19$ grows indefinitely as n approaches ∞.

Incidentally, Robinson's treatment of infinitesimals does not include the granulars and differs in certain significant ways from the discussion given here. In a later section, we will compare the two. For a full treatment of calculus from Robinson's perspective, see H. J. Keisler (1976a, 1976b).

Calculus Becomes Arithmetic

Within the granular numbers, defined merely by the first infinitesimal ∂, one can do all of calculus without the notion of a limit. Calculus simply becomes arithmetic.

The derivative can be defined as follows:

$$f'(x) = R\left(\frac{f(x + \partial) - f(x)}{\partial}\right)$$

Compare this to the definition for limits.

$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

Instead of the variable Δx in the limit definition, we have a fixed constant ∂. Instead of a limit operator, $\lim_{\Delta x \to 0}$, we have a real-approximation operator R that has the effect of ignoring all infinitesimals. The results are always the same. Indeed, for any proof with limits, there is a corresponding proof with ∂'s.

Compare the treatment of $f(x) = x^2$ with limits and with ∂'s. With limits we have

$$\begin{aligned}\lim_{\Delta x \to 0} \frac{(x + \Delta x)^2 - x^2}{\Delta x} &= \lim_{\Delta x \to 0} \frac{x^2 + 2x\Delta x + \Delta x^2 - x^2}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{2x\Delta x + \Delta x^2}{\Delta x} \\ &= \lim_{\Delta x \to 0} 2x + \Delta x\end{aligned}$$

At this point, we have to apply the special case of the BMI for limits and treat Δx, which never literally gets to zero, as if it did. The (metaphorical) limit of $2x + \Delta x$ as Δx approaches zero is $2x$.

The derivation is similar, but a bit simpler, with infinitesimals.

$$\begin{aligned}R\left(\frac{(x + \partial)^2 - x^2}{\partial}\right) &= R\left(\frac{x^2 + 2x\partial + \partial^2 - x^2}{\partial}\right) \\ &= R\left(\frac{2x\partial + \partial^2}{\partial}\right) \\ &= R(2x + \partial) \\ &= R(2x) + R(\partial) \\ &= 2x + 0 \\ &= 2x\end{aligned}$$

As should be clear, one can construct an algorithm to convert any derivation from limit notation to granular notation and vice versa. In the granulars, there are no limits, just arithmetic and the real-approximation operator.

Integrals

The definition of the integral in the granular numbers is correspondingly straightforward:

$$\int_a^b f(x)dx = R\left(\sum_{n=1}^{H(b-a)} f(x_n) \cdot \partial\right)$$

The integral divides up the interval from a to b into subintervals of length ∂. $f(x_n)$ is the height of the curve at the nth division into subintervals. The product $f(x_n) \cdot \partial$ is the area of the rectangle of width ∂ and height $f(x_n)$. The sum is the sum of all the rectangles in the interval from a to b (see Figure 11.3).

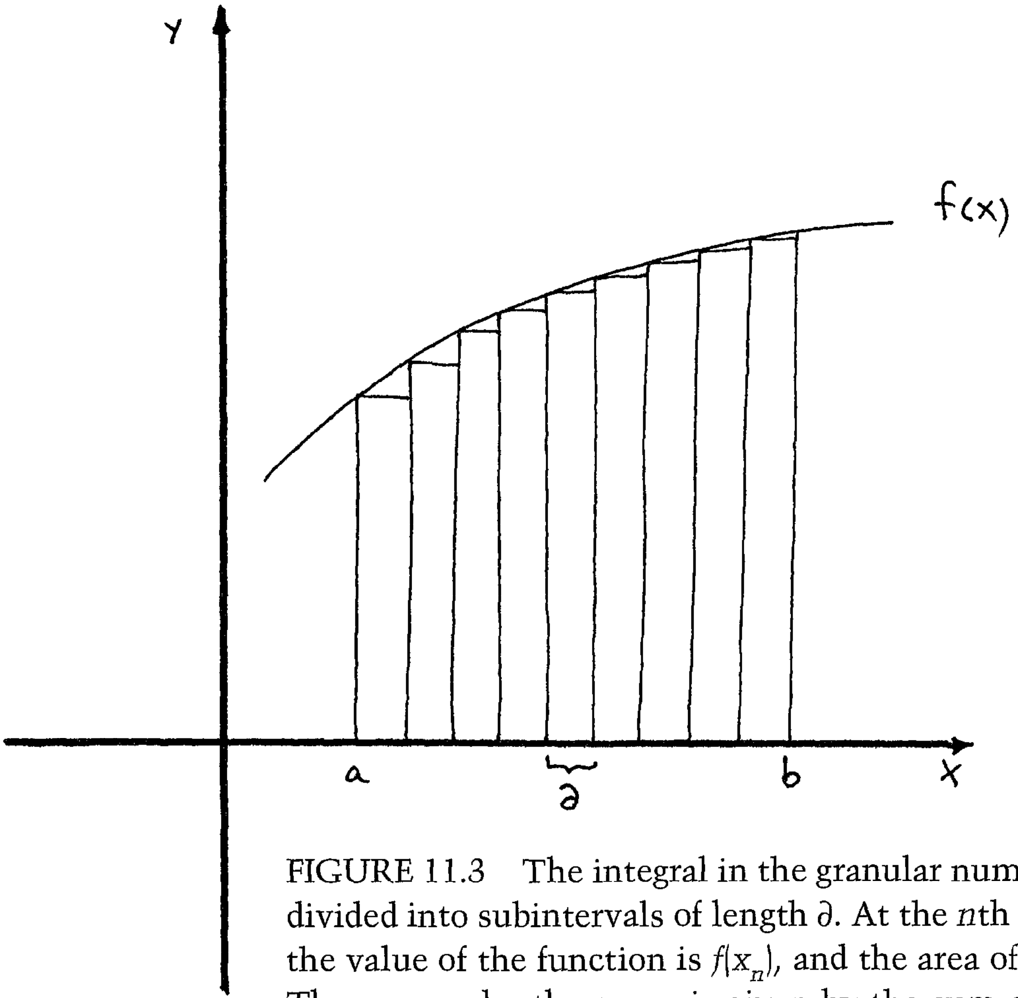


FIGURE 11.3 The integral in the granular numbers: The interval $[a, b]$ is divided into subintervals of length ∂. At the nth division into subintervals the value of the function is $f(x_n)$, and the area of each rectangle is $f(x_n) \cdot \partial$. The area under the curve is given by the sum of the areas of all the rectangles in the interval $[a, b]$.

Since $\partial H = 1$, there are H subintervals for every interval of length 1. The number of such subintervals in the interval of length $(b - a)$ is $H \cdot (b - a)$.

This sum will differ from the area under the curve by an infinitesimal. The operator R eliminates that infinitesimal difference. The result is the same as in standard calculus.

For example, consider the integral of $f(x) = 2x$ from $x = a$ to $x = b$. For the sake of simplicity, suppose that a equals zero and b is positive (see Figure 11.4). It is obvious from the figure what the right answer should be. Since the length b and the height $2b$ form a rectangle of area $2b^2$, the area of the triangle under the curve is half that: b^2. The definition of the integral in terms of an infinite sum of rectangles of infinitesimal width will give exactly the same result. Here's how:

Let us add up the area of the columns of rectangles under the curve in Figure 11.4. Each column has a width of ∂. There are $H \cdot b$ columns. The first column is composed of only one rectangle of area $\partial \cdot 2\partial = 2\partial^2$. The second column is composed of two rectangles of area $2\partial^2$. The total area of the second column is $2 \cdot 2\partial^2$. Similarly, the third column is composed of three rectangles and has the area of $3 \cdot 2\partial^2$. In general, the nth column has the area of $n \cdot 2\partial^2$. And the last

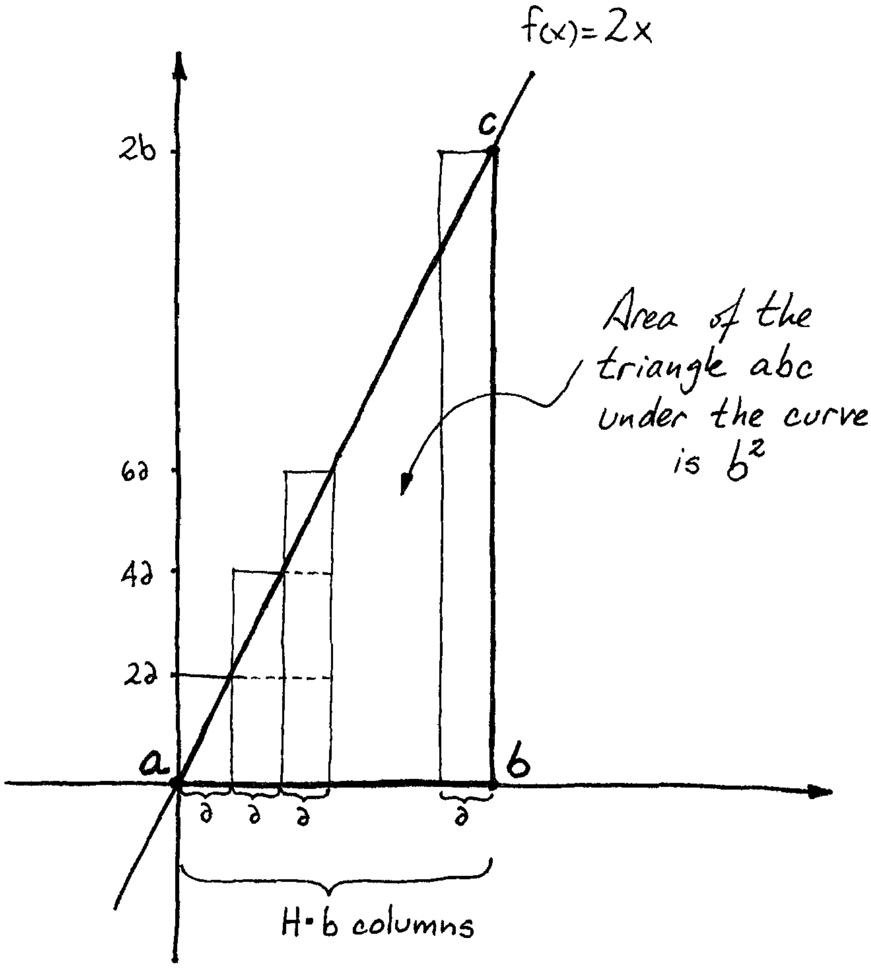


FIGURE 11.4 The integral of $f(x) = 2x$ from a to b in the granular numbers (where $a = 0$ and b is positive). The interval $[a, b]$ is divided into $\mathrm{H} \cdot b$ columns, each with the width ∂. In general, the n^{th} column has the area of $n \cdot 2\partial^2$, and the last column has the area of $\mathrm{H}b \cdot 2\partial^2$. The sum of all columns is $2\partial^2 \cdot (1 + 2 + \ldots + \mathrm{H}b)$, which after a little algebra yields $b^2 + \partial b$.

column has the area of $\mathrm{H}b \cdot 2\partial^2$. The sum of all the columns therefore is $2\partial^2 \cdot (1 + 2 + \ldots + \mathrm{H}b)$. That is,

$$\sum_{n=1}^{\mathrm{H}b} 2\partial^2 \cdot n$$

It is well known that the sum of the first n integers is $\frac{n(n+1)}{2}$. Thus the sum of the areas of the columns is

$$2\partial^2 \cdot \frac{\mathrm{H}b(\mathrm{H}b + 1)}{2}$$

which equals

$$2\partial^2 \cdot \frac{\mathrm{H}^2b^2 + \mathrm{H}b}{2}$$

which equals

$$\partial^2 \, (\mathrm{H}^2b^2 + \mathrm{H}b),$$

which equals

$$\partial^2 \mathrm{H}^2 b^2 + \partial^2 \mathrm{H} b.$$

Since $\partial\mathrm{H} = 1$, we have

$$b^2 + \partial b.$$

Applying the real part operator R, we get

$$R(b^2 + \partial b),$$

which equals

$$R(b^2) + R(\partial b),$$

which equals

$$b^2 + 0 = b^2.$$

This is essentially the same as doing the definite integral via limits. The corresponding limit expression for this integral would be

$$\int_a^b f(x)dx = \lim_{\Delta x \to 0} \sum_{i=1}^{\infty} f(x_i) \cdot \Delta x$$

With granular numbers, there is only one sum and it is computed by simple arithmetic. With limits, there is an infinity of finite sums. The reason is that limits are defined in terms of sequences of sums, with the number of sums approaching infinity.

Remember that in the granular numbers, ∂ and H are fixed numbers whose product is 1. The number of divisions into columns under the curve is a fixed number, $\mathrm{H}(b - a)$. There is just one figure with a sum of column areas to be computed, whereas in the case of limits there is an infinity of such figures and sums, since the number of divisions into columns—namely, n—approaches infinity.

In the granular numbers, the infinitesimal widths and the infinite number of divisions are fully arithmetized. In the case of the limit, there is no such arithmetization. With limits, we do not even know how many elements of width Δx are being summed in each sum, since Δx is a variable, nor do we know the exact width of any Δx.

As Δx gets small, the number of elements being summed gets indefinitely large, but we do not know exactly how large, because there is no arithmetization of infinity. Within the granular numbers, there is precise arithmetization of everything.

The Hyperreals

The system of granular numbers is vast. It dwarfs the real numbers. Half the granulars are infinitesimal and half are huge. Do the infinitesimal granulars and the huge granulars exhaust all the infinitesimal and huge numbers? The answer is no. There are more infinitesimal and huge numbers outside the granulars.

Recall that we formed the granulars in such a way that they constitute the smallest extension of the real numbers to infinitesimals. The way we did this was to apply the Basic Metaphor of Infinity to the process of forming numbers, $1/n$, such that "$1/n$ is a number greater than zero and less than $1/(n-1)$, satisfying the first nine axioms for real numbers." The BMI completes this process by yielding a unique infinitesimal number, which we called ∂. The rest of the granulars arose by arithmetic operations on ∂ and the reals.

Recall also that we had previously characterized the set of all infinitesimals by applying the BMI to a different process—a process of forming *sets,* not individual numbers. The sets were of the form: "the set of all numbers bigger than zero and less than $1/n$, satisfying the first nine axioms for the real numbers." When the BMI completes that process, it yields the set of *all* infinitesimals. The multiplicative inverses of those infinitesimals are included, because the axioms for the real numbers must be satisfied and those axioms require inverses. Those inverses are *all* the huge numbers. This set of *all* infinitesimals and huge numbers is called the *hyperreals*.

It is easy to see that the granulars do not by any means exhaust the hyperreals. The reason is that once we form the first infinitesimal via the BMI, we can use the BMI again to characterize more infinitesimals that are not in the granulars.

Here's how we do it. Recall the following: $\partial = 1/\mathrm{H}$. It is infinitesimal, because it is divided by an infinite integer. $\partial^{\mathrm{H}} = 1/\mathrm{H}^{\mathrm{H}}$. Now, ∂^{H} is infinitely smaller than even ∂ is, because it equals 1 divided by H^{H}, which is infinitely larger than H is. Within the granular numbers, we can go on forming successive powers of powers, such as

$$\partial^{\mathrm{H}^{\mathrm{H}}} \text{ and } \partial^{\mathrm{H}^{\mathrm{H}^{\mathrm{H}}}}$$

We will write these as $\partial^{\mathrm{H}\ (2\text{ levels})}$ and $\partial^{\mathrm{H}\ (3\text{ levels})}$. In general, we can form an arbitrary number of levels of powers, $\partial^{\mathrm{H}\ (n\text{ levels})}$.

We can now form an unending process for creating infinitesimal numbers greater than zero:

> Form $\partial^{\mathrm{H}\ (1\text{ level})}$; raise $\partial^{\mathrm{H}\ (1\text{ level})}$ to the power H, yielding $\partial^{\mathrm{H}\ (2\text{ levels})}$; . . . ; raise $\partial^{\mathrm{H}\ (n-1\text{ level})}$ to the power H, yielding $\partial^{\mathrm{H}\ (n\text{ levels})}$.

We now let this be the unending process in the BMI. The BMI conceptualizes this as a completed process with an endpoint and a new result: $\partial^{H\,(H\text{ levels})}$, an infinitesimal smaller than any granular infinitesimal, yet greater than zero. We will call it the second infinitesimal and write it as $\partial^{H\,(H\text{ levels})} = {}^{2}\partial$, or "delta super-two".

Using ${}^{2}\partial$, we can now form another layer of granular numbers, with infinitesimals smaller than any of the first granulars and huge numbers larger than any in the first layer of granulars. Once we have the layer of granulars defined by ${}^{2}\partial$, we can use the BMI in the same way again to form ${}^{3}\partial$, and so on indefinitely. Every layer defined by a ${}^{n}\partial$ is a subset of the hyperreals, which is the set of *all* infinitesimals and huge numbers.

Each time we get another ${}^{n}\partial$—the nth infinitesimal number—we can create a numeral for it—say, "${}^{n}\partial$". We started with the numerals "0," "1," "2," "3," "4," "5," "6," "7," "8," "9" for the real numbers and added "∂" to give a numeral system that could symbolically represent every granular number. We then added "${}^{2}\partial$," "${}^{3}\partial$," "${}^{4}\partial$," . . . , "${}^{n}\partial$" to give a finite numeral system in which every number in the first n layers of granulars could be symbolically represented. But there is no end to the ${}^{n}\partial$'s, and so there is no end to the numerals that would have to be created to represent these numbers in a system of numerals. To represent all the layers of granulars in numerals would require more than a finite number of symbols. Thus, it is impossible to design a numeral system to symbolically represent all infinitesimals and huge numbers.

Recall that we can do calculus just fine with only one layer of granulars, so only one added numeral—"∂"—suffices. But if for theoretical reasons we wanted to have a numeral system for all the infinitesimals, we would not be able to achieve that goal.

Beyond All Granulars

The formation of more and more ${}^{n}\partial$'s via the BMI is an unending process itself. We can now apply the BMI to that unending process. That is, we can apply the BMI recursively. Here is the unending process:

From ${}^{n-1}\partial$, apply the BMI to yield ${}^{n}\partial$.

Applying the BMI to this unending process yields a metaphorical completion of the process. The result is a new infinitesimal number smaller than any of the ${}^{n}\partial$'s—a superinfinitesimal: ${}^{H}\partial$. This, too, is of the form $1/k$ (where k is an integer). It is smaller than any number in any layer of granulars, satisfies the first nine axioms for the real numbers, combines freely with all reals and granulars in every layer, and produces a new set of numbers that are linearly ordered with respect to all the other granulars.

We can now begin the process all over again, producing another infinity of layers, applying the BMI recursively to that, and on and on endlessly. Each time we use the BMI, we create another layer of infinitesimal and huge numbers that is a subset of the hyperreals.

This gives us a sense of just how big the hyperreals are and what kind of substructures exists within the hyperreals. Because *all* the hyperreals can be linearly ordered (since they satisfy the first nine axioms for the real numbers), they can all be mapped onto the line via the metaphor Hyperreal Numbers Are Points on a Line. Notice that the reals are such a small subset of the hyperreals that they are barely noticeable. Virtually the whole line is taken up by nonreals.

Calculus in the Hyperreals

The field known as nonstandard analysis does calculus using the hyperreals, not the granulars. Because there can be no numeral system for the hyperreals, there can be no finite symbol system for the arithmetic of the hyperreals. Where we used arithmetic and arithmetic notation to do calculus, those scholars using the hyperreals must take another approach. Their approach is to use variables over the infinitesimals and huge numbers (which they call *infinite numbers*).

In H. J. Keisler's calculus text based on Robinson's treatment (Keisler, 1976a, 1976b), Keisler uses ε and δ as symbols for *variables* over the infinitesimals in the hyperreals, and H and K as *variables* over the huge, or "infinite," numbers. (As variables, they are symbolized in italics.) Since nonstandard analysis does not recognize the existence of the granular numbers, it cannot name individual infinitesimals and huge numbers at the granular level. Moreover, it cannot have statements in granular arithmetic like $H \cdot \partial = 1$, which implies that $H^2 \cdot \partial = H$ and $H \cdot \partial^2 = \partial$. For Keisler, the product $K \cdot \varepsilon$ of a variable K over an infinite number and a variable ε over the infinitesimals is "indeterminate." The reason this has to be true for Keisler, as for Robinson, is clear. The K is a variable and so, in granular terms, could be varying over any huge number at all (e.g., H or H^2), whereas the ε could be varying over any infinitesimal at all (e.g., granular ∂ or ∂^2). Thus, Keisler's "$K \cdot \varepsilon$" could be varying over the granular numbers $\mathrm{H} \cdot \partial$, $\mathrm{H}^2 \cdot \partial$, or $\mathrm{H} \cdot \partial^2$. The first is 1, the second is H, and the third is ∂ in the granulars. Keisler, who cannot "see" the granulars, can only conclude that the result in the hyperreals is indeterminate—not clearly finite, infinite, or infinitesimal.

We can now begin to see the conceptual disadvantages of using the hyperreals without the granulars. Imprecision results. The numbers cannot be named. The only way to even discuss the product of an infinite and an infinitesimal number in hyperreals is with variables, but the result can only be called indeterminate.

There are further problems with the way Robinson happened to develop his conception of the hyperreals. He made a three-way category distinction between "infinitesimal," "finite," and "infinite" numbers. He also made a distinction between the "standard part," analogous to what we called the real-approximation operator, and the "nonstandard part" of the hyperreal numbers. The "infinitesimals" have an absolute value smaller than any positive real number. This makes zero an "infinitesimal" for Robinson and Keisler. In their development of calculus, they must distinguish between zero and nonzero infinitesimals so that there is never division by zero. In the granular numbers, we keep zero separate from the infinitesimal granulars, for reasons that will become clear in the next chapter.

Here's how Robinson and Keisler characterize differentiation using variables ranging over nonzero infinitesimals. The derivative is defined as follows, using "st" for the standard-part operator:

$$f'(x) = \mathrm{st}\left(\frac{f(x + \Delta x) - f(x)}{\Delta x}\right)$$

Here, the Δx's are variables over nonzero infinitesimals. From here on, the differential calculus is just like it is for the granular numbers. And as with the granular numbers, there is a correlate for every proof in standard differential calculus. Keisler defines the definite integral—the area under a curve between two points—using variables over infinitesimals and infinite numbers. Given the interval on the x-axis $[b, a]$ with length $b - a$, that length is divided into H subintervals, each of length δ, but where H is some unspecified "hyperinteger" and δ is some unspecified infinitesimal. In Keisler's text, the result is a set of H rectangles, each of width δ and height $f(x)$. Then the definite integral is defined as the standard part of the sum of the areas of that rectangle. In short, the strategy is again to use variables over hyperreals rather than specific hyperreal numbers. The results, again, are exactly the same as in granular calculus and standard calculus.

A Comparison

Calculus can be done equally well either the standard way with limits, or using granulars with a real-approximation operator, or using hyperreals with a standard-part operator. From our cognitive perspective, the granulars have certain advantages over the other two. Compared with real numbers and limits, the granulars allow us additionally to see that infinite sums are not equal to their limits, unless the real-approximation operator applies. This allows us to better understand the *nature* of infinite sums. The granulars also allow us to precisely calculate infinite sums with the same limit and compare their size. Part of this job can be done with the hyperreals. In the hyperreals, you could show that there is a difference between the limit and the sum to H terms, where H is not a fixed number but a variable over hyperreal integers. But you cannot precisely calculate the exact infinite sum as a specific number, nor can you calculate the difference between the infinite sums as a specific number.

Compared with the hyperreals, the granulars allow us to use a numeral system and to do calculus in terms of the arithmetic of particular granular numbers. Using the granulars, every expression in calculus is an arithmetic expression with particular numbers, rather than an algebraic expression with variables. Moreover, the granulars have a conceptual advantage: The precise structure of the number system you are using is well defined and revealed through the notation.

The Mathematical Importance of Ignoring Differences

The study of infinitesimals teaches us something extremely deep and important about mathematics—namely, that *ignoring certain differences is absolutely vital to mathematics!* This idea goes against the view of mathematics as the supreme exact science, that science where precision is absolute and differences, no matter how small, should never be ignored.

Whether we are using real numbers and limits, the granular numbers with the real-approximation operator, or the hyperreals with the standard-part operator, *calculus is defined by ignoring infinitely small differences.* This has not always been an uncontroversial view. In the eighteenth century, the Irish philosopher and bishop George Berkeley, for example, rejected calculus as a form of mathematics on the grounds that mathematics required *exact* precision and calculus ignored infinitesimals.

Berkeley's argument is revealing. As Davis and Hersh (1981, p. 244) observe, discussing Berkeley,

> After all, dt is either equal to zero, or not equal to zero. If dt is not zero, then $32 + 16dt$ is not the same as 32. If dt is zero, then the increment in distance ds is also zero, and the fraction ds/dt is not $32 + 16dt$ but a meaningless expression $0/0$.

Berkeley's own comment (in *The Analyst*, 1734) was as follows:

> For when it is said, let the increments vanish, i.e., let the increments be nothing, or let there be no increments, the former supposition that the increments were something, or that there were increments, is destroyed, and yet a consequence of that supposition, i.e., an expression got by virtue thereof, is retained. Which is a false way of reasoning.

Speaking of Newton's "fluxions," or derivatives, Berkeley wrote,

> What are these fluxions? The velocities of evanescent increments. And what are these same evanescent increments? They are neither finite quantities, nor quantities infinitely small, nor yet nothing. May we not call them the ghosts of departed quantities?

Bishop Berkeley was being a literalist, as usual. As we have seen, there is always a metaphor involved in differentiation—which we have called the BMI—whether in the form of limits or of infinitesimal numbers. Literally—that is, without accepting any use of the BMI—Berkeley was right. Differentiation *does* require ignoring certain differences—infinitely small differences. Using limits to define infinite sums also involves ignoring infinitely small differences. Thus, accepting 0.9999... = 1 also requires ignoring infinitely small differences. If you think the very definition of mathematics requires absolute precision—never ignoring differences, no matter how small—then you, too, with Berkeley, should reject calculus as a form of mathematics.

But calculus is central to mathematics, as is the study of infinite series, as are dozens of uses of the BMI. Let us think for a moment about what would be lost to mathematics if infinitesimal differences were *not* ignored, at least in certain well-defined situations.

We can see this most clearly with the granular numbers. Suppose the derivative—the rate of change of a function—were still defined by the metaphor that Instantaneous Change Is Average Change Over an Infinitely Small Interval. But now suppose that instead of taking the real approximation of the resulting expression, we leave in the ∂'s. Let us call this the cumulative derivative "$f!$", in contrast to the normal derivative f'.

NORMAL DERIVATIVE f' $f'(x) = R\left(\frac{f(x+\partial) - f(x)}{\partial}\right)$	CUMULATIVE DERIVATIVE $f!$ $f!(x) = \frac{f(x+\partial) - f(x)}{\partial}$
$f'(x^2) = 2x$	$f!(x^2) = 2x + \partial$
$f'(x^3) = 3x^2$	$f!(x^3) = 3x^2 + 3x\partial + \partial^2$
$f'(\sin(x)) = \cos(x)$	$f!(\sin(x)) = \frac{\sin(x+\partial) - \sin(x)}{\partial}$
$f'(e^x) = e^x$	$f!(e^x) = \frac{e^{(x+\partial)} - e^x}{\partial}$

First, with cumulative derivatives of polynomials, the ∂ expressions accumulate. Second, there is no function that is its own cumulative derivative, since the ∂ terms keep accumulating. For example, e^x would no longer be a function that is its own derivative. Indeed, there would be no such function! Third, consider periodic functions like sine and cosine. Whereas the derivative of $\sin(x) = \cos(x)$, the *cumulative* derivative of $\sin(x) \neq \cos(x)$, but equals a much longer expression. Moreover, while the second derivative of $\sin(x) = -\sin(x)$, the second *cumulative* derivative of $\sin(x) \neq -\sin(x)$, but equals another very long expression. And the fourth *cumulative* derivative of $\sin(x)$ is not $\sin(x)$ itself but an extremely long expression. Thus, the values of sine and cosine do not repeat after four derivatives but become increasingly complex.

If there were only cumulative derivatives, a great many of the beauties of classical mathematics would cease to exist, as would the usefulness of classical calculus in actual computation. In short, ignoring infinitesimal differences of the right kind in the right place is part of what makes mathematics what it is.

The Closure Engine

Calculus is ubiquitous in advanced mathematics curricula. But calculus with infinitesimals is barely taught anywhere, despite the fact that calculus was first formulated with infinitesimals and infinitesimals were used for doing calculus for two hundred years. Imaginary and complex numbers are a commonplace in high school mathematics classes. Infinitesimal numbers are not. Hardly anyone has ever heard of them. Why?

The main impetus for extending our number system over centuries has come from the demands of closure (see Chapter 4). Zero and negative numbers were needed for closure under addition, rational numbers for closure under multiplication, the real numbers for closure under self-multiplication, and the complex numbers for closure with the square roots of negative numbers. In each case, closure was needed to solve basic polynomial equations: $x + 1 = 1$; $x + 5 = 3$; $7x = 2$; $x^2 = 2$: $x^2 + 1 = 0$.

The infinitesimal numbers are not needed to solve basic polynomial equations. They are not necessitated by the drive for closure. Indeed, they are not needed for any practical mathematical purpose such as solving equations and doing calculations. Moreover, they have a certain stigma. Given the construction of natural numbers by the successor operation—the successive addition of 1 starting with 1—you cannot reach the infinite integers, which are inverses of certain infinitesimals. This is counterintuitive for most people. How can there be an "integer" that you can't count up to? If you don't admit the existence of the infinitesimals, you don't have to face this problem.

What about the granulars? We invented the granulars by applying the BMI as shown. In our search through the literature, we could not find any mention of a number system generated by a "first infinitesimal" and fully symbolizable by the addition of one numeral, in which all of calculus could be done as arithmetic. Why not? Certainly the remarkable mathematicians who have worked on the hyperreals could have developed such a number system easily. Why didn't they?

We believe that the reason has to do with the formal Foundations of mathematics movement and its values (see Chapter 16). Technically, the hyperreals arise within the confines of formal foundations because of an important technical property of model theory: namely, the existence of nonstandard models—models that happen to satisfy a set of axioms even though you didn't intend them to. Henkin found in the late 1940s that there was a model containing *all* the hyperreals, which unexpectedly satisfied the first nine axioms for the real numbers.

But suppose you don't want a model containing *all* the hyperreals. Suppose you want a model containing the granulars and you want to use the techniques of model theory to get it. Your model would have to contain a particular entity—the "first infinitesimal," which we have called ∂. The way that mathematicians generate such particular entities is via additional axioms. But what was attractive about the hyperreals was that you could get *all* infinitesimals without adding any extra axioms. Why add extra axioms to get some infinitesimals when you can get them all for free?

Nonstandard models are, of course, things that classical mathematicians tend not to want to talk about, and calculus has been one of the most classical of the classical subject matters of mathematics.

These are some of the reasons that infinitesimals are generally not taught. We will discuss others in the next chapter.

Two Types of Infinite Numbers

Cantor's transfinite cardinals and ordinals are infinite numbers. So are the huge numbers in the granulars and the infinite numbers in the hyperreals. Do they have anything to do with each other? They appear to be two different types of infinite numbers. What does it mean for there to be two different types of infinite numbers?

The answer is quite interesting: The different types of infinite numbers differ in their conceptual structures. The transfinite numbers have a conceptual structure that is imparted by

- the BMI applied to the formation of sets of natural numbers to form the set of all natural numbers,
- Cantor's metaphor, and
- the results that power sets have "more" members (in Cantor's sense) than the sets they are formed from.

The huge granulars and the infinite hyperreals have structure imparted to them by first using the BMI to form infinitesimal numbers and then taking the inverses of the infinitesimals—which are guaranteed to exist, since the infinitesimals must satisfy the first nine axioms for the real numbers.

Conceptually, these are two utterly different structures, leading to two utterly different notions of "infinite numbers." It is not surprising that Cantor did not believe in infinitesimals. After all, the infinitesimals provided a different notion of infinite number than his transfinite numbers—one that characterized infinity and degrees of infinity in a completely different fashion.

How can there be two different conceptions of "infinite number," both valid in mathematics? By the use of different conceptual metaphors, of course—in this case, different versions of the BMI.

Some Contributions of Mathematical Idea Analysis to the Discussion of Infinitesimals

As we observed, Abraham Robinson characterized infinitesimals in formal mathematical terms. What is different in the mathematical idea analysis of infinitesimals?

First, it makes the whole idea much less mysterious. Mathematical idea analysis shows that the concept is grounded in our everyday experience with specks. And via the use of the BMI, it shows how the infinitely small is linked to the infinitely large. Moreover, it shows explicitly how huge numbers differ from transfinite numbers. It provides a characterization of the granular numbers with an explicit notation that turns calculus into simple arithmetic. Finally, the analysis shows how ordinary human ideas give rise naturally to such an apparently abstruse notion.

Part IV

Banning Space and Motion: The Discretization Program That Shaped Modern Mathematics

12

Points and the Continuum

You might think that points, lines, and space are simple concepts. They aren't. There are no deeper concepts in mathematics. The complexity of these concepts are reflected in three results that we will discuss in the course of this chapter.

1. "The real-line" is not a line.
2. "Space-filling curves" do not fill space.
3. "The Continuum hypothesis" is not about the continuum.

These results are not in any way paradoxical or contradictory. The expressions in quotes all refer to concepts that are characterized using conceptual metaphors from within mathematics. Those concepts are at odds with our ordinary concepts of curves, lines, and space as expressed by the language not in quotes. There is nothing surprising about this; any technical discipline develops metaphors that are not in the everyday conceptual system. In order to teach mathematics, one must teach the difference between everyday concepts and technical concepts, making clear the metaphorical nature of the technical concepts.

This chapter is about the role of the Basic Metaphor of Infinity in the modern conceptualization of space, lines, and points in mathematics. The BMI, as we shall see, is the crucial link between discrete mathematics (set theory, logic, arithmetic, algebra, and so on) and "continuous" mathematics (geometry, topology, analysis, and so on). This relationship is profound and anything but obvious. Indeed, it is the source of a great deal of confusion not only among students and the lay mathematical public, but in the philosophy and epistemology of mathematics itself. We will try, from the perspective of cognitive science, to sort out those conceptual confusions as we look into how the BMI has permitted a reconceptualization of the continuous in terms of the discrete.

Two Conceptions of Space

Space has been conceptualized in two very different ways in the history of mathematics. Prior to the mid-nineteenth century, space was conceptualized as most people normally think of it—namely, as naturally continuous. Here is how we all think about space in everyday life.

NATURALLY CONTINUOUS SPACE

> Space is absolutely continuous. Space does not consist of objects. Rather, it is the background setting that objects are located *in*. Space exists independently of, and prior to, any objects located in space. Planes, too, are absolutely continuous. They, too, are not made up of objects but have locations on them where objects can be situated. Similarly, a line or curve is absolutely continuous, like the path traced by a moving point. Lines and planes also exist independently of, and prior to, any objects located on them.
>
> Points are locations in space, on lines, or on planes. They are not objects that can exist independently of the line, plane, or space where they are located. Dimensionality is a property of a space, a plane, or a line.

Why Naturally Continuous Space Became a "Problem" for Mathematics

Descartes's invention of analytic geometry changed mathematics forever. His central metaphor, Numbers Are Points on a Line (see Case Study 1), allowed one to conceptualize arithmetic and algebra in geometric terms and to visualize functions and algebraic equations in spatial terms. The conceptual blend of the source and target domains of this metaphor lets us move back and forth conceptually between numbers and points on a line, and between n-tuples of numbers and points in n-dimensional spaces. Just as it lets us visualize functions as curves, it lets us conceptualize higher-dimensional spaces in terms of n-tuples of numbers. Ultimately, it led to a precise calculus of change—especially of movement and acceleration—which began with a geometric visualization of change in spatial terms (tangents of curves) and resulted in calculus, the arithmetization of the idea of change.

Two mental constructs were crucial in this development: the association of discrete numbers with discrete points, and the association of discrete symbols with discrete numbers. Arithmetization and symbolization allowed for precise calculation and the constructions of algorithms for differentiation and integration. The study of change became arithmetizable and therefore formalizable. Calculus became as precise and calculable as arithmetic.

What made this possible was Descartes's Numbers-As-Points metaphor, which matched discrete numbers (having discrete symbols) to discrete points on lines and in space. Thus began what we will call the program of *discretization*. Analytic geometry and calculus began the process of reconceptualizing naturally continuous space and naturally continuous change in terms of the discrete: points, numbers, symbols, and algorithmic rules for calculation.

As we shall see, much of modern mathematics has been defined as the progressive discretization of mathematics.

THE DISCRETIZATION PROGRAM SINCE THE LATE NINETEENTH CENTURY

1. The arithmetization program, which sought to fully arithmetize calculus to eliminate any notion of naturally continuous space.
2. The formalization program, which sought to reconceptualize mathematics as the manipulation of discrete symbols.
3. Symbolic logic, which sought to discretize reason itself, using discrete symbols and precisely formulated algorithms that employed only discrete symbols.
4. Logicism, which sought to discretize all of mathematics through the claim that mathematics could be "reduced" to symbolic logic and the theory of sets, where sets and members of sets are both discrete.
5. Point-set topology, which sought to reconceptualize all understanding of naturally continuous space in terms of discrete points, sets of points, and discrete, symbolizable operations on sets of points.

What propelled the discretization program was the success of analytic geometry and calculus and the idea that anything not formalizable was "vague," "intuitive" (as opposed to "rigorous"), and imprecise. In late-nineteenth-century Europe, mathematics had gained an important stature: the discipline that defined the highest form of reason, with precise, rigorous, and indisputable methods of proof. The discretization program was seen as crucial to preserving that stature for mathematics.

Even so-called intuitionists and constructivists were part of the discretization program. Indeed, they were even more radical discretizers, since they insisted on "constructive" methods alone: algorithmic proofs that were "direct" (disallowing reductio ad absurdum proofs, which use a negative of a negative to "prove" a positive) and finite algorithms (rather than those using what we have called the BMI). This was a further discretizing, in the direction of an even more constrained formalization and an even narrower concept of "rigor."

The discretization program not only made sense in social and historical terms, but it made important mathematical sense as well. It could be seen as asking: What are the limits of discretization? How far can one go in discretizing naturally continuous space? How many naturally continuous concepts (like change or likelihood) can be reasonably discretized and brought into the realm of discrete algorithmic processes? After Gödel, this became the question "What is computable by discrete algorithmic processes?"—a vitally important question in the age of digital computers.

Conceptualizing the Naturally Continuous in Terms of the Discrete

From the perspective of cognitive science, the discretization program is fraught with difficulty from the outset. The continuous and the discrete are conceptual opposites. To conceptualize the continuous—naturally continuous space, motion, and change—in terms of its opposite, the discrete, is at the very least a formidable metaphorical enterprise. It is at worst conceptually impossible, and at best extremely challenging. What we, as cognitive scientists, have been particularly impressed by is how far the mathematical community has come with this program. As we shall see, the discretization program has not been completed and may not be completable. But the progress it has made is stunning. Through the remarkably creative use of conceptual metaphors, a huge amount of new discretized mathematics has been created. Our task in this chapter and the next two will be to give some detailed initial understanding of what the cognitive processes implicit in the program have been, what the triumphs are, and what has been left out or left undone, possibly because it is undoable.

Carrying Out the Discretization Program

There is a central metaphor at the heart of the discretization program:

A SPACE IS A SET OF POINTS

Source Domain A SET WITH ELEMENTS		*Target Domain* NATURALLY CONTINUOUS SPACE WITH POINT-LOCATIONS
A set	→	An n-dimensional space—for example, a line, a plane, a 3-dimensional space
Elements are members of the set.	→	Points are locations in the space.
Members exist independently of the sets they are members of.	→	Point-locations are inherent to the space they are located in.
Two set members are distinct if they are different entities.	→	Two point-locations are distinct if they are different locations.
Relations among members of the set	→	Properties of space

This is a radical departure from the commonplace concepts of lines, planes, and space. Here is a description of it.

THE SET-OF-POINTS CONCEPTION OF SPACE

A space is just a set of elements with certain relations holding among the elements. There is nothing inherently spatial about a "space." What are called "points" are just elements of the set of any sort. They are discrete entities, distinct from one another.

Like any members of sets, the points exist independently of any sets they are in. Spaces, planes, and lines—*being* sets—do not exist independently of the points that constitute them.

A line is a set of points with certain relations holding among the points. A plane is a set of points with other relations holding among the points. A geometrical figure, like a circle or a triangle, is a subset of the points in a space, with certain relations among the points. There is thus nothing inherently spatial about a circle or triangle. For example, a circle is just a subset of the elements of the space with certain relations to one another.

What we ordinarily understand as the properties of spaces and figures are characterized, according to this metaphor, as relations among elements in the set, or as functions that assign numbers to elements or n-tuples of elements. For example, the dimension of a space is a number assigned by a "dimension function" to a set of elements called "points." A distance between "points" is a number assigned by a function to pairs of set elements. The definitions of "dimension" and "length" are given by formal statements characterizing these functions.

What is the curvature of a line or curve at a point? If the curve is a set of points, then the degree of "curvature" at each point is a number assigned by a function to each point in the set. For example, each point in a circle has the same curvature assigned to it—namely, the number equal to $1/r$, where r is the radius. A "straight" line is a set of points meeting the axioms for a line, where the curvature function assigns the number 0 to each point.

Since a "point in space" is metaphorically conceptualized as just an element of a set, any kind of mathematical entity that can be a member of a set can be seen as a "point," provided that an abstract "distance function" can be defined over the set. For example, there are "function spaces" in which a set of functions is seen as a set of "points" constituting a "space" with a distance "metric"—a function assigning a nonnegative real number to each pair of "points" (i.e., functions).

Given the metaphor A Space Is a Set of Points, "points" are not necessarily spatial in nature but can be any kind of mathematical entities at all. Spaces, lines, and planes are therefore not inherently spatial in nature. They are sets of elements that meet certain axioms and to which certain functions apply, like those assigning numbers indicating metaphorical "dimensionality" and "curvature," which are also formal mathematical notions and not conceptually spatial notions.

To see what this means, consider what a circle is, given these metaphors. A "circle" is a set of "points" in a "plane" that are all at the same distance from a single point C called the "center." That is, for every point P in the set of points constituting a "circle," there is a "distance" function that maps the ordered pairs of points (C, P) onto a single real number, called the "radius." The "points" constituting the "circle," the "center," and the "plane" need not be spatial at all; they can be any entities whatever, provided they bear the appropriate relations.

Comparing the Two Conceptions of Space

The Space-As-Set-of-Points metaphor, which is ubiquitous in contemporary mathematics (but which did not exist two centuries ago) provides a conception of points, lines, planes, and space that is quite different from our ordinary one and anything but obvious to unsophisticated students of mathematics. Indeed, the two conceptions are inconsistent. In one, spaces, lines, and planes *exist independently of* points, while in the other they *are constituted by* points. In one, properties are inherent; in the other they are assigned by relations and functions. In one, the entities are inherently spatial in nature; in the other, they are not.

The first—naturally continuous space—is our normal conceptualization, one we cannot avoid. It arises because we have a body and a brain and we function in the everyday world. It is unconscious and automatic. The second has been consciously constructed to suit certain purposes. It is a reconceptualization of the first, via conceptual metaphor.

Even professional mathematicians think using the naturally continuous concept of space when they are functioning in their everyday lives and communicating with nonmathematicians. It takes special training to think in terms of the Set-of-Points metaphor. Moreover, one must learn which kinds of mathematical problems require which metaphors.

The set-of-points conception is the one taken for granted throughout contemporary mathematics. When you are thinking about mathematics, you are dealing simultaneously with two utterly different conceptualizations of our most basic geometrical concepts. It is crucial to keep them straight. Failing to do so may lead to apparent paradoxes. Becoming a professional mathematician requires learning how to operate in this dual fashion.

Indeed, one of the most paradoxical-sounding concepts in twentieth-century mathematics is a result of just such a confusion: the concept of so-called *space-filling curves*. We shall see shortly that such "curves" do not "fill space" at all. But in order to see exactly why, we will have to look at what a point is and how points are related to lines, planes, and numbers.

How Is a Point Conceptualized?

To understand how modern mathematics differs from classical mathematics, we must understand how points, lines, and spaces are conceptualized in the discretization program, where spaces are conceptualized as *being* sets of points. Let us begin with the question of how points are commonly conceptualized, and later turn to the question of what conceptualization is needed to understand discretized mathematics.

Euclid defined a surface as "that which has length and breadth only," a line as "breadthless length," and a point as "that which has no part." Euclid used the ordinary concept of a *lack*: A surface lacks thickness, a line lacks breadth and thickness, and a point (which is made up of no parts) lacks all of these.

In modern mathematics, the lack of a feature is conceptualized metaphorically as the presence of that feature with value zero. Thus, a lack of length is conceptualized as a *particular length*—zero. From the perspective of the modern discretized mathematics, a "point" is an abstract object. Three functions—length-, width-, and thickness-functions—might each assign the number 0 to the "point." This is one way to render Euclid's idea in terms of the Properties Are Functions metaphor.

How are we to understand a point as an object with length, width, and depth equal to zero? Or, to make it simpler, a point in a plane as an object with length and width equal to zero? A common way to conceptualize points is to start with a blob, or a disc, and then make it smaller and smaller and smaller until it is as small as it can get. That's how most people learn to think about what a point is. In other words, a point is infinitely small. And to conceptualize *infinite* smallness, one needs to make use of the Basic Metaphor of Infinity (see Chapter 8).

Start with a disc of some unit radius 1. Keep shrinking the size of the disc to half its previous diameter. The result is an indefinitely large set of discs:

Shrink the disc from diameter 1 to diameter $1/2^1$.
Shrink the disc from diameter $1/2^1$ to diameter $1/2^2$.
$\vdots$
Shrink the disc from diameter $1/2^{n-1}$ to diameter $1/2^n$.

This is an unending process. It produces an infinite sequence of discs, each smaller than the previous one. Let it be the unending process in the BMI. The BMI supplies a metaphorical completion to the process and a metaphorical final result: a "disc" of diameter "0." Such a "disc" is a point in the plane. Of course, instead of nested discs in the plane, we could have used nested intervals on the line (see Chapter 9).

This is a common way to conceptualize a point, and every such conceptualization uses some such version of the BMI to create an infinitely small entity. But this application of the BMI produces a conceptual problem. Here is the frame semantics for a disc.

THE FRAME FOR A DISC

Roles: Center, Circumference, Interior, Diameter, where Center ≠ Circumference ≠ Interior
Parts: Interior, Center, Circumference
Constraints: (a) Center is in Interior. (b) Distance from Center to the Circumference is the same for all points of the Circumference.

Conceptually, a disc has the following parts: a center, a circumference, and an interior, with the center in the interior and distinct from the circumference—

and both center and circumference distinct from the interior. For this to be true, the diameter has to be larger than zero. If the diameter is zero, the center, interior, and circumference are no longer distinct. They collapse into the same entity, violating the frame semantics for what a disc is. Literally, a disc cannot be a disc and have diameter zero.

But one of the interesting things about human conceptual systems is that we can form metaphorical blends (as discussed in Chapter 2). That is how we typically use the BMI applied to discs. We form a blend of the Disc frame and a Line-Segment frame (see Figure 12.1). At each stage of the BMI, we have a pair of frames (Disc, Line-Segment), where the length in the Line-Segment frame is set identical to the diameter in the Disc frame.

This linking of the two frames forms a conceptual blend. In the BMI, the length in the Line-Segment frame gets shorter and shorter, resulting in the diameter of the disc getting smaller and smaller. At the final resultant state of the BMI, the length in the Line-Segment frame is zero. At this final stage, there are still a pair of frames (Disc, Line-Segment), with the diameter in the Disc frame still set equal to the length in the Line-Segment frame. That is, we are conceptualizing a "disc" with zero diameter, even though, if we consciously put together all that we know, the constraints of the Disc frame are violated. However, as human beings, we do not always put together all that we know; we can focus our attention separately on the Disc frame (a point is a disc) and the Line-Segment frame (the length is zero). What is logically a contradiction may not be recognized by a functioning human being unless attention is focused on both at once.

Notice that the following table has two-headed arrows. This is to differentiate the tables for blends from the tables we have used to characterize conceptual metaphors which have unidirectional arrows.

THE DISC/LINE-SEGMENT BLEND

Element 1 THE DISC FRAME		*Element 2* THE LINE-SEGMENT FRAME
A disc, with roles: center, circumference, interior, and diameter, where center ≠ circumference ≠ interior	↔	A line segment, with roles: endpoint *A*, endpoint *B*, center, interior, and length, where endpoint *A* ≠ endpoint *B* ≠ center ≠ interior
Diameter	↔	Length
Center	↔	Center
Opposite points on circumference	↔	Endpoints *A* and *B*

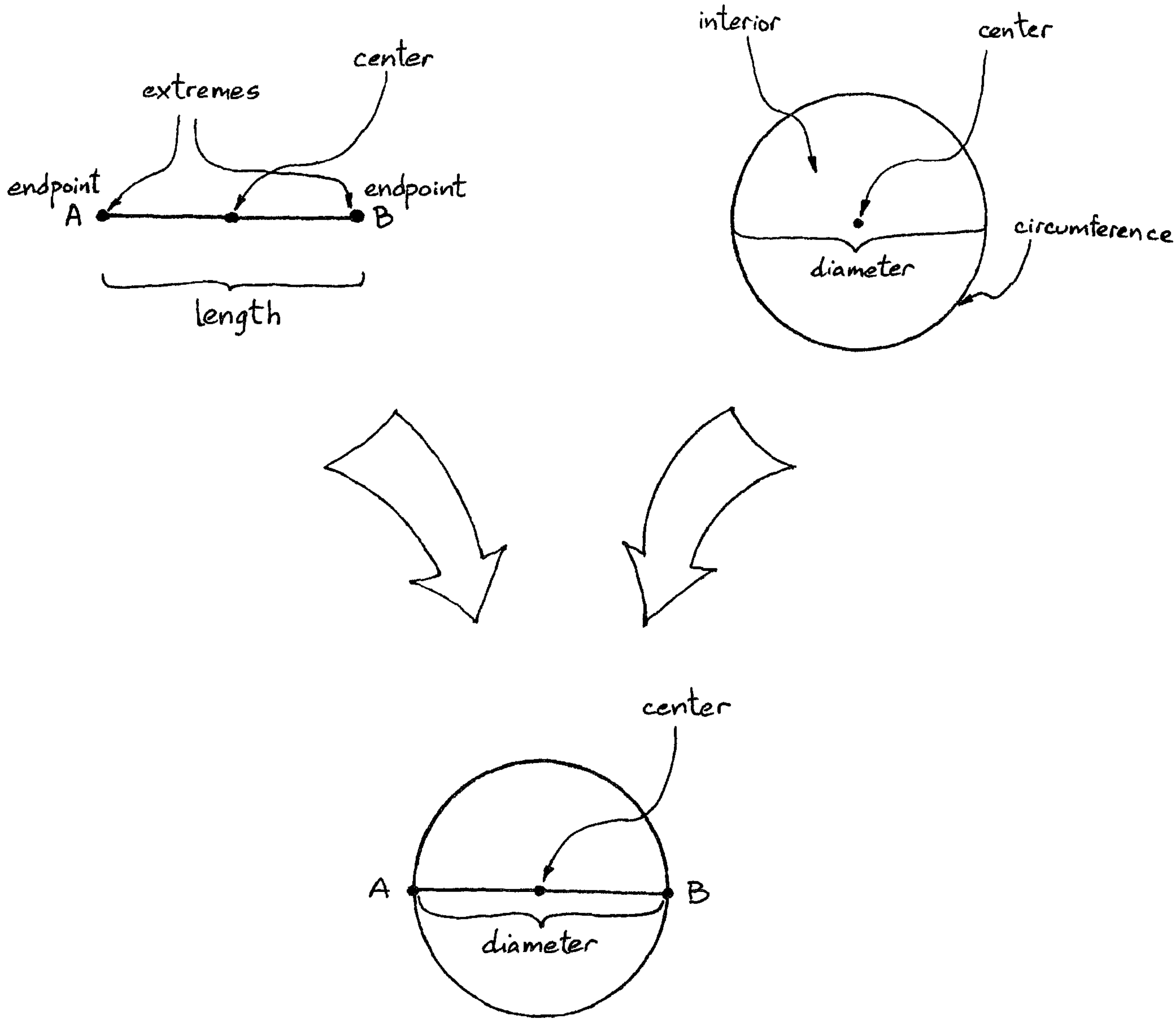


FIGURE 12.1 The Line-Segment frame (top left) and the Disc frame (top right) form a conceptual blend (bottom) in which the line segment is made identical to the diameter of the disc, while the center of the line segment is made identical to the center of the disc. In the blend there are new inferences that weren't available in the frames taken separately.

A point, as we conceptualize it, is a metaphorical entity. It is conceptualized via the Basic Metaphor of Infinity and the Disc/Line-Segment blend. If you try to think of it as a literal entity—one you could encounter in the world—contradictions will arise.

The Infinitesimal Point

Points in textbooks are very often drawn as little discs (see Figure 12.2). This is not a coincidence. It is common to think of a point as a disc made as small as possible.

This is a special case of the Basic Metaphor of Infinity, which applies to produce something infinitely small. But as we have seen, there are two kinds of

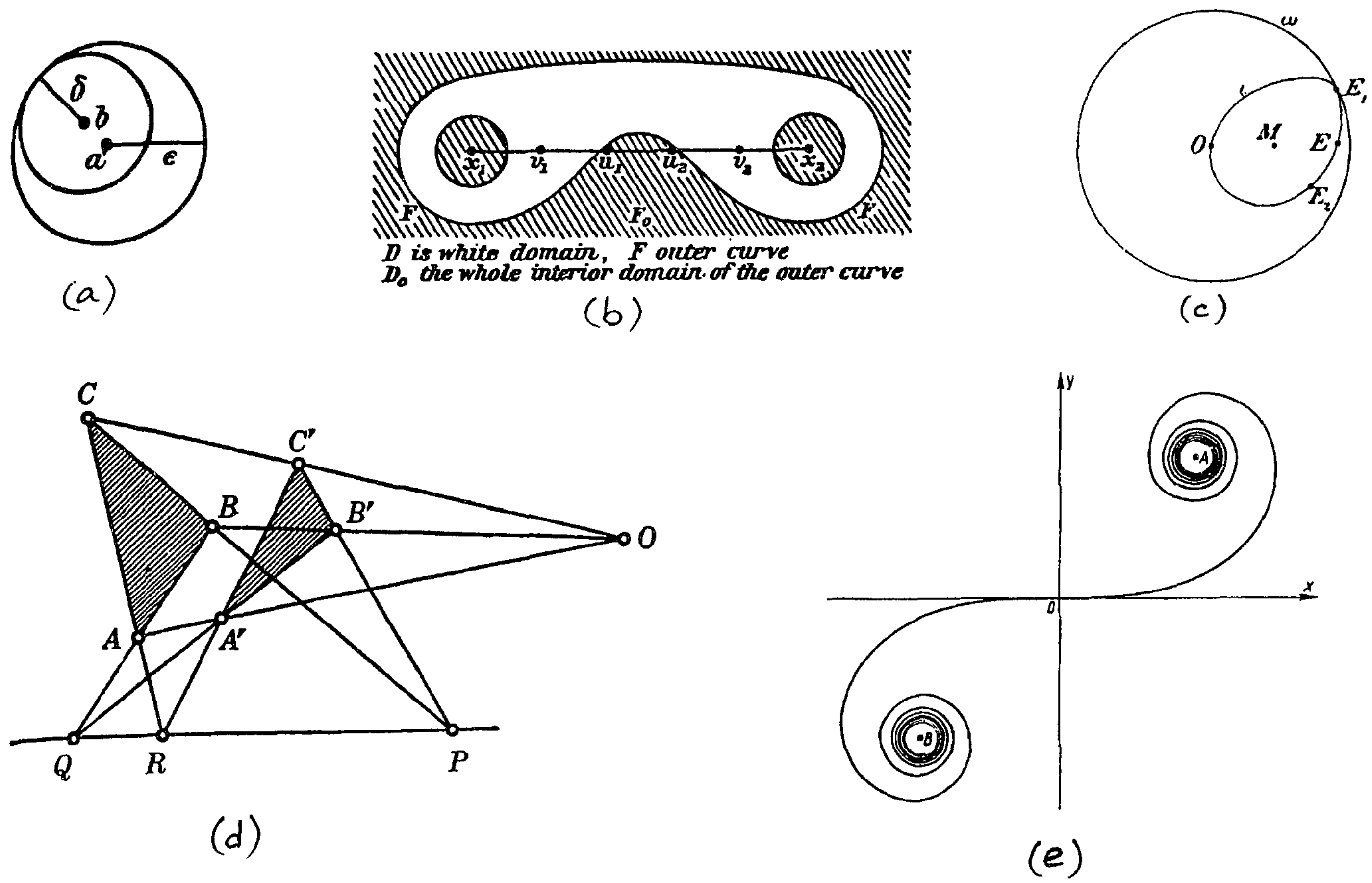


FIGURE 12.2 Points as discs. Here are examples from mathematics books in which points are represented as small discs. Drawings (a) and (b) are taken from a text on the topology of plane sets of points (Newman, 1992); (c), from Hilbert's (1971) famous *Foundations of Geometry;* (d), from Courant and Robbins's (1941) classic *What Is Mathematics?;* and (e), from an Eastern European handbook of mathematics (Bronshtein & Semendyayev, 1985).

things that are "infinitely small": things of size zero and things of an infinitesimal size. To understand the concept of a point better, it is useful to contrast a point of zero diameter with a point that has an infinitesimal diameter. This can be produced by the BMI as follows.

Let us start again with a disc of diameter 1. Again we shrink the disc to half its diameter. But we do one thing different in the process: We keep the diameter greater than zero at all stages in this special case of the BMI. The unending process, then, is:

Shrink the disc from diameter 1 to diameter 1/2, which is > 0.
Shrink the disc from diameter $1/2^{n-1}$ to diameter $1/2^n$, which is > 0.

The BMI completes the process, producing a disc not with diameter zero but with an infinitesimal diameter (the granular $1/2^H$ as discussed in Chapter 11), which is smaller than any real number! We now have a disc that has an infinitely small diameter but is still a normal disc, with its center distinct from its circumference and interior. If we call this a "point," it will have radically different properties from Euclid's "point."

Child psychologists Jean Piaget and Bärbel Inhelder, in experiments as far back as the 1940s, established the following for children around the age of seven or younger. When you tell a child to imagine a disc (or a dot or circle) and make it smaller and smaller and smaller until it's as small as it can get, the child will conceptualize something that is as small as it can be but is still a bona fide disc, with a center separate from the circumference (Piaget & Inhelder, 1948). Suppose a child is told: Start with one of various figures—a circle, triangle, or square—and make it as small as it can get. What the child gets is a point. Now make the point larger and larger. What does it look like? The answer is the same figure the child started with—the circle, triangle, or square. The figures do not lose their integrity and collapse to the mathematical point. Moreover, when children were asked to shrink an object with volume—say, a ball—to a point, they said that points had volume (see Núñez, 1993c).

We can now see why many children are confused about what a point is. The procedure they are told to use to understand a point leads to a contradiction. They avoid the contradiction and come up with something like the infinitesimal point. This is not the Euclidean point at all.

Does this matter? Is the infinitesimal point all that different from Euclid's point? As long as both are infinitely small, what difference can it make?

All the difference in the world.

If points conceptualized as infinitesimal discs touch, then they must touch *at a point*. What is that point? Well, you could say that that point is an infinitesimal disc at the next smallest level of infinitesimals—what we called ${}^2\partial$ in the previous chapter. But then the same problem arises again at that level of infinitesimals. It's turtles all the way down. There is no final characterization of a point, if it is an infinitesimal point.

It is very hard not to think of a point in one of these two ways—either as a zero-diameter disc or as an infinitesimal disc. There are two immediate problems. Either there is a contradiction between the very concept of a disc and the length of its diameter being zero, or there is no way to characterize a point without infinite regress.

Yet people do think in these ways, as we shall see immediately.

Do the Points on a Line Touch?

We have received two kinds of answers to this question. The most common answer, sometimes from professionals who have studied college mathematics, is something like, "Yes. Of course they touch. If they didn't touch, the line wouldn't be continuous. There would be gaps between the points."

The other answer, usually given by mathematicians, is, "Of course not. If two points on a line touched, there would be no distance between them and so they would be the same point."

Those in the first group seem to be thinking of points as being like either zero-diameter discs or infinitesimal discs—infinitesimal discs visualized as little dots or beads on a string. If the line is to be continuous—if it is not to have gaps—then the points have to touch, don't they? The image is of a line made up of discs of zero or infinitesimal diameter, with the circumference of one disc bumping up against the circumference of the next. Such an image of points is inconsistent with a conception of points of zero diameter, which are indeed the same point if they "touch."

But the image of points of zero diameter raises a problem: If two such points touch, they are the same point, but if they don't, there must be a gap between them. And if there is a gap, isn't the line discontinuous—not just in a place or two, but everywhere?!

What Is "Touching?"

Suppose a friend touches your shoulder with his hand. Then there is some location on your shoulder where the distance between your shoulder and his hand is zero. Does this mean that your shoulder and his hand share a common part? No, of course not. Your shoulder and his hand are distinct. There is just no distance between them. Using the metaphor for arithmetizing the concept of "no distance," the distance between shoulder and hand is zero.

Is this what is meant in discretized mathematics by two geometric entities "touching"—that the distance between them is zero? The answer is no.

Imagine two circles touching "at one point." If one adopts the Spaces Are Sets of Points metaphor, then each circle is a set of points. Let us call it a circle-set. When two circles "touch," they are sharing a point in common. That is, there is a single object—a point *P*—that is a distinct, existent entity (independent of any sets it is a member of) and a member of both circle-sets. That point plays a role in constituting both circle-sets.

In other words, when two circles "touch," it is not the case that the two circles are distinct sets of points, with zero distance between two points—Point *A* in the first circle and Point *B* in the second. Points *A* and *B* are the *same point*—the same object in both circle-sets.

This brings us to another way in which discretized space differs from our ordinary conception of space. We normally conceptualize geometric figures as if they were objects *in* space; for example, it is common to think of a circle in the plane as if it were something like a circle painted on the floor. The floor exists independently of the circle; you can erase the circle and the floor is still there—with not one point of the floor changed. The circle is painted on top of the floor and is distinct from the floor.

But this is not the case in discretized geometry. There, geometric figures are sets of points *that are part of the space itself!* The circle "in" the plane consists of a subset of the same points that make up the plane. Discretized space is not something to be "filled"; it is constituted by point-objects. Figures are not "in" space; they are part of space. For this reason, the only way in discretized mathematics to model our ordinary idea of figures "touching" is for a point on one figure and a point on the other figure to *be the same point*. The same is true for points on a line. The only way two points can "touch" is if they are the same point.

This is not like our ordinary notion of touching at all. It would be like saying that two lovers' lips can touch only if they shared common skin. The image is a bit creepy. But that's the metaphor that mathematicians are forced to use in order to carry out the program of discretizing space.

When we think or speak of figures as "touching," we are using our ordinary everyday, naturally continuous notion of space, where geometric figures are metaphorically thought of as objects "in" space and points are conceptualized in their normal way as locations. This is what we are doing, for example, when we think of a tangent line as "touching a curve *at* one point." In thinking this way, we are using the commonplace everyday conceptual metaphor that Geometric Figures Are Objects in Space; we are using our normal metaphor system, which was not developed for discretized mathematics. To think in terms of discretized space is to think using metaphors developed for technical reasons in the discretization program.

This is why mathematicians tend to wince when asked if two points on a line can touch and then say, no. They wince because, in the discretized mathematics in which they are trained, "points" are not the kinds of things that can either touch or not touch. Since they take the Set-of-Points metaphor as the correct way of conceptualizing space, they see the question as an improper question to ask. If points are just elements of sets that can be taken as anything at all—if they are not inherently spatial and not seen technically as objects in space—then the idea of their touching or not touching makes no sense.

As noted, it is common for people who first approach the study of mathematics to conceptualize points as infinitesimally small discs, either zero-diameter discs or infinitesimal discs. Indeed, this is how most students are taught what a point is, and how diagrams in textbooks represent points—namely, as small disclike dots. Yet this conception of a point is radically at odds with the discretization program. From the perspective of that program, points are very different entities—entities that cannot literally touch.

For this reason, many mathematics educators consider the understanding of points as infinitely small discs—either zero or infinitesimal—as a "misconception." Yet this "misconception" is a perfectly natural concept—indeed, an unavoidable one. It is the only natural way that most people have of understanding what a point is. If you are going to teach what a point is in discretized mathematics, you will have to cope with this aspect of normal human cognition. Your students will inevitably understand a point in this way initially. It is important that you and they both understand that such a conception of a point is natural, but that it does not fit the metaphors defining discretized mathematics—metaphors that defy our ordinary intuitions about space.

How Can We Conceptualize a Point in Discretized Mathematics?

It is not the job of mathematics to characterize how people think, or how people conceptualize (or should conceptualize) mathematical ideas. That is the task of the cognitive science of mathematics as a subject matter. The job of mathematicians taking part in the discretization program is to provide precise symbolic expressions "defining" all the notions used in discretizing space: limit points, accumulation points, neighborhoods, closures, metric spaces, open sets, and so on. Our job as cognitive scientists is to comprehend how these "definitions" in symbols are conceptualized, to give an account in terms of human cognition of the ideas that the symbols are meant to express. The question we turn to now is how a point, given the metaphor Spaces Are Sets of Points, is conceptualized within discretized mathematics.

The key to understanding how space is conceptualized in the discretization program is the BMI. The reason is that the strategy of the discretization program was to replace naturally continuous space with infinite sets of points. The theoretical mechanisms that were used all make implicit use of the BMI.

Let us look at the details.

The Centrality of the BMI in the Discretization of Space

The program for discretizing naturally continuous space had a strategy:

1. Pick out the necessary properties of naturally continuous space that can be modeled in a discretized fashion and model them.
2. Model enough of those necessary properties to do classical mathematics as it was developed using naturally continuous space.
3. Call the discretized models "spaces," and create new discretized mathematics replacing naturally continuous space with such "spaces."

For this strategy to be successful, there have to be some small number of such necessary properties that can be successfully modeled in a discretized fashion. The remarkable thing about the discretization program is that such a small number of discretizable properties of naturally continuous space were found and precisely formulated.

This does *not* mean that such properties *completely* model the everyday concept of naturally continuous space. As far as we can tell, it would be impossible to model a naturally continuous concept completely in terms of its opposite. But complete modeling is not necessary for mathematical purposes. All that is necessary is sufficient modeling to achieve a purpose: in this case, being able to discretize classical mathematics with a new notion of a "space" and then create new discretized mathematics using that notion.

Let us now look at the small number of necessary properties of naturally continuous space that have successfully been discretized and are sufficient for the purposes of classical mathematics.

Consider the following properties of naturally continuous space as we normally conceptualize it:

1. *The metric property:* Our everyday concept of "distance" in naturally continuous space has the following necessary properties:
   - The distance between any two distinct locations is greater than zero.
   - If the distance between two locations is zero, "they" are the same location.
   - The distance from location A to location B is the same as the distance from B to A.
   - The distance from A to B is less than or equal to the distance from A to another location C plus the distance from C to B.
2. *The neighborhood property:* "Nearness" implicitly uses the concept of distance. Though nearness is relative to context, in any context, one can pick a distance that is "near." Call that distance *epsilon*. Everything that is closer than epsilon to a point is "near" that point. A "neighborhood" of a point P is the collection of all points near P, in that sense of "near."

3. *The limit point property:* Near every point, there are an infinity of other points, no matter how close you take "near" to be.
4. *The accumulation point property:* Given any point, you can find at least one other point near it.
5. *The open set property:* In any naturally continuous space, there are bounded regions (conceptualized via the Container schema, as discussed in Chapter 2). The interior of such a bounded region (the container minus the boundary) is an open set containing all the spatial locations in that bounded region.

Within discretized mathematics, each of these properties is made precise, starting with the Spaces Are Sets of Points metaphor. These properties have been picked out by mathematicians for the purpose of constructing a discretized mathematics; as such, they are not necessarily the properties that most of us would ordinarily think of.

Given the metric property, the other properties can all be conceptualized via the BMI. Our strategy in demonstrating this will be to construct a special case of the BMI that can be used to characterize all these concepts at once, showing their relationships and just how they implicitly make use of the BMI. We will do this by defining the *infinite nesting property for sets* in terms of a special case of the BMI, and then showing that these concepts can all be characterized using the infinite nesting property.

The basic idea of the infinite nesting property is this.

- Consider a set S with a distance metric—that is, a relation $d(x,y)$ holding among all the members of the set. The distance metric has the following properties: $d(x,y) = d(y,x)$; $d(x,y) = 0$ if and only if $x = y$; and $d(x,y) + d(y,z) \geq d(x,z)$.
- Consider an element C, which may or may not be in S but to which the distance metric applies.
- Characterize an "epsilon disc" around C as a set of points in S within a distance epsilon of C.
- Using the BMI, construct an infinity of nested epsilon discs around C containing members of S.

If this is possible for S and C, we will say that S has the infinite nesting property with respect to C. The main objectives are:

- First, to show how this property can be seen as a special case of the BMI.
- Then, to show how *limit points* and *accumulation points* can be straightforwardly defined as special cases of this special case of the BMI.

We begin by characterizing the concept of an epsilon disc in terms of a conceptual frame, with the semantic roles "distance function," epsilon, and C, the center of the disc.

THE EPSILON-DISC FRAME

A set S with a distance function, or "metric," d.
An element C, either in S or not, to which the distance metric applies.
Epsilon: a positive number.
An epsilon disc around C: a subset of members, x, of S such that $d(C, x) <$ epsilon.

We then characterize what it means for a set S to have the infinite nesting property with respect to an element C, using a special case of the BMI.

NESTED DISCS

Target Domain ITERATIVE PROCESSES THAT GO ON AND ON		*Special Case* A SET OF ELEMENTS WITH THE INFINITE NESTING PROPERTY
The beginning state (0)	⇒	The ε_0-disc: the epsilon-disc frame, with epsilon = ε_0 D_1: the set with the ε_0-disc as a member
State (1) resulting from the initial stage of the process	⇒	Choose ε_1 arbitrarily smaller than ε_0. Result: the set $D_2 = D_1 \cup \{$the ε_1-disc$\}$
The process: From a prior intermediate state (n–1), produce the next state (n).	⇒	Given an ε_{n-1}, choose ε_n arbitrarily smaller than ε_{n-1}. Form $D_{n+1} = D_n \cup$ {the ε_n-disc}.
The intermediate result after that iteration of the process (the relation between n and n–1)	⇒	The set D_{n+1}, containing all ε_i-discs, for $0 \leq \varepsilon_i \leq n$
"The final resultant state" (actual infinity "∞")	⇒	**The set D_∞ containing all ε_i-discs, for $0 \leq \varepsilon_i \leq n$, where n is finite.** **Property 1: Every ε_i-disc in D_∞ contains an infinite number of members of S.** **Property 2: Every ε_i-disc in D_∞ contains both C and a member x of S such that $x \neq C$.**

Let us now look at a set S that has the infinite nesting property with respect to element C, which may or may not be in S.

- Being an ε_i-disc in D_∞ characterizes the notion of being "a neighborhood of C in S." That is, it characterizes the *neighborhood* property.
- Property 1 characterizes the *limit point* property.
- Property 2, when C is in S, characterizes the *accumulation point* property.
- An *open set* is a subset, O, of S such that every member of O is the center C of an ε_i-disc in D_∞.

A neighborhood of an element C is usually characterized simply as an epsilon disc around C. However, the prototypical uses of neighborhoods are in $\mathfrak{R}^n$, the n-dimensional Cartesian space with points as n-tuples of real numbers. As we saw in Chapter 9, the real numbers are defined relative to some use of the BMI—either least upper (greatest lower) bounds, infinite intersections, infinite decimals, infinite polynomials, or some other use. These uses of the BMI to characterize real numbers all impose the infinite nesting property on $\mathfrak{R}^n$, since the infinite nesting property can be seen as a generalization over all those cases: They all involve infinite nesting of some sort. Suppose one takes into account these uses of the BMI in characterizing the reals, as one should. The conceptualization of a neighborhood of C in $\mathfrak{R}^n$ will then implicitly involve a use of the infinite nesting property. Any such neighborhood of C in $\mathfrak{R}^n$ will therefore be an epsilon disc—that is, an ε_i-disc in D_∞, as defined above. Such an epsilon disc will contain an infinity of other epsilon discs around C in $\mathfrak{R}^n$, which is just the property that neighborhoods of a point C have in $\mathfrak{R}^n$. That is why the BMI is implicit in the concept of a neighborhood in $\mathfrak{R}^n$.

The infinite nesting property is a conceptual property intended to show how the BMI is used in the cognitive characterization of all the other properties we have given. It is intended not to be part of discretized classical mathematics as formulated in contemporary mathematics but, rather, to characterize a cognitively plausible account of how such properties can be characterized using the mechanisms of human conceptual systems—for example, frames, conceptual metaphors, conceptual blends, image schemas, and so on.

We will delay until the next chapter an account of how the properties just discussed are used to characterize central notions like continuity in discretized mathematics. For the sake of the remainder of this chapter, there are a number of morals to bear in mind.

The Morals So Far

Here's what we can conclude at this point in our discussion.

- Don't confuse the ordinary, naturally continuous concept of space with the discretized concept characterized by the Spaces Are Sets of Points metaphor.
- Don't confuse the discretized notion of a point with either a spatial location, an infinitesimal disc, or a zero-diameter "disc."
- The discretized notion of a geometric figure (e.g., a circle) contradicts our ordinary notion of a geometric figure in naturally continuous space. The discretized concept of a geometric figure—the one used in contemporary formal mathematics—is therefore *not* a generalization over the concept of a figure used in our ordinary conceptualization. It is a very different kind of conceptual entity, with very different properties.
- Both our ordinary concept of a point and the concept of a point in discretized mathematics make implicit use of the Basic Metaphor of Infinity.
- Conceptualizing a point metaphorically as a disc of zero diameter or an interval of zero length contradicts our everyday conceptual system.
- In thinking about contemporary discretized mathematics, be aware that your ordinary concepts will surface regularly and that they contradict those of discretized mathematics in important ways.
- Trying not to think about your ordinary concepts is like trying not to think of an elephant. You just can't do it when words like "point," "line," and "space" are used. Just be aware that your everyday concepts are there unconsciously and can interfere with your understanding of the metaphorical mathematical concepts.

What Is a Number Line?

Actually, there are two number lines. There is the one you learn in grammar school, where the line is just the ordinary, everyday, naturally continuous concept of a line. This is the line you understand when you examine graphs in the newspaper for temperature readings or stock fluctuations, say. This number line is conceptualized via the metaphor Numbers Are Points on a Line.

NUMBERS ARE POINTS ON A LINE
(FOR NATURALLY CONTINUOUS SPACE)

Source Domain POINTS ON A LINE		*Target Domain* A COLLECTION OF NUMBERS
A Point P on a line	→	A Number P'
A point O	→	Zero
A point I to the right of O	→	One
Point P is to the right of point Q	→	Number P' is greater than number Q'
Point Q is to the left of point P	→	Number Q' is less than number P'
Point P is in the same location as point Q	→	Number P' equals number Q'
Points to the left of O	→	Negative numbers
The distance between O and P	→	The absolute value of number P'

This metaphor constitutes our nontechnical understanding of numbers as points on a line. The number line we learn in elementary school is a conceptual blend—the Number-Line blend—of the source and target domains of this metaphor, in which the entities are *simultaneously numbers and points* (see Figure 12.3).

But there is a second number line as well, one requiring a technical understanding of mathematics, where space is fully discretized—that is, conceptualized using the Spaces Are Sets of Points metaphor. Those with such a technical understanding have a conceptual blend of the source and target domains of that metaphor, which we will call the Space-Set blend.

THE SPACE-SET BLEND

Target Domain NATURALLY CONTINUOUS SPACE WITH POINT-LOCATIONS SPECIAL CASE: THE LINE		*Source Domain* A SET WITH ELEMENTS
The line	↔	A set
Points are locations on the line.	↔	Elements are members of the set.
Point-locations are inherent to the line they are located on.	↔	Members exist independently of the sets they are members of.

Two point-locations are distinct if they are different locations.	$\leftrightarrow$	Two set members are distinct if they are different entities.
Properties of the line	$\leftrightarrow$	Relations among members of the set

In the Space-Set blend, space is conceptualized as a set of elements in which necessary properties of natural continuous space are conceptualized as formal relations on elements of the set.

Those who conceptualize space in this way have a more technical version of the metaphor that Numbers Are Points on a Line. In their metaphor, the source domain is the Space-Set blend, where space is discretized. That more technical

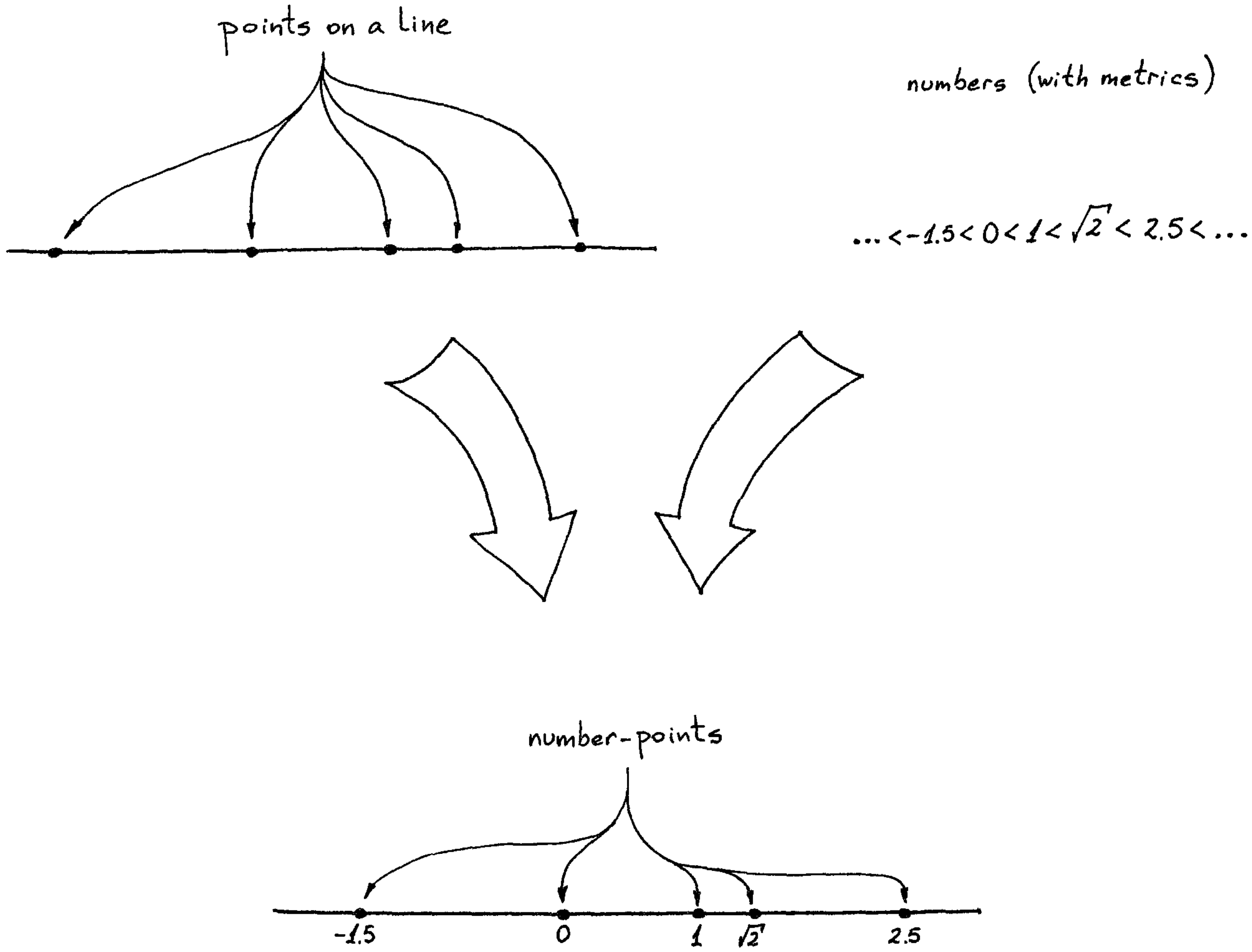


FIGURE 12.3 The Number-Line blend. The metaphor Numbers Are Points on a Line sets up a correspondence between points on a line and numbers. When the source and target domains of this metaphor are both active, the result is a conceptual blend in which the corresponding elements of the metaphor become single entities—number-points. The figure illustrates the source domain of points on a line at the top left and the target domain of numbers (ordered and with a metric) at the top right. The resulting blend is shown at the bottom.

metaphor can be stated as follows, taking the real numbers as special cases of numbers and the line as a special case of a space.

NUMBERS ARE POINTS ON A LINE (FULLY DISCRETIZED VERSION)

Source Domain THE SPACE-SET BLEND				*Target Domain*
NATURALLY CONTINUOUS SPACE: THE LINE		SETS		NUMBERS
The line	↔	A set	→	A set of numbers
Point-locations	↔	Elements of the set	→	Numbers
Points are locations on the line.	↔	Elements are members of the set.	→	Individual numbers are members of the set of numbers.
Point-locations are inherent to the line they are located on.	↔	Members exist independently of the sets they are in.	→	Numbers exist independently of the sets they are in.
Two point-locations are distinct if they are different locations.	↔	Two set members are distinct if they are different entities.	→	Two numbers are distinct if there is a nonzero difference between them.
Properties of the line	↔	Relations among members of the set	→	Relations among numbers
A point O	↔	An element "0"	→	Zero
A point I to the right of O	↔	An element "1"	→	One
Point P is to the right of point Q.	↔	A relation "$P > Q$"	→	Number P is greater than number Q.
Points to the left of O	↔	The subset of elements x, with $0 > x$	→	Negative numbers
The distance between O and P	↔	A function d that maps $(0,P)$ onto an element x, with $x > 0$	→	The absolute value of number P

This is a complex metaphor for conceptualizing numbers in terms of discretized space. The discretized number line is a conceptual blend of the source and target domains of this metaphor. Since the source domain is a metaphorical blend itself, the complete blend is a blend of *three domains*. One can see why this is not so easy to teach or to learn.

It is crucial for what is to follow that the reader understand the vast difference between these two metaphorical conceptions of the number line. In the discretized version based on the Space-Set blend, the source domain is fully discretized. The "points" in "space" are discrete elements in a set. In the naturally continuous version, the source domain contains point-locations in a naturally continuous space—a continuous medium. Change in the discretized version can be represented only as a sequence of discrete set elements, while change in the first is a naturally continuous, holistic path with discrete point-locations on it.

Why the "Real Line" Is Not a Line

When mathematicians speak technically of the "real line," they do not mean our ordinary, naturally continuous line with real numbers sprinkled along it; that is the nontechnical "real line." The technical "real line" is fully discretized, via the Space-Set blend. The "real line" is a set of discrete elements (of an unspecified nature) constrained by various relations among the elements. There is no naturally continuous line in the technical "real line." From the perspective of our ordinary concept of a line as naturally continuous, the "real line" of discretized mathematics is not a line.

A Crucial Property of Naturally Continuous Number Lines

There is a crucial difference between the two conceptions of the number line. In the Discretized Number-Line blend, there is a one-to-one correlation between the point-elements of the set in the source domain and the numbers in the target domain. In this mapping, everything in the source domain is mapped: All the elements of the set are mapped onto all the numbers.

This is not true in the naturally continuous Number-Line blend. There, only certain point-locations on the naturally continuous line are mapped onto numbers. But the whole naturally continuous line—the *medium* in which the points are located—is not mapped. Since the point-locations are zero-dimensional and have no magnitude, no point-location "takes up" any amount of that medium. Point-locations don't take up space. (Recall that in the Space-Set blend, the points *constitute* the space. That is not true here.) This means that no matter what number systems point-locations are mapped onto, the medium of space that is left unmapped will have an unlimited supply of point-locations "left." The point-locations of a naturally continuous number line are never exhausted.

This is not true of the discretized number line, which is a conceptual blend of three conceptual domains—space, sets, and numbers. The set domain plays the central role in the blend. Because sets constitute the source domain of the Spaces Are Sets of Points metaphor, the target domain of space is structured by the ontology of sets. In the Numbers Are Points on a Line metaphor, the "points" are set elements. The kind of number system that the set maps onto depends on the formal relations (e.g., the axioms for the real numbers in Chapter 9) that constrain what the members of the set can be and how they are related to one another.

The real-number system is "complete." Once the reals are characterized, the members of the set are completely determined. This means that the "points" constituting the line are fully determined by the axioms that define the real numbers. For this reason, it is true of the Discretized Real-Number-Line blend that *the real numbers exhaust it.* Of course they do. Since there can be a point in this blend only if a real number maps onto it, the real numbers determine what the points in this blend can be.

But in the naturally continuous real-number line, *the real numbers do not exhaust the points on the line.* The point-locations that map onto the real numbers are not all the possible point-locations on the line. The reason is that this line is a naturally continuous background medium. There is no limit at all to the number of zero-length point-locations that can be assigned to such a background medium. For example, in the case of the naturally continuous hyperreal line, the real numbers are relatively sparse among the hyperreals on that line. On the hyperreal line, a huge number of points surround each real number. The naturally continuous line has no problem whatsoever accommodating all the hyperreal numbers in excess of the real numbers.

Thus, the question of whether or not the real numbers "exhaust the line" is not formulated precisely enough; it depends, unsurprisingly, on what you take the "line" to be.

The Real, Granular, and Hyperreal-Number-Line Blends

Here are two statements one might be tempted to make:

- It is only with the nontechnical, nondiscretized, naturally continuous number line that the real numbers do not exhaust the points on the line. Since the line is naturally continuous, there are always more points of zero length.

- On the line as understood technically by professional mathematicians—the discretized number line, which is a set of points—the real numbers always *do* exhaust the points on the line.

Tempting as such a pair of claims might be, they are not true.

The reason is quite simple. The hyperreal and granular numbers (see Chapter 11) form linearly ordered sets (since they satisfy the first nine axioms for the real numbers). Consider the hyperreals once more. The sets that model the hyperreals are linearly ordered. For this reason, there can be a discretized hyperreal-number line. Just let a set modeling the hyperreals be the set in the metaphor that Spaces Are Sets. Then there will be a discretized hyperreal-number line. On this number line, the real numbers are incredibly sparse. Therefore, the real numbers do not exhaust even a discretized number line—the technical number line used by mathematicians.

Why "Space-Filling Curves" Do Not Fill Space

Just what is a "space-filling curve"?

Imagine a function with the following properties.

- Its domain is the closed interval [0, 1] of the discretized real-number line (i.e., $\Re^1$).
- Its range is the unit square of $\Re^2$, that portion of the Cartesian plane with x- and y-coordinates both within [0, 1].
- The mapping is surjective (onto, but not one-to-one) and is "continuous."

Looking ahead to Chapter 13, we will take "continuity" for discretized functions to mean that the function *preserves closeness*. This can technically be achieved by a condition like the following: The inverse image of the function f^{-1} maps open sets in the range onto open sets in the domain. Since neighborhoods are open sets, what this conditions says is that neighborhoods around each $f(x)$ come from neighborhoods around x.

In other words, the "continuous" discretized line [0, 1] is to be mapped via a "continuous" function onto a "continuous" curve in the square. (Technically, a "curve" can be a straight line.)

Now the question arises: Can such a mapping from the unit line to a curve in the unit square "fill" the entire square? Giuseppe Peano constructed the first such

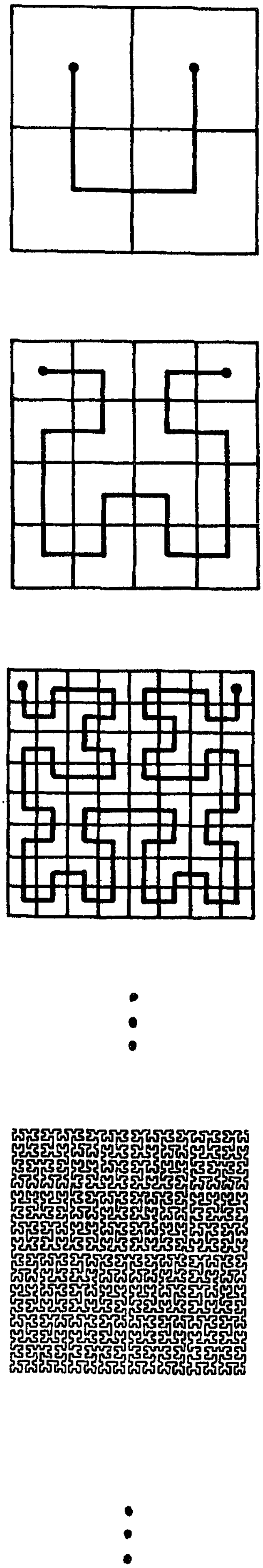


FIGURE 12.4 Hilbert's "space-filling" curve, as described in the text. Shown here are some of the first few members of the sequence of curves, which the interval [0, 1] on the real line is mapped onto. As the sequence gets longer, the resulting curves are said to "fill" more and more of the real-valued points in the unit square. Since the function is real-valued, the points with hyperreal values remain "unfilled." Hilbert's curve, characterized via the BMI, "fills" only the real-valued points in the unit square. The moral here is that when spaces are understood as sets of points, the term "space-filling" means "mapping onto every member of the relevant set of points." "Space" does not mean naturally continuous space. Therefore, Hilbert's "space-filling" curve doesn't *fill space.*

curve in 1890, with others following: David Hilbert in 1891, E. H. Moore in 1900, W. Sierpinsky in 1912, and George Pólya in 1913 (for details, see Sagan, 1994). To give you an idea of what such a "curve" is like, Hilbert's "space-filling" curve is shown in Figure 12.4. Hilbert's curve, like Peano's and all the others, is constructed via an infinite sequence of ordinary curves, as follows. Consider the simple curve in the top box—the first step in the construction of the function. The mapping at this step has the following parts:

1. When x varies from 0 to 1/3 in the interval [0, 1/3], $f(x)$ varies continuously along the left vertical line going from top to bottom.
2. When x varies from 1/3 to 2/3 in the interval (1/3, 2/3], $f(x)$ varies continuously along the horizontal line from left to right.
3. When x varies between 2/3 and 1 in the interval (2/3, 1], $f(x)$ varies continuously along the right vertical line from bottom to top.

At stage 2, the square is divided into 16 smaller squares. The line [0, 1] is divided into 15 subintervals: [0, 1/15], (1/15, 2/15], . . . , (14/15, 1].

1. When x varies from 0 to 1/15 in the interval [0, 1/15], $f(x)$ varies continuously along the upper left horizontal line going from left to right.
2. When x varies from 1/15 to 2/15 in the interval (1/15, 2/15], $f(x)$ varies continuously along the connecting vertical line from top to bottom.

And so on.

At each stage n, there are 4^n subsquares and $4^n - 1$ divisions of the unit line [0, 1] mapped continuously onto $4^n - 1$ connected line segments going from the center of one subsquare to the center of the adjacent subsquare. At each stage, the mapping is "continuous."

As we go from stage to stage, the curve gets longer and longer, with the interval [0, 1] being mapped onto more and more points of the square "filling up" progressively more of the square.

The limit of this sequence of functions is defined in the following way.

- The process defines a sequence of functions f_1, f_2, . . . from [0, 1] onto the square.
- Notice that a given point p in the domain [0, 1] will be mapped onto different points $f_1(p)$, $f_2(p)$, . . . at successive stages of the process.
- For each such point p, there will therefore be a sequence of points $f_1(p)$, $f_2(p)$, . . . in the square.
- Each such sequence will converge to a limit. Let us call that limit $f_\infty(p)$. (Of course, this can be conceptualized via the BMI.)
- Now consider the set of all $f_\infty(x)$, for all x in [0, 1]. It is a function from [0, 1] to the unit square. (This is the set of final resultant states of the BMI.)
- It is "continuous" since every open set in the unit square is mapped by the inverse function $f_\infty^{-1}(x)$ onto an open set in [0, 1].
- Finally, $f_\infty(x)$ "fills" the unit square. That is, for every point in the unit square, there is at least one point in [0, 1] that maps onto it.

This is what Hilbert proved.

Note that as the subsquares get smaller, the boundaries of the subsquares get closer and closer to the center of each subsquare. Consequently, as n gets larger, the points in each subsquare not on the curve get closer and closer to points on the curve. As n approaches infinity, what happens? Look at the distance within each subsquare between points in the subsquare on the curve and points in the subsquare *not* on the curve. As n approaches infinity, that distance approaches zero. At limit stage, where we have the function f_∞, that distance *is* zero. The subsquares have zero area, so that no subsquare can have a point in the plane that is not on the curve. In other words, every point in the square *is* a point on the curve, because the subsquares have become the points on the curve!

At first, this seems bizarre. A curve is one-dimensional. It is infinitely thin. A plane is two-dimensional. How can a one-dimensional curve fill a two-dimensional space?

First, notice that the curve is a metaphorical "curve." Each limit operation for each point in [0, 1] is an instance of the Basic Metaphor of Infinity. The range of the function f_∞ is an infinite set, each member of which is conceptualized via the BMI. In that sense, the function is *infinitely metaphorical*—every one of its infinity of values arises via the Basic Metaphor of Infinity.

Second, notice that the domain and range of the function are metaphorical, since they are not a naturally continuous line segment and a naturally continuous square but, rather, are discretized—produced via the Spaces Are Sets of Points metaphor.

Third, notice that the "continuity" of each f_i and of f_∞ is also metaphorical. It is not the natural continuity of a path of motion. It is the metaphorical "continuity" for discrete "space" characterized in terms of preservation of closeness, as noted. (We will give a detailed analysis of continuity in chapters 13 and 14.)

Fourth, notice that from the perspective of natural continuity, the space of the unit square is not "filled." A naturally continuous square is a holistic background medium. The range of the function is only the set of points in the unit square with real-valued coordinates.

However, the naturally continuous unit square must be able to accommodate not just points with real-valued coordinates but also points with hyperreal-valued coordinates. Thus, the real-valued point-locations do not exhaust the naturally continuous holistic background space of the unit square and therefore the "space-filling curve" does not "fill" the naturally continuous unit square.

Fifth, suppose that we rule out the naturally continuous plane and think only in terms of the discretized plane. Suppose further that we assume that the function maps the real numbers within the hyperreal interval [0, 1] onto the real-valued coordinates within the hyperreal unit square.

In this case, the function f_∞ is exactly the same as it was before. It maps the same real numbers in the interval [0, 1] onto the same points with real-valued coordinates in the unit square. The difference is that now most points in the hyperreal unit square have coordinates that are *not* real-valued but hyperreal-valued. Since the function f_∞ (as we have defined it) does not apply to hyperreal values in the interval [0, 1], it does not "fill" any of the points in the unit square with hyperreal coordinates. Indeed, since the real-valued points in the hyperreal unit square are incredibly sparse, most of the square is not "filled."

In summary, the so-called space-filling curve f_∞ does not "fill space" at all. It only maps discrete real-valued points in the interval [0, 1] onto discrete points with real-valued coordinates in the unit square. That's all.

Moreover, it does not "fill space" under either assumption of what "space" is—naturally continuous space or discretized space.

What Is "The Continuum"?

Before the nineteenth century, the *continuum* referred to the naturally continuous line. The discretization program changed the meaning of the term. In discretized mathematics, "the continuum" became a discretized "line" characterized via the Spaces Are Sets of Points metaphor; that is, it became a set of abstract elements. Given the discretized number line, "the continuum" was seen as a set of numbers.

The question as to whether the real numbers "exhaust" the continuum is the same question as whether the real numbers exhaust the points on a line. It is not a well-put question. It depends on what you mean by "the continuum." The term can mean all of the following:

1) A *naturally continuous line.*
2) A *discretized line*—a "line" conceptualized via the Spaces Are Sets of Points metaphor, consisting of a set of elements that characterize *all* the points on the line.
3) A *Discretized Number-Line blend,* where the "points" are elements of sets, which in turn correspond one-to-one to numbers. The numbers then determine what the metaphorical "points" are. Suppose the number system is "complete," as the real numbers are; that is, they are closed under arithmetic operations and contain the limits of all the infinite sequences. Since the numbers determine what the points are, a complete set of numbers exhausts *all* the points defined by those numbers. And since the real numbers are complete and define what a "point" is in a Discretized Number-Line blend, they, of course, exhaust all the points in that blend. That does not mean that they exhaust all the points on a naturally continuous line.

The claim dating from the late nineteenth century that the real numbers "exhaust the continuum" is usually a claim about case 3. However, many people assume that the claim is about case 1 and somehow true of case 1. This is, of course, a mistake arising from the confusion about what "the continuum" means.

Unfortunately, the mathematical community has done little to disabuse their students and the public of this misunderstanding. As a result, many people have been led to believe (incorrectly) that it has been proved mathematically that the naturally continuous line—taken *to be* the continuum—can somehow be exhausted by the real-number points, which are discrete entities without any length. This seems a wondrous, even mystical idea. Indeed, it is. But it isn't true. And the continued used of the term "continuum" doesn't help clarify the situation, since it suggests the line that most people understand in their ordinary lives—the naturally continuous line.

This misunderstanding can be traced back at least to the German mathematician Richard Dedekind (1831–1916), one of the founders of the discretization program. When he conceptualized a real number as a *cut* on the line, the line he had in mind was naturally continuous (case 1), not just a set of elements. (For cognitive evidence of this claim, see the next chapter.) On the other hand, it was Dedekind who defined real numbers in terms of sets. The image of the "cut" and the definition in terms of sets do not give a consistent picture. It leads to a confusion between case 1 and case 2.

As far as we can tell, that misunderstanding has continued to the present day, even in some segments of the mathematical community (see Weyl, 1918/1994; Kramer, 1970; Longo, 1998).

Is the "Continuum Hypothesis" About the Continuum?

George Cantor, as we saw in Chapter 10, used the term "*C*" for "continuum" to name the cardinality of the real numbers. Most likely, he did this because he took it for granted that the real numbers exhausted the points on the continuum. This meant to him that the real numbers could be put in one-to-one correspondence with the points on the continuum and, by Cantor's metaphor, implied that there were the same number of real numbers as there were points on the continuum—strongly suggesting that even Cantor believed (incorrectly) that the real numbers exhausted the points on the naturally continuous line.

Suppose that you take "the continuum" to mean the naturally continuous line. And suppose you then ask the question, "How many points can fit on the continuum?" The answer, as we have seen, is "As many as you like." If by "the continuum" you mean the one-dimensional, naturally continuous, holistic background space, and if by "point" you mean a location in that background medium with zero length, then you don't have to stop with the real numbers;

you can go into the hyperreals. The number of such potential point-locations is unlimited.

Cantor's Continuum hypothesis was about the cardinality of the set of real numbers—the transfinite *number* of real numbers. He had proved that there were "more" reals than rationals or natural numbers (in the Cantor-size sense). As we saw, he had named the cardinality of the natural and rational numbers "$\aleph_0$" and the cardinality of the reals "C". He called the next largest transfinite number "$\aleph_1$". The Continuum hypothesis said that there was no set of cardinality larger than that of the natural numbers ($\aleph_0$) and smaller than that of the reals (C). That, if true, would mean that the cardinality of the reals was $\aleph_1$. In other words, the Continuum hypothesis can be stated as: $C = \aleph_1$.

Is the Continuum hypothesis about the continuum? It is certainly about the real numbers. It is not about the naturally continuous line. For those people who believe that the term "the continuum" refers to the naturally continuous line, the Continuum hypothesis is not about the continuum.

There is nothing strange or paradoxical or wondrous about this. It is simply a matter of being clear about what you mean.

Closure and the Continuum

The passion for closure pushed the mathematical community to find *all* the linearly ordered numbers that could be solutions to polynomial equations—that is, the system of all the linear numbers closed under the operations of addition and multiplication, including all the limits of sequences of such numbers.

The principle of closure then had another effect. It tended to keep the mathematical community from looking further at ordered number systems that were not required for closure—systems such as the hyperreals. Though infinitesimals were used successfully in mathematics for hundreds of years, the drive for closure of a linear number system led mathematicians to stop with the reals. Closure took precedence over investigating the infinitesimals.

Achieving closure for the real numbers meant "completing" the integers and the rationals. Since numbers were seen as points on the line, that meant "completing" the line—the entire "continuum," finding a "real" number for every possible point. The cardinality of the real numbers—the numerical "size" as measured by Cantor's metaphor—became the "number of points on the line." And the Continuum hypothesis, which is about numbers, was thought of as being about points constituting the naturally continuous line.

What was hidden was that "completeness"—closure with limit points—under the discretization program wound up *defining* what a "point" on a "line" was to be.

The Difference That Mathematical Idea Analysis Makes

All this becomes clear only through a detailed analysis of the mathematical ideas involved. Without such an analysis, false interpretations of crucial ideas are perpetuated and mystery reigns.

- The naturally continuous line is not clearly distinguished from the line as a set of discrete "points," which are not really points at all but arbitrary set members.
- The real numbers are not clearly distinguished from the points on the line.
- The so-called real line is misnamed. It isn't a line.
- So-called space-filling curves are also misdescribed and misnamed. They do not "fill space."
- What is called the continuum is not distinguished adequately from a naturally continuous line.
- And perhaps most important, the discretization program goes undescribed and taken as a mathematical truth rather than as a philosophical enterprise with methodological and theoretical implications. The myth that the discretization program makes classical mathematical ideas more "rigorous" is perpetuated. That program abandons classical mathematical ideas, replacing them with new, more easily symbolizable but quite different ideas.

What suffers isn't the mathematics itself but the understanding of mathematics and its nature.

Mathematical idea analysis removes the mystery of "space-filling" curves. It clarifies what is meant by the "real line" and the "continuum." And most important, it places in clear relief the mathematical ideas used in the philosophical program of discretizing mathematics—a program that introduces new ideas, rather than formalizing old ones, as it claims to do.

Students of mathematics—and all of us who love mathematics—deserve to have these mathematical ideas made clear.

13

Continuity for Numbers: The Triumph of Dedekind's Metaphors

THE BASIC METAPHOR OF INFINITY, as we have just seen, is at the heart of the discretization program for reconceptualizing the continuous in terms of the discrete. From the cognitive perspective, it is the BMI that permits the appropriate conceptualization in purely discrete terms of infinite sets, infinite sequences, limits, limit points, points of accumulation, neighborhoods, open sets, and (as we shall see in this chapter) *continuity*.

To appreciate the remarkable achievement of this enterprise and the fundamental role of metaphor in it, let us turn to a historic moment in modern mathematics. The turning point came in the early 1870s, with the classic work of Richard Dedekind and Karl Weierstrass. They successfully launched the movement toward discretization, a reaction against the use of geometric methods via analytic geometry. Descartes had seen the real numbers as points on a naturally continuous line. Newton, linking physical space with naturally continuous mathematical space, had created calculus in a way that depended on geometric methods.

Dedekind showed, through a dramatic use of conceptual metaphor, that the real numbers did not have to be seen as points on a naturally continuous line. And through an implicit use of the Basic Metaphor of Infinity, he showed how to construct the real numbers using sets (infinite sets, of course) of discrete elements. Weierstrass, making another implicit use of the BMI, showed how to make calculus discrete, eliminating all naturally continuous notions of space and movement.

An important dimension of the discretization program was the concept of "rigor." This meant the use of discrete symbols and precisely defined, systematic algorithmic methods, allowing calculations that were clearly right or wrong. They could be written down step by step and checked for correctness. The prototypical cases were the methods of calculation in arithmetic. Those methods of calculation provided for certainty and precision, which were taken as the hallmarks of mathematics in nineteenth-century Europe.

Geometry involved visualization and spatial intuition. But mathematics in nineteenth-century Europe increasingly became a field that valued certainty, order, and precise methods over imagination and visualization. Indeed, visualization and spatial intuition were eventually disparaged. By the end of the century, they were regarded negatively—that is, not for what they could contribute but for how they threatened certainty, objectivity, precision, and the sense of order those mathematical virtues brought. Geometric methods came to be seen as vague, imprecise, not independently verifiable, and therefore unreliable. What could be worse than unreliability in a field that valued certainty, reliable methods, and objective truth?

It was thus imperative in nineteenth-century European mathematics to arithmetize calculus and eliminate geometry from the idea of number in arithmetic. Calculus had to be rescued from geometric methods and "put on a secure foundation"—a discrete and fully symbolizable and calculable foundation. If Weierstrass could arithmetize calculus—reconceptualize calculus as arithmetic—he could eliminate from calculus all visualization and spatial intuition, and thereby bring it into the realm of "rigorous" methods.

This meant conceptualizing naturally continuous space in terms of discrete entities and holistic motion in terms of *stasis* and *discreteness*. From a cognitive perspective, the task required conceptual metaphor—a way to conceptualize the continuous in terms of the discrete. Combined with other fundamental metaphors, the Basic Metaphor of Infinity permitted the understanding of naturally continuous space in discrete terms.

The central issue was continuity. The place and time were Germany in 1872.

Dedekind Cuts and Continuity

Like many other mathematicians of his era, Dedekind was profoundly dissatisfied with the way calculus had been developed—namely, through geometric notions like secants and tangents. Calculus was, after all, part of the subject matter of arithmetic functions and so ought properly to be understood in terms of arithmetic alone and not geometry. But continuous functions were understood at the time in geometric terms, as naturally continuous curves. How, he asked, could continuity be understood in arithmetic terms? The key, he believed, was the real numbers.

Here is Dedekind writing in his 1872 classic, *Continuity and Irrational Numbers* (translated into English in 1901).

> The statement is so frequently made that the differential calculus deals with continuous magnitude, and yet an explanation of this continuity is nowhere given; even the most rigorous expositions of the differential calculus do not base their proofs upon continuity but, with more or less consciousness of the fact, they either appeal to geometric notions or those suggested by geometry, or depend upon theorems which are never established in a purely arithmetic manner. (p. 2)

Dedekind further assumed that all arithmetic should come out of the natural numbers. The arithmetic of the rational numbers had been characterized in terms of the arithmetic of the natural numbers.

> I regard the whole of arithmetic as a necessary, or at least natural consequence of the simplest arithmetic act, that of counting. . . . The chain of these numbers . . . presents an inexhaustible wealth of remarkable laws obtained by the introduction of the four fundamental operations of arithmetic. (p. 4)

The idea here is that the full development of arithmetic should come from the same laws governing the natural numbers.

> Just as negative and fractional rational numbers are formed by a new creation, and as the laws of operating with these numbers must and can be reduced to the laws of operating with positive integers, so we must endeavor completely to define irrational numbers by means of the rational numbers alone. (p. 10)

In other words, irrational numbers must be understood in terms of rationals, which in turn must be understood in terms of the natural numbers.

> . . . [N]egative and rational numbers have been created by the human mind; and in the system of rational numbers there has been created an instrument of even greater perfection. . . . [T]he system *R* [the rational numbers] forms a well-arranged domain of one dimension extending to infinity on two opposite sides. What is meant by this is suffi-

> ciently indicated by my use of expressions from geometric ideas; but just for this reason it will be necessary to bring out clearly the corresponding purely arithmetic properties in order to avoid even the appearance as if arithmetic were in need of ideas foreign to it. . . . (p. 5)

Even though numbers were at that time understood in geometric terms, the geometry had to be eliminated in order to characterize the *essence* of arithmetic. Dedekind then describes in great detail (pp. 6–8) what we have called the Number-Line blend (see Chapter 12), and goes on to say,

> This analogy between rational numbers and the point of a straight line, as is well known, becomes a real correspondence when we select upon the straight line a definite origin or zero-point and a definite unit length for the measurement of segments. (pp. 7–8)

Dedekind's notion of "real correspondence" is fateful. He means a one-to-one correspondence—not merely that there is a point for every number but also that *there must be one and only one number for every point!* What Dedekind has described is essentially what we have called the Discretized Number-Line blend, which is a conceptual blend that uses two metaphors: Spaces Are Sets and Numbers Are Points on a Line. The blend, as we saw in the previous chapter, contains three domains precisely linked by these metaphorical mappings: Space, Sets, and Numbers.

Here we see Dedekind in the process of constructing this metaphorical blend. In the process, he goes back and forth, attending first to one domain and then another: first to space and then to numbers and sets. Back and forth. If naturally continuous space is to be "eliminated," the crucial elements of space must be picked out and modeled by corresponding entities in the domains of numbers and sets.

Dedekind's activity in constructing his version of the Discretized Number-Line blend is taken for granted in his reasoning from here on. He observes that when the rational numbers are associated with points on the line, there are points not associated with numbers, which creates a problem.

> Of the greatest importance, however, is the fact that in the straight line l there are infinitely many points that correspond to no rational number. If the point p corresponds to the rational number a, then, as is well-known, the length op is commensurable with the invariable unit of measure used in the construction, i.e., there exists a third length, a so-called common measure, of which these two lengths are integral multiples. But the ancient Greeks already knew and had demonstrated that there are lengths incommen-

> surable with the given unit of length. If we lay off such a length from the point o upon the line, we obtain an endpoint which corresponds to no rational number. Since it further can be easily shown that there are infinitely many lengths that are incommensurable with the unit of length, we may affirm: The straight line *l* is infinitely richer in point-individuals than the domain *R* of rational numbers in number-individuals.
>
> If now, as is our desire, we try to follow up arithmetically all phenomena in the straight line, the domain of rational numbers is insufficient and it becomes absolutely necessary that the instrument *R* constructed by the creation of the rational numbers be essentially improved by the creation of new numbers such that the domain of all numbers shall gain the same completeness, or as we may say at once, the same *continuity*, as the straight line. (p. 9)

Here we have the heart of Dedekind's link between the real numbers and continuity. The line is absolutely continuous. When arithmetic functions are characterized in the Cartesian plane, continuous functions are represented, metaphorically, by continuous lines (see Case Study 1). A function maps numbers (which are discrete) onto other numbers. If a function in the Cartesian plane is a continuous curve, then the collection of *numbers* corresponding to points on a line must be *continuous,* too. This is a metaphorical inference in the version of the Discretized Number-Line blend that Dedekind was constructing. The naturally continuous line is part of the space domain in the blend. The line in the blend is (naturally) continuous. Points on the line are mapped one-to-one to numbers via the metaphors that Numbers Are Points on a Line and A Line Is a Set of Points. The inference in Dedekind's developing version of the blend is clear: If the line, seen as made up of points, is continuous, then the set of numbers must be continuous, too.

> The above comparison of the domain *R* of rational numbers with a straight line has led to the recognition of the existence of gaps, of a certain incompleteness or discontinuity of the former, while we ascribe to the straight line completeness, absence of gaps, or continuity. (p. 10)

This is a crucial inference! The rational numbers, by themselves, have no *gaps.* In the set of rational numbers, rational numbers are all there is. Rational numbers are ratios of integers. Among the ratios of integers, there are no gaps; that is, every ratio of integers is in the set of rational numbers (see Figure 13.1).

Where does the very idea of a "gap" in the rational numbers come from? The answer is metaphor and conceptual blending. When the Number-Line blend is formed on the basis of the metaphor that Numbers Are Points on a Line, a

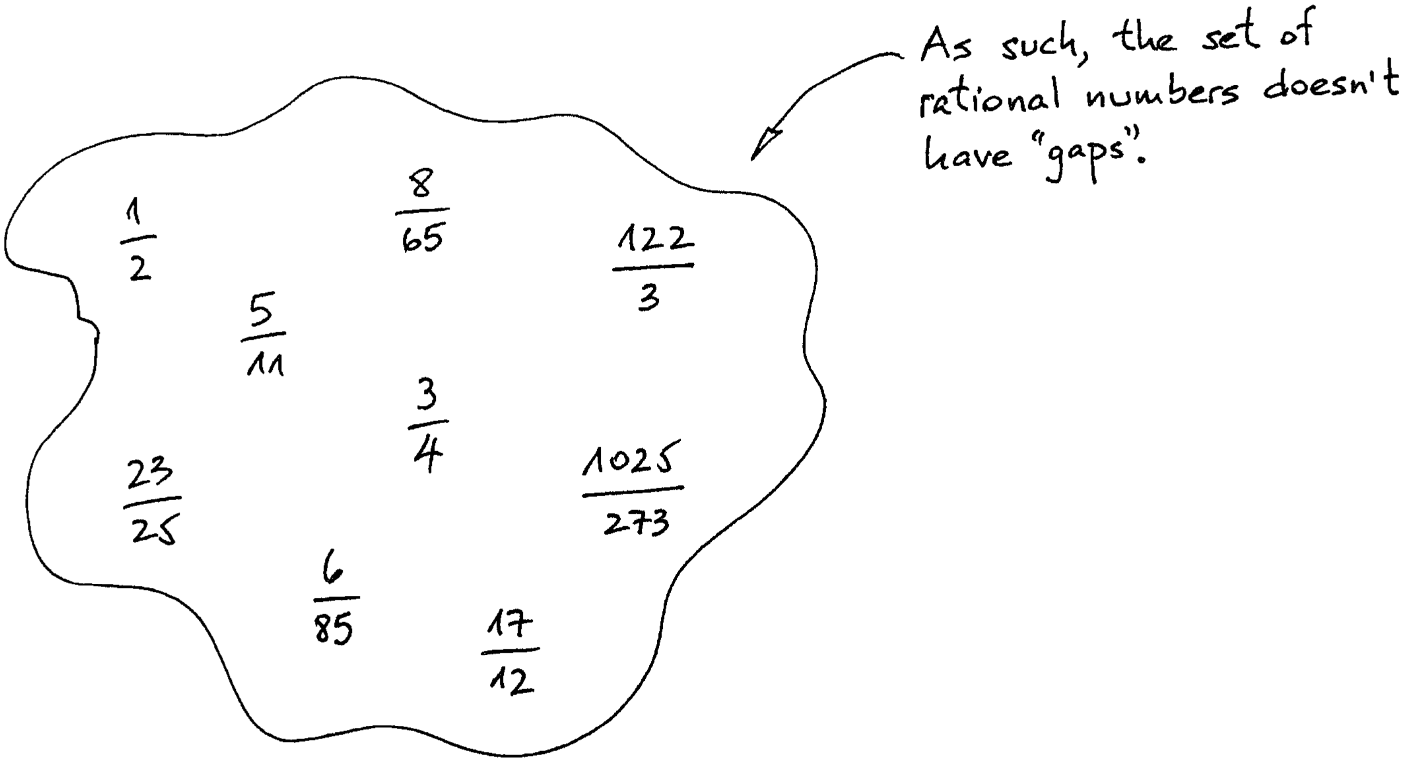


FIGURE 13.1 The set of all rational numbers—and nothing but rational numbers. Within *this* set, by itself, there are no "gaps" between the rational numbers, since there are no other entities being considered. For there to be a "gap," other entities must be under consideration. So-called gaps between the rational numbers appear under two additional conditions: (1) in the Number-Line blend, where all points on a line are assumed to correspond to a number; and (2) with the added idea of closure, where arithmetic operations (e.g., taking the square root) must yield a number as output (when such a number can be well-defined).

metaphorical correspondence is set up between numbers and points. Since this conceptual blend of space and arithmetic is used for measurement and built into measuring instruments (like rulers), it is taken for granted as objectively true: There *is* a true correspondence between points and numbers, as seen in the act of measuring and in the instruments for doing so.

Dedekind accepts this metaphorical conceptual blend as a truth—as many mathematicians have since the Greeks—because it is instantiated in physical measuring instruments. Those instruments have been (for centuries) constructed in order to be consistent with that conceptual blend. The instrument and the conceptual blend are, of course, different things. The instrument—as a measuring instrument—makes no sense in the absence of the conceptual blend of lines and numbers. And of course, there are things in the blend that are not in the instrument. For example, real numbers are not on any physical measuring tape or thermometer. Measuring tapes and thermometers are physical instruments in the world, conceived to be consistent with, and interpreted through, conceptual metaphorical blends. But that does not mean that all the entities in the conceptual blend (like the real numbers, in this case) are in the instrument or in the world. To believe so is to impute reality to something conceptual. That is just what Dedekind was doing when he concluded that there was a "gap" in the rational numbers. *It is only in the Number-Line blend that a "gap" exists*—a "gap" in points on a naturally continuous line that are paired with numbers. The "gap" is in the space domain, where there are points on a naturally continuous line that are not paired with rational numbers. This "gap" makes sense only in the metaphorical conceptual blend that Dedekind was constructing and we have inherited.

Measurement and Magnitude

As we have just seen, the Number-Line blend is not an arbitrary blend. It arises from an activity that has a physical as well as a conceptual part—the activity of measuring real objects in the world. In measuring, numbers are associated with pointlike locations in the physical world. Those physical points are not like metaphorical geometric points of length zero. They are physical positions of small but positive physical dimensions. And in measuring along a line, two distinct physical positions correspond to two different distances from a chosen origin point. Those distances are measured by numbers, and so each distinct "point" corresponds to a different number. This intuitive understanding is characterizable as a folk theory.

THE FOLK THEORY OF MEASUREMENT AND MAGNITUDE

- A line segment from one point to another has a magnitude.
- When a line segment is assigned as a unit of measure and set equal to the number 1, then any magnitude can be measured by a positive number.
- The Archimedean principle: Given numbers A and B (where A is less than B) corresponding to the magnitudes of two line segments, there is some natural number n such that A times n is greater than B.

Notice that the Archimedean principle is built into our everyday folk theory of measurement and magnitude. It entails that, given a measuring stick, even a very small one, you can measure a magnitude of any finite size by marking off the length of the measuring stick some finite number of times. This folk theory, including the Archimedean principle, is implicit in Dedekind's account of the ancient Greek ideas about number lines, as well as in his version of the Number-Line blend.

All this is taken for granted when Dedekind says, "we ascribe to the straight line completeness, absence of gaps, or continuity." Here he is creating yet another metaphor—one that has come down to us today: *Continuity Is Gaplessness*—or, more precisely, Continuity for a Number Line Is Numerical Completeness, the absence of gaps between numbers. Not gaps between points per se, but gaps between *numbers!*

Here is the great break with the everyday concept of natural continuity, the concept that had been used in mathematics for two millennia! Dedekind's new metaphor, *Continuity Is Numerical Completeness*, represented a change of enormous proportions. The "natural" continuity imposed by motion is gone. The metaphor that A Line Is a Set of Points is implicitly required. Continuity no longer comes from motion but from the completeness of a number system. Since each number is discrete and each number is associated one-to-one with the points on the line, there is no longer any naturally continuous line independent of the set of points. *The continuity of the line—and indeed of all space—is now to come from the completeness of the real-number system, independent of any geometry or purely spatial considerations at all.* Dedekind asks,

> In what then does this continuity consist? Everything must depend on the answer to this question, and only through it shall we obtain a scientific basis for the investigation of all continuous domains. By vague remarks on the broken connection in the smallest parts, obviously nothing is gained; the problem is to indicate a precise characteristic of continuity that can serve as a basis for valid deductions. . . . It consists in the following. . . . [E]very point *p* of the straight line produces a separation of the same into two portions such that every point on one portion lies to the left of every part of the other portion. I find the essence of continuity in the converse, i.e., in the following principle:
>
> "If all points of the straight line fall into two classes such that every point of the first class lies to the left of every point of the second class, then there exists one and only one point which produces this division of all points into two classes, this severing of the straight line into two portions." (p. 11)

This is Dedekind's celebrated characterization of the real numbers in terms of what has come to be called *the Dedekind cut*. This is a profound mathematical idea! It seems obvious at first, but as we shall see, it isn't. To show the true richness of Dedekind's idea, we need to apply the tools of mathematical idea analysis.

. . . . .

Dedekind's Idea of the Cut is built out of three conceptual parts: the Cut frame, the Geometric Cut metaphor, and the Arithmetic Cut metaphor. Dedekind begins with the Number-Line blend for the rational numbers and defines a new and original structure on it, which we will call *Dedekind's Cut frame.*

DEDEKIND'S CUT FRAME

The Number-Line blend for the rational numbers, with a point C (the "cut") on the line, dividing all the rationals into two disjoint sets, A and B, such that every member of A is to the left of, and hence less than, every member of B.

Given this precise notion of a "cut" defined relative to the Rational-Number-Line blend, Dedekind goes on to propose a new geometric metaphor for irrational numbers. By this means, he extends the Rational-Number-Line blend to metaphorically create the Real-Number-Line blend.

DEDEKIND'S GEOMETRIC CUT METAPHOR

Source Domain THE RATIONAL-NUMBER-LINE BLEND, WITH THE CUT FRAME		*Target Domain* THE REAL-NUMBER-LINE BLEND, WITH THE CUT FRAME
Case 1: A has a largest rational R, or B has a smallest rational R. C (the "cut")	→	R
Case 2: A has no largest rational and B has no smallest rational. C (the "cut")	→	I, an irrational number

The idea here is this: Either the "cut" will fall at a rational point or it will not. If it falls at a rational point, that rational number will either be the smallest number in B or the largest number in A. In that case, we conceptualize the rational number R as being the cut C. This is not really anything new. The innovation of the metaphor comes in case 2, where A has no largest rational and B, no smallest rational. The "cut" in the rational number line then falls "between" the rationals. The great innovation is to metaphorically "define" an irrational number as *being* the "cut" between the two sets of rationals. Notice that there are no irrational numbers in the source domain of the metaphor but only a "cut"—a point on the line in the blend. The irrationals are conceptualized metaphorically as *being* such cuts.

The brilliance of Dedekind's Geometric Cut metaphor is that the cut characterizes a *unique* irrational number. The uniqueness is not obvious. It follows from three assumptions:

- Assumption 1: Dedekind's Number-Line blend, in which there is a one-to-one correspondence between points and numbers. This guarantees that there will be a number for every point.
- Assumption 2: the Archimedean principle, which holds in his version of the Number-Line blend. This guarantees that the number will not be infinitesimal.

These two assumptions guarantee that if there are two irrational points characterized by (A, B), then there has to be a finite distance between them.

- Assumption 3: the denseness property for the rationals—namely, that there is a rational number on every finite interval in the Rational-Number-Line blend.

Assumption 3 guarantees that if there were two distinct irrational points defined by (A, B) and separated by a finite distance, then there would be a rational point between them, and hence not in either A or B. But this contradicts the assumption that A and B jointly contain *all* the rational points. Given this contradiction, it follows that the irrational point I must be unique.

Notice that it is only by making assumptions 1 and 2 that this conclusion follows. Assumptions 1 and 2 are not necessary assumptions for doing mathematics, as we have seen in previous chapters. Dedekind thus got his uniqueness result by making assumptions that others do not, and need not, make.

But Dedekind achieved only half his job with the Geometric Cut metaphor. What he did was to extend the Rational-Number-Line blend to the Real-Number-Line blend, creating the reals from the rationals using only one frame and one metaphor. But his ultimate goal was to define the real numbers in terms of the rationals, using no geometry at all. Having constructed the Real-Number-Line blend, which has geometry in it, he now had to construct a metaphor to eliminate the geometry—that is, to conceptualize his Real-Number-Line blend only in terms of numbers and sets. Below is the conceptual metaphor with which he did so. The basic idea here is to conceptualize the cut C in the Number-Line blend in terms of the ordered pair of sets of rational numbers (A, B).

DEDEKIND'S ARITHMETIC CUT METAPHOR

Source Domain RATIONAL NUMBERS AND SETS		*Target Domain* THE REAL-NUMBER-LINE BLEND, WITH THE GEOMETRIC CUT METAPHOR DEFINING IRRATIONALS
An ordered pair (A, B) of sets of rational numbers, where $A \cup B$ contains all the rationals and every member of A is less than every member of B	$\rightarrow$	The point C (the "cut") on the line dividing all the rationals into two sets A and B, such that every member of A is to the left of, and hence less than, every member of B
Case 1: A has a largest rational R, or B has a smallest rational R. $(A, B) = R$	$\rightarrow$	Case 1: A has a largest rational R, or B has a smallest rational R. C (the "cut") $= R$
Case 2: A has no largest rational, and B has no smallest rational. $(A, B) = I$, not a rational.	$\rightarrow$	Case 2: A has no largest rational, and B has no smallest rational. $C = I$, an irrational number.

Here the geometric cuts are "reduced" via the metaphor to sets and rational numbers alone. In each case, the set (A, B) is mapped onto the geometric cut C. The cut C is therefore conceptualized in terms of *sets and numbers alone*. Irrational numbers, which were shown to be uniquely characterized via the Geometric Cut metaphor, are now also uniquely characterized via the Arithmetic Cut metaphor.

In short, the Dedekind Arithmetic Cut metaphor states: *A real number is an ordered pair of sets (A, B) of rational numbers, with all the rationals in A being less than all the rationals in B, and all the rational numbers are in A ∪ B.*

Here is the triumph of metaphor at a crucial moment in the birth of the discretization program. We can see in its full beauty the depth and subtlety of Dedekind's metaphorical idea of the cut.

The Geometric Basis of Dedekind's Completeness for the Real Numbers

Dedekind's two cut metaphors, taken together, have an extraordinary entailment: The real numbers characterized in this way constitute *all* the real numbers! This entailment does not come from arithmetic alone; it relies on the

Geometric Cut metaphor. This is an important point and is worth going into in some detail.

Dedekind assumed that for each point on a line, there was a real number that could measure the distance of that point from an origin point (given the choice of a unit). This defines what we will call the *Measurement Criterion for Completeness* of the real numbers. If you can characterize numbers for all such points, then you have characterized *all* the real numbers. The Measurement Criterion for Completeness is a geometric criterion.

Dedekind sought an arithmetic characterization of the real numbers—a characterization that would meet this geometric criterion. If he could find it, he could be sure he had characterized *all* the real numbers. To meet this geometric criterion, Dedekind needed *both* the Geometric and Arithmetic Cut metaphors.

It is important to realize that the Measurement Criterion for Completeness of the real numbers uses assumptions 1 and 2 above: Dedekind's Number-Line blend and the Archimedean principle. It is the Number-Line blend that gets you from (a) assigning numbers to all the points measured to (b) characterizing all the real numbers. It is the Archimedean principle that guarantees that two points will not be separated by an infinitesimal.

Thus, Dedekind's characterization of the completeness of the real numbers depended on geometry after all, since he took a geometric criterion to define what "completeness" was to be. Notice that he did not use the algebraic criterion explicitly in arguing for completeness, though he knew from the work of Carl Friedrich Gauss and others that the real numbers constituted a complete ordered field. The algebraically complete real numbers are those that are both (1) roots of polynomial equations and (2) linearly ordered. Apparently, algebraic completeness was not enough for Dedekind. He did not argue for the completeness of the real numbers on algebraic grounds (*that* had been proved by Gauss) but only on geometric grounds.

It is somewhat ironic that the man given credit for arithmetizing the completeness of the real numbers did so using a *geometric* criterion and a *geometrically* based metaphor.

Continuity as Numerical Completeness

Dedekind's Arithmetic Cut metaphor is stated in such a way as to guarantee that there will be no "gaps" in the real numbers. Every real number meeting the Measurement Criterion for Completeness will be characterized in terms of the rational numbers via the Arithmetic Cut metaphor.

"Continuity" for the real-number line can then be conceptualized purely in terms of numbers, which are discrete entities, via the following metaphor: Continuity for the Number Line Is Arithmetic Gaplessness. The result is a metaphorical "continuum" of discrete entities: the "real-number continuum"—that is, the "gapless" sequence of all real numbers defined metaphorically using only numbers and sets as "cuts"—with the geometry eliminated by the Arithmetic Cut metaphor.

Continuity for *space* can now be characterized in terms of *metaphorical "continuity"* for numbers. Dedekind adds,

> The assumption of this property of the line is nothing else than an axiom by which we attribute to the line its continuity, by which we find continuity in the line. If space has at all a real existence, it is *not* necessary for it to be continuous; many of its properties would remain the same even were it discontinuous. And if we knew for certain that space was discontinuous, there would be nothing to prevent us from filling up its gaps, in thought, and thus making it continuous; this filling would consist in the creation of new point-individuals and would have to be effected in accordance with the above principle. (p. 12)

It is remarkable that he says, *"If space has at all a real existence. . . ."* If he can eliminate space from mathematics, maybe it doesn't exist. He goes on, *"If space has at all a real existence, it is* not *necessary for it to be continuous; many of its properties would remain the same even were it discontinuous."* Here it is clear that by "continuous" space he means our ordinary concept of naturally continuous space. If he can create a mathematics that does as well as the old mathematics, but is discrete rather than naturally continuous, maybe real space (if it exists!) is also discrete. Maybe the world fits the mathematics he is creating . . . .

The bottom line for Dedekind is this: Continuity for space in general is now to be conceptualized in terms of continuity for numbers. Space must therefore be made up of discrete points, which are not mere locations in a naturally continuous medium. The points *constitute* the space, the plane, the line, and all curves and surfaces and volumes. *What makes lines and planes and all space continuous is their correspondence to numbers in a complete, "continuous" number system.*

And what do Dedekind's metaphors—his continuity metaphor and his two cut metaphors—imply about infinitesimal numbers? The metaphorical entailment is immediate: *Infinitesimal numbers cannot exist!* If the real numbers use up all points on the line, and if the real numbers form a complete system—with the principle of closure and the measurement criterion requiring no further extensions—then there is no room in the mathematical universe, either the universe of arithmetic or that of geometry, for infinitesimals. As we have seen, this comes from the Archimedean principle, which is built into the folk theory of measurement and magnitude, which in turn is implicit in Dedekind's concept of the number line. The Archimedean principle was absolutely necessary both to guarantee the uniqueness of the real numbers as Dedekind constructed them and to guarantee the completeness of the real numbers according to the measurement criterion.

Infinitesimal numbers had been used for two centuries by hundreds of mathematicians to do calculus. But the arithmetization of calculus that Dedekind envisioned could not make use of infinitesimal numbers, because it used a version of the Number-Line blend with an implicit Archimedean principle. If calculus was to be arithmetized in terms of numbers alone and not geometry, only the real numbers would do. Moreover, Dedekind's Continuity metaphor did not even leave room for infinitesimals in space. All space was to be structured by the real numbers alone.

What we see in these pages from Dedekind is one of the most important moments in the history of modern mathematics. Not only is arithmetic freed from the bonds of the Cartesian plane, but the very notion of space itself becomes reconceptualized—and mathematically redefined!—in terms of numbers and sets. In addition, the real-number line becomes reduced to the arithmetic of rational numbers, which had already been reduced to the arithmetic of natural numbers plus the use of classes. This sets the stage for the Foundations movement in modern mathematics, which employs the Numbers Are Sets metaphor to reduce geometry, the real line, the rational numbers, and even the natural numbers to sets. All mathematics is discretized—even reason itself, in the form of symbolic logic.

All that was needed was the reduction of calculus to arithmetic. Such a program of reduction had begun with the work of Augustin-Louis Cauchy in 1821. Cauchy had given preliminary accounts of the real numbers, continuity, and derivatives in purely arithmetic terms using what has come to be called the epsilon-delta characterization of a limit. It remained for Karl Weierstrass to complete the job.

14

Calculus Without Space or Motion: Weierstrass's Metaphorical Masterpiece

BETWEEN 1821 AND THE 1870S, the need for an arithmetization of calculus had reached crisis proportions. The reason had to do with the discovery of a number of functions called monsters. What made these functions "monsters" was a common understanding of arithmetic functions and calculus in terms of three interlocking ideas forming a conceptual paradigm, which we will call the *Geometric Paradigm.*

We will discuss the "monsters" below. In order to appreciate them and their effect on the mathematical community, it is important first to have a clear idea of the geometric paradigm, and then to understand the arithmetization of calculus that came to replace it. Here is an outline of the geometric paradigm.

THE GEOMETRIC PARADIGM
• The Cartesian plane, with Descartes's metaphor that A Mathematical Function Is a Curve in the Cartesian Plan
• Newton's geometric characterization of calculus in terms of a sequence of secants with a tangent as a limit
• The understanding of a curve in terms of natural continuity and motion

Central to the geometric paradigm was the notion of a curve. An excellent characterization was given by James Pierpont, professor of Mathematics at Yale, in addressing the American Mathematical Society. Pierpont correctly and insightfully listed what a cognitive scientist would now refer to as eight "prototypical" properties of what he called a *curve*—that is, a line in three-dimensional space that is either straight or curved (Pierpont, 1899, p. 397):

Pierpont's Prototypical Properties of a Curve

1. It can be generated by the motion of a point.
2. It is continuous.
3. It has a tangent.
4. It has a length.
5. When closed, it forms the complete boundary of a region.
6. This region has an area.
7. A curve is not a surface.
8. It is formed by the intersection of two surfaces.

Take note of the year: 1899—a quarter of a century after the successful launching of the discretization program by Dedekind and Weierstrass. Pierpont, as we shall see shortly, was both attracted and repulsed by the program—attracted by the prospect of increased rigor, repulsed by the loss of the tools of visualization and spatial intuition. Pierpont's reaction was like that of many of his colleagues. The discretization program was still controversial, but it had won.

The geometric paradigm had formed the intuitive basis of what mathematicians of that era took for granted as being *the* characterization of algebraic and trigonometric functions. It characterized the conceptual framework of mathematicians of the day; the assumption was that all functions should fit this paradigm.

Discretization and "Triumph" over the Monsters

Part of the reason why discretization won out over the geometric paradigm had to do with what were called *monster functions*. What made a monster function a "monster" was that it violated the geometric paradigm, which characterized the idea of what a function should be. The monsters violated the very conceptual framework that mathematicians had learned to think in. That framework was very largely geometric, while the functions themselves were functions from numbers to numbers. In the functions themselves, there was no inherent geometry.

The geometric paradigm allowed mathematicians of the day to use their geometric intuitions in the study of functions from numbers to numbers. The monsters were cases where the geometric paradigm failed. The failure was falsely, we believe, attributed to the failure of "intuition"—geometric intuition. The "failure" was one of expecting too much of a particular paradigm—one that works perfectly well for the cases it was designed to work for, but which cannot be generalized beyond those. What made the geometric paradigm appear to be a failure was a search for *essence*—for a single unitary characterization of all numerical functions.

1872 Again, Elsewhere in Germany: Weierstrass and the Arithmetization of Calculus

The person usually credited with devising an acceptable replacement for the geometric paradigm was Karl Weierstrass, whose ideas from lectures are cited by H. E. Heine, in his *Elemente* in 1872. This was the same year that Dedekind published his *Continuity and Irrational Numbers.* Weierstrass's theory was an updated version of Cauchy's earlier account (which had certain flaws).

Weierstrass, like Dedekind and Cauchy, sought to eliminate all geometry from the study of numbers and functions mapping numbers onto numbers, including derivatives and integrals in calculus. This project required many changes in the geometric paradigm, essentially the same changes proposed by Dedekind for pretty much the same reasons:

- Natural continuity had to be eliminated from the concepts of space, planes, lines, curves, and geometric figures. Geometry had to be reconceptualized in terms of sets of discrete points, which were in turn to be conceptualized purely in terms of numbers: points on a line as individual numbers, points in a plane as pairs of numbers, points in n-dimensional space as n-tuples of numbers.
- The idea of a function as a curve defined in terms of the motion of a point had to be completely replaced. There could be no motion, no direction, no "approaching" a point. All these ideas had to be reconceptualized in purely static terms using only real numbers.
- The geometric idea of "approaching a limit" had to be replaced by static constraints on numbers alone, with no geometry and no motion. This is necessary for characterizing calculus purely in terms of arithmetic.
- Continuity for space was to be reconceptualized as "continuity" for numbers.
- Continuous functions also had to be reconceptualized purely in terms of numbers.

- Calculus had to be reformulated without either geometric secants and tangents or infinitesimals. Only the real numbers could be used.

How Weierstrass Did It

How do you accomplish such a complete paradigm shift? How do you utterly change from one mode of thought to another, from one system of concepts to another? The answer is, via conceptual metaphor.

Implicit in Weierstrass's theory are some of the conceptual metaphors and blends we have already discussed: Spaces Are Sets of Points, Numbers Are Points on a Line, The Number-Line blend (including the folk theory of measurement and magnitude), Continuity for a Number Line Is Arithmetic Gaplessness. Like Dedekind, Weierstrass metaphorically conceptualized the real numbers as aggregates of rationals, though in a way that was technically different than Dedekind's cuts or Cantor's infinite intersections of nested intervals (see Boyer, 1968, p. 606; also see Chapter 9).

Weierstrass's achievement was to add to these conceptual metaphors and blends a new metaphor of his own, a way to metaphorically reconceptualize the continuity of naturally continuous space itself and continuous functions in it. That is, he had to come up with a remarkable metaphor, one conceptualizing the continuous in terms of the discrete. Via this metaphor, he had to be able to *redefine* what "continuity" could mean for sets of discrete elements.

Continuity

Weierstrass's redefinition of continuity for a function over sets of discrete numbers was perhaps his greatest metaphorical magic trick. Think of the job he had to do. He had to eliminate one of our most basic concepts—the natural continuity of a trajectory of motion—and replace it with a concept that involved no motion (just logical conditions), no continuous space (just discrete entities), no points (just numbers), with functions that are not curves in a plane (but, rather, sets of ordered pairs of numbers). And in the process he had to convince the mathematical community that this was the *real* concept of "continuity."

Weierstrass's strategy was this:

Look at continuous functions, conceptualized as motions through continuous space. Look at what is true of discrete point-locations along those continuous paths of motion. Try to reformulate in discrete, static terms what is true of the discrete point-locations on the continuous paths. Using the Number-Line blend, think of points converging to a spatial limit as numbers converging to a numerical limit.

Weierstrass saw that, with continuous functions, points that are close in the domain of the function tend to be close in the range of the function, though what counts as "close" might be different in the range than in the domain, and different at different values for the function. His idea was this:

- For a given point *f*(*a*) in the range, pick a number that defines a standard of closeness for *f*(*a*). Call it epsilon.
- Then, for the corresponding point *a* in the domain, pick another number that defines a standard of closeness for *a*. Call it delta.
- Suppose the function is naturally continuous over point-locations in naturally continuous space. Then points that are "close" to *f*(*a*) (within distance epsilon) in the range will have been mapped from points that are "close" (within a corresponding distance delta) in the domain.

That is, for every standard of closeness in the range (epsilon) there will be corresponding standard of closeness in the domain (delta), and this will be true for every point-location *f*(*a*) in the range and its corresponding point-location *a* in the domain. This is where the classical quantification statement "For every epsilon, there exists a delta" comes from and why it is important.

The Weierstrass program for discretizing continuity came with a minimal standard of adequacy: Begin with all of the prototypical classical continuous curves in the geometric paradigm. Replace these curves with sets consisting of the discrete points on each curve. To be minimally adequate, Weierstrass's new "definition" of "continuity" had to assign the judgment "continuous" to each set of points corresponding to a classical naturally continuous curve.

Weierstrass now used the Number-Line blend, in which there is a correspondence between point-locations and numbers, based on the metaphor Numbers Are Points on a Line. In the Number-Line blend, every choice of *point-locations* corresponds to a choice of *numbers*, and every *distance between point-locations* corresponds to an arithmetic *difference between numbers*. Using these metaphorical correspondences, Weierstrass observed that the idea of continuity itself, the defining property of naturally continuous space, could be stated in entirely discrete terms, using numbers in place of their corresponding point-locations in the Number-Line blend.

Weierstrass's observation that naturally continuous functions *preserved closeness* for point-locations, via the Number-Line blend, is characterized by the following conceptual metaphor:

WEIERSTRASS'S CONTINUITY METAPHOR

Source Domain NUMBERS		*Target Domain* NATURALLY CONTINUOUS SPACE
Discrete numbers	→	Discrete point-locations
Sets of numbers	→	Curves
Numbers in sets of numbers	→	Point-locations on continuous curves
Functions seen as mapping discrete numbers within sets to discrete numbers within sets	→	Functions seen as mapping points on continuous curves to points on continuous curves
Preservation of numerical closeness for functions over discrete numbers	→	Continuity for functions over continuous curves

Weierstrass's Implicit Use of the BMI: Epsilon-Discs and Infinite Nested Intervals

The discerning reader will note that these ideas can be conceptualized using the Epsilon-Disc frame and the special case of the BMI for infinite nested intervals, discussed in Chapter 12. Each number defines a standard of closeness, which is the radius of an epsilon disc. The full range of standards of closeness for the real numbers is infinite. This infinity can be conceptualized via the BMI using the special case of the infinite nesting property (see Chapter 12). The connection between the epsilon and delta standards of closeness can be conceptualized by a correspondence between the epsilon discs on the $f(x)$-axis and those on the x-axis.

Thus, using the Number-Line blend, the Epsilon frame, and two linked instances of the BMI for infinite nested intervals, we have the cognitive apparatus necessary to characterize Weierstrass's basic idea for redefining continuity in terms of arithmetic alone.

Limits for Functions

To arithmetize calculus, Weierstrass next had to arithmetize the concept of a limit for functions. We have already seen continuity defined in terms of linked standards of closeness across the x- and $f(x)$-axes. And we have seen how these can be conceptualized using the linked version of the BMI for infinite nested intervals. With this apparatus, the notion of a limit follows immediately.

THE LIMIT L OF A FUNCTION $f(x)$ AS x APPROACHES a

- An infinite set of nested discs with L as a center on the $f(x)$-axis.
- A corresponding infinite set of nested discs with a as a center on the x-axis.
- The radius of each disc on the $f(x)$-axis is $|L - \varepsilon_i|$.
- The radius of each corresponding disc on the x-axis is $|a - \delta_i|$.
- As each radius gets smaller on the $f(x)$ axis, the corresponding radius gets smaller on the x-axis.

We are now in a position to see how these ideas were written down by Weierstrass in common mathematical notation.

THE TRANSLATION INTO COMMON MATHEMATICAL NOTATION

- The infinite nested sequence of discs around L is covered by the quantifier expression "For every $\varepsilon > 0$."
- The correspondence across the two nested sequences of discs is given by the expression "There exists a $\delta > 0$" following the above quantifier expression.
- The radius of each disc on the $f(x)$-axis around L is "$|f(x) - L|$."
- The radius of a corresponding disc on the x-axis around a is "$|x - a|$".

Given these translations, the concepts just discussed can be understood as covered by the following formal definition.

WEIERSTRASS'S NOTATION FOR HIS CONCEPT OF LIMITS

Let a function f be defined on an open interval containing a, except possibly at a itself, and let L be a real number. The statement

$$\lim_{x \to a} f(x) = L$$

means that for every $\varepsilon > 0$, there exists a $\delta > 0$, such that if $0 < |x - a| < \delta$, then $|f(x) - L| < \varepsilon$.

Given this notation for Weierstrass's concept of the limit, we can use it to provide a statement expressing Weierstrass's notion of Continuity in common mathematical notation.

WEIERSTRASS'S NOTATION FOR HIS CONCEPT OF CONTINUITY

A function f is continuous at a number a if the following three conditions are satisfied:

1. f is defined on an open interval containing a,
2. $\lim_{x \to a} f(x)$ exists, and
3. $\lim_{x \to a} f(x) = f(a)$.

Here Weierstrass is exploiting the linkage between his concept of continuity and his concept of a limit. Note that the "function f [is] defined on an open set of real numbers containing a, except possibly at a itself." The assumption of an open set of real numbers presupposes something crucial. It builds in a hidden case of the BMI, as we saw in Chapter 12. The reason is that the open set of reals containing a has the nested interval property with respect to a. In other words, an open set containing a presupposes the concept of an infinite nested sequence of epsilon discs around a—conceptualized, of course, via the BMI.

The Result

Weierstrass's characterization of limits for functions and of continuity is a statement "redefining" what was meant by limits and continuity in the geometric paradigm. In the redefinition, there is no motion, no time, no "approach," no "limit." There are only numbers. "a" and "L" are numbers; "x," "δ" and "ε" are variables over numbers. There are no points or variables over points. Instead of motion through space, we have only sets of pairs of numbers. In the notation for the concept of a limit, there is no motion, no approaching, no endpoint; there is only a static logical constraint on numbers, a constraint of the form:

For every u, there exists a v such that, if $G(v)$, then $H(u)$.

Take this statement and make the following replacements: $u = \varepsilon$, $v = \delta$, $G(v) = 0 < |x - a| < v$, and $H(u) = |f(x) - L| < u$. The result will be Weierstrass's discretized account of continuity. Just numbers and logic. There is nothing here from geometry—no points, no lines, no plane, no secants or tangents. In place of the natural spatial continuity of the line, there is just the numerically gapless set of real numbers. There are no infinitesimals, just static real numbers. The function is not a curve in the Cartesian plane; it is just a set of ordered pairs of real numbers.

How does Weierstrass replace Newton's geometric idea of approaching a limit, if there is no motion, no space, no approach, and no limit? Weierstrass's work of genius was a clever use of conceptual metaphor.

Having reconceptualized "limit" and "continuity" in purely arithmetic terms without any geometry, Weierstrass had no difficulty characterizing a derivative in the same way. The Newtonian derivative was characterized using the metaphor that Instantaneous Speed Is Average Distance Traveled over an Infinitely Small Interval of Time. Newton had conceptualized this metaphorically in geometric terms: Taking the derivative (the instantaneous rate of change) of a function as a tangent, he conceptualized the tangent as a secant cutting across an infinitely small arc of the curve. That metaphor allowed him to calculate the tangent as the limit of an infinite series of values of the secant. Newton had arithmetized the tangent as:

$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

Given his new metaphor for approaching a limit, Weierstrass simply appropriated Newton's arithmetization of the idea of average change over an infinitely small time, jettisoning the idea and keeping the arithmetic.

Weierstrass's static, purely arithmetic idea of the derivative is as follows.

> The derivative function $y = f'(x)$ is a set of pairs (x, y) meeting the condition: For every neighborhood of size ε around y, there is a neighborhood of size δ around x such that $\frac{f(x + \delta) - f(x)}{\delta}$ is in the neighborhood of size ε around y.

In other words, $y = f'(x)$ if y *preserves closeness* to $\frac{f(x + \delta) - f(x)}{\delta}$, for arbitrarily small values of δ. "Neighborhood" here is defined in arithmetic and set-theoretic terms as an open set of numbers. The ingenuity here is that y is defined not directly but in terms of what it is indefinitely close to.

Here again, there is no geometry, no motion, no curves, no secants or tangents, no approaching, and no limit. There are just numbers and logical constraints on numbers. $\frac{f(x + \delta) - f(x)}{\delta}$ is the arithmetic expression of Newton's idea of instantaneous change as average change over an infinitely small interval. But the idea and the geometrization are gone. Only the arithmetic is left. The ε-δ condition expresses *preservation of closeness*.

The Hidden Geometry in Weierstrass's Arithmetization

Weierstrass is given credit for having arithmetized calculus—and in a very important way, he did. Yet the geometry is still there. Calculus is about the concept of change. Change, in conceptual systems around the world as well as in classical calculus, is conceptualized in terms of motion. Motion in mathematics is conceptualized in terms of the ratio of distance to time. Time, in turn, is metaphorically conceptualized in terms of distance. Ratios are conceptualized in terms of the arithmetic operation of division. The Newton-Leibniz metaphor for instantaneous change is average change of distance over an interval (time conceptualized as distance) of an infinitely small size. This is arithmetized by the expression $\frac{f(x+\delta)-f(x)}{\delta}$. It is this collection of metaphors that links the concept of change to $\frac{f(x+\delta)-f(x)}{\delta}$.

In these metaphors, there is implicit geometry: the ratio of *distance* to *time*, where time is itself conceptualized metaphorically as *distance*. If mathematics is taken to include the ideas that arithmetic expresses—that is, if calculus is taken to be about something—namely, change—then Weierstrass did not eliminate geometry at all. From the conceptual perspective, he just hid it. From the perspective of mathematical idea analysis, no one *could* eliminate the geometry metaphorically implicit in the very concept of change in classical mathematics.

Weierstrass and the Monsters

It is now time to discuss the monsters, those functions that did not fit the geometric paradigm. It was the monsters that created the urgent need for a new nongeometric paradigm—a need that the work of Dedekind and Weierstrass filled. Weierstrass was said to have "tamed the monsters."

What makes a function a monster? Suppose you believed with Descartes, Newton, and Euler that a function could be characterized in terms of naturally continuous geometric curves in a classical Cartesian plane. Your understanding of a function would then be characterized by your understanding of curves. As Pierpont pointed out, the prototypical curve has the following properties:

1. It can be generated by the motion of a point.
2. It is continuous.
3. It has a tangent.
4. It has a length.
5. When closed, it forms the complete boundary of a region.
6. This region has an area.
7. A curve is not a surface.
8. It is formed by the intersection of two surfaces.

Monsters are functions that fail to have *all* these properties.

Imagine yourself assuming that a function *was* a curve in the Cartesian plane with all these properties. Then imagine being confronted by the monster functions we are about to describe. What we will do in each case is

- describe one or two monsters.
- show how each does not fit the geometric paradigm.
- describe what Weierstrass's paradigm says about each one.

It is often said that Weierstrass's account of "continuity" is (1) a formalization and (2) a generalization of "continuity" in the geometric paradigm—that is, natural continuity. We will also ask, in each case, whether this is true.

Here Come the Monsters

Monster 1: $f(x) = \begin{cases} \sin(1/x) & \text{for } x \neq 0 \\ 0 & \text{for } x = 0 \end{cases}$

Monster 2: $f(x) = \begin{cases} x \sin(1/x) & \text{for } x \neq 0 \\ 0 & \text{for } x = 0 \end{cases}$

These are represented in Figures 14.1 and 14.2, respectively.

As x gets close to zero, $1/x$ grows to infinity (when x is positive) and minus infinity (when x is negative). As x approaches zero and $1/x$ approaches positive or negative infinity, both monster functions oscillate with indefinitely increasing frequency. That is, there are more and more oscillations in the functions as x approaches zero from either the positive or negative side.

Monster 1 oscillates between –1 and 1 all the way up to (but not including) zero (see Figure 14.1). Monster 2 is more constrained. It oscillates between two straight lines each at a 45-degree angle from the x-axis and intersecting at the origin. But since x gets progressively smaller as it approaches zero, the function goes through progressively smaller and smaller oscillations (see Figure 14.2).

What happens when these monsters confront the geometric paradigm, in which all functions are supposed to be curves? Do monsters (1) and (2) have all the properties (1 through 8) of prototypical curves? Suppose we ask if the curves for these functions can be generated by the motion of a point. The answer is no. Such a point would have to be moving in a direction at every point, including zero. But as each function approaches zero, it oscillates—that is, changes direction—more and more often. The definition of the function has been gerrymandered a bit to artificially include a value at $x = 0$, because $1/x$ would not be

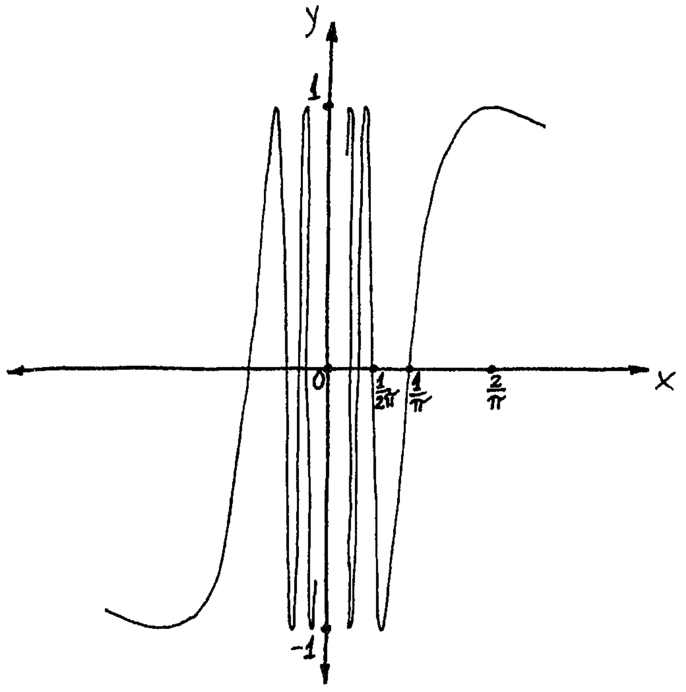


FIGURE 14.1 The graph of the function $f(x) = \sin(1/x)$.

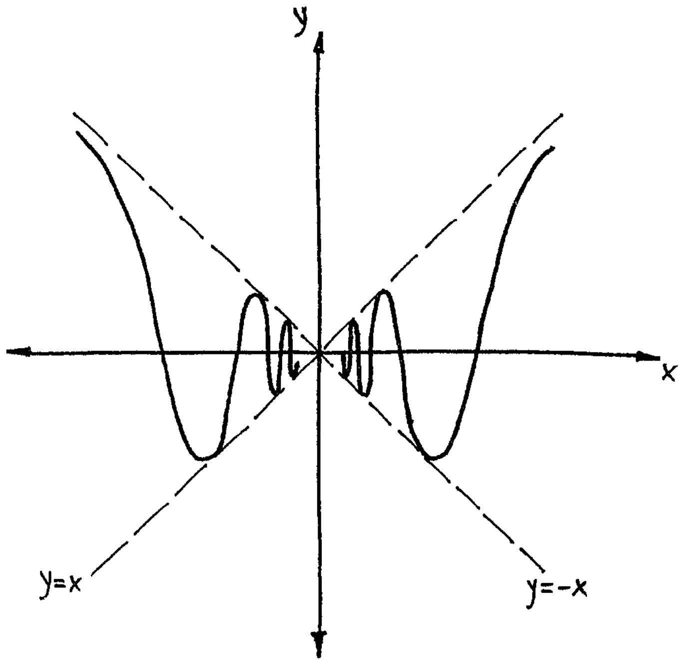


FIGURE 14.2 The graph of the function $f(x) = x \sin(1/x)$.

defined for $x = 0$. The nongerrymandered part of the function does not "pass" through the origin at all. It just keeps getting closer and closer, changing direction more and more, but never reaching the origin. The added point makes it seem as if the function "passes" through the origin, but there is no motion. It is just a point stuck in there.

At the origin—the gerrymandered point—the oscillation, or change of direction, approaches infinity. What direction is the curve going in as it "reaches" the origin? There is no well-defined direction. Does it approach from above or below? There is no answer, because *at* the origin it isn't coming from any one direction. For this reason, the function cannot be a curve traced by the motion of a point.

Since no direction can be assigned to the function as it passes through zero, the function cannot be said to have any fixed tangent at zero. Additionally, consider an arc of each function in a region including zero—for example, an arc between $x = -0.1$ and $x = 0.1$. What is the length of such an arc? Because the function oscillates infinitely, the arc does not have a fixed length. In short, monsters (1) and (2) fail to have properties 1, 3, and 4 of curves.

What about property 2? Are the monster functions "continuous"? Continuity for a prototypical curve means "natural continuity." If natural continuity is characterized, as Euler assumed, by the motion of a point, then the answer is no. Since neither monster function can be so characterized, neither is naturally continuous. Thus, they also fail to have property 2. Since they lack half the properties of prototypical curves, and since properties 5 and 6 don't apply to them, they lack four out of six of the relevant properties of curves. That is what makes them monsters from the perspective of the geometric paradigm.

From the perspective of the Weierstrass paradigm, the purely arithmetic definitions of "continuity" (i.e., preservation of closeness) and differentiability apply to such cases exactly as they would to any other function. In the geometric paradigm, neither case has a derivative at the origin, since neither has a tangent there. On the Weierstrass account, there is no derivative at zero, either (the limit does not exist), but for an arithmetic reason: In neighborhoods close to zero, the value of $\frac{\sin\frac{1}{x+\delta} - \sin\frac{1}{x}}{\delta}$ and $\frac{(x+\delta)\cdot\sin\frac{1}{x+\delta} - x\cdot\sin\frac{1}{x}}{\delta}$ vary so wildly that closeness to any value at zero is not preserved. The condition "$y = f'(x)$, if for every $\varepsilon > 0$, there exists a $\delta > 0$, such that $\left|\frac{f(x+\delta) - f(x)}{\delta} - y\right| < \varepsilon$" is not met for any y at $x = 0$.

But what about Weierstrass's continuity—that is, preservation of closeness? Here the two monsters differ. Monster 1 does not preserve closeness at the origin. If you pick an epsilon less than, say, 1/2, there will be no delta near zero that will keep $f(x)$ within the value 1/2. The reason is that as x approaches zero, $f(x)$ oscillates between 1 and –1 with indefinitely increasing frequency—and so cannot be held to within the value of 1/2 when x is anywhere near zero. Since preservation of closeness is what Weierstrass means by "continuity," Monster 1 is not Weierstrass-continuous. In this case, preservation of closeness matches natural continuity: Both are violated by Monster 1.

Monster 2 is very different for Weierstrass. Because $f(x)$ gets progressively smaller as x approaches zero, it does preserve closeness at zero. If you pick some number epsilon much less than 1, then for every delta less than epsilon, the value of $f(x)$ for Monster 2 will stay within epsilon. Since preservation of closeness is Weierstrass's metaphor for continuity, his "definition of continuity" designates Monster 2 as "continuous" by virtue of preserving closeness.

This does not mean that Monster 2 is naturally continuous while Monster 1 is not. Neither is naturally continuous. It means only that Monster 2 preserves closeness while Monster 1 does not.

What are we to make of this? In the geometric paradigm, Monster 2 is not continuous. In the Weierstrass paradigm, Monster 2 *is* "continuous." There is no contradiction here, just different concepts that have confusingly been given the same name. However, there are two very important theoretical morals here.

- Weierstrass "continuity" is not just a formalization of the "vague" concept of natural continuity. If one were just a formalization of the other, it could not be the case that one clearly holds but the other clearly doesn't. The two concepts of "continuity" are simply very different concepts. They have different cognitive structure.

- Weierstrass "continuity" is not a generalization over the concept of natural continuity. It cannot be a generalization in this case, since Monster 2 *is* Weierstrass "continuous" but *is not* naturally continuous.

"Space-Filling" Monsters

We discussed so-called space-filling curves, like the Hilbert curve, in Chapter 12 (see Figure 12.4). The Hilbert curve—let us call it Monster 3—and other curves like it were high on the list of monsters.

Monster 3 violates properties 1, 2, 3, 4, 7, and 8 of the prototypical curve. (Properties 5 and 6 are inapplicable.) It cannot be generated by a moving point. The reason is that such a point must move in some direction at each point of the "curve," but in the Hilbert curve there is never a particular direction that the curve is going in. It changes direction at every point. Therefore, it can have no single direction of motion at any point.

Since the Hilbert curve cannot be generated by a moving point, it is not naturally continuous in Euler's sense. Moreover, it has no direction at any point, nor does it have a tangent at any point. Additionally, it has no specific finite length for any arc. Every arc is of indefinitely long length. Thus it fails properties 1 through 4.

Suppose you take the geometric paradigm to include only real numbers and not infinitesimals. Since the Hilbert curve maps onto all the real points in the square, it maps onto the points on a surface, in violation of property 7. For the same reason, it cannot be formed by an intersection of surfaces, in violation of property 8. This monster is therefore even more monstrous than Monsters 1 and 2. You can see why anyone brought up to think in terms of the geometric paradigm would find such cases "pathological."

As we have just seen, the Hilbert space-filling curve is not a naturally continuous one-dimensional curve; therefore, it is not "continuous" in the geometric paradigm. But what about the Weierstrass paradigm? Does the Hilbert curve preserve closeness?

The Hilbert curve does preserve closeness. Points close to one another on the unit interval get mapped onto points close to one another in the unit square. Again, a monster that is not continuous in the geometric paradigm *is* "continuous" in the Weierstrass paradigm, showing once more that the two notions of continuity are different concepts. Weierstrass continuity is neither a formalization nor a generalization of natural continuity.

More Monsters

Here are two well-known functions, which we will call Monsters 4 and 5.

$$\text{Monster 4: } f(x) = \begin{cases} 1 \text{ if } x \text{ is irrational} \\ 0 \text{ if } x \text{ is rational} \end{cases}$$

$$\text{Monster 5: } g(x) = \begin{cases} 1 \text{ if } x \text{ is irrational} \\ \text{undefined if } x \text{ is rational} \end{cases}$$

The geometric paradigm can make no sense of these functions, since they are nothing at all like a prototypical curve. As far as natural continuity is concerned, Monster 4 is nowhere naturally continuous. At every point p, there is a point infinitely close to p where the function "jumps." Monster 5, which is undefined over the rationals, is defined over a domain that is not naturally continuous—the irrational numbers with "holes" at the rational points. The graph of Monster 5 will therefore also have "holes" at rational points and not be naturally continuous.

In the Weierstrass paradigm, Monster 4 is not "continuous," because it does not preserve closeness. No matter what point x you pick on the real-number line, if you take an $\varepsilon = 1/2$ there will be no δ that will keep the value of f within $1/2$.

But Monster 5 *is* "continuous" in the Weierstrass paradigm. The function g does preserve closeness. Even though the function is now defined over a naturally discontinuous domain, the values of the function do not "jump"; rather, they are always the same. Thus, for any ε, $g(x)$ will stay within ε of 1, no matter what δ you pick.

Again, this example shows that the two notions of "continuity" are very different concepts. Weierstrass "continuity" is shown, once more, to be neither a formalization nor a generalization of natural continuity.

Pierpont's Address

The discretization program is more than a century old. But whereas it is now largely accepted without question, it was very much an issue of debate a century ago. It is instructive to see how an influential American mathematician argued in favor of the Weierstrass paradigm a hundred years in the past.

In 1899, Professor Pierpont felt compelled to address the American Mathematical Society to try to convince his colleagues of the necessity of Weierstrass's arithmetization of calculus. The list we gave of the eight properties of curves is taken from Pierpont's address. Pierpont presented the list and went through his discussion of various monster functions, arguing, as we have done, that they cannot make sense as curves in the geometric paradigm. He also invoked the rigor myth again and again.

> The notions arising from our intuitions are vague and incomplete . . . The practice of intuitionists of supplementing their analytical reasoning at any moment by arguments drawn from intuition cannot therefore be justified. (Pierpont, 1899, p. 405)

He expressed the idea that only arithmetization is rigorous, while geometric intuitions are not suitable for "secure foundations."

> There are, however, a few standards which we shall all gladly recognize when it becomes desirable to place a great theory on the securest foundations possible . . . What can be proved should be proved. In attempting to carry out conscientiously this program, analysts have been forced to arithmetize their science. (p. 395)

The idea that numbers do not involve intuition comes through clearly.

> The quantities we deal with are numbers; their existence and laws rest on an arithmetic and not on an intuitional basis . . . and therefore, if we are endeavoring to secure the most perfect form of demonstrations, it must be wholly arithmetical. (p. 397)

What was interesting about Pierpont is that he knew better. He knew that ideas are necessary in mathematics and that one cannot, *within mathematics*, rigorously put ideas into symbols. The reason is that ideas are in our minds; even mathematical ideas are not entities within formal mathematics, and there is no branch of mathematics that concerns ideas. The link between mathematical formalisms using symbols and the ideas they are to represent is part of the study of the mind—part of cognitive science, not part of mathematics. Formalisms using symbols have to be understood, and the study of that understanding is outside mathematics per se. Pierpont understood this:

> From our intuition we have the notions of curves, surfaces, continuity, etc. . . . No one can show that the arithmetic formulations are *coextensive* with their corresponding intuitional concepts. (pp. 400–401; original emphasis)

As a result he felt tension between this wisdom and the appeal of the arithmetization of calculus.

Pierpont was torn. He understood that mathematics was irrevocably about ideas, but he could not resist the vision of total rigor offered by the arithmetization program:

> The mathematician of today, trained in the school of Weierstrass, is fond of speaking of his science as "die absolut klare Wissenschaft" [the absolutely clear science]. Any attempts to drag in metaphysical speculations are resented with indignant energy. With almost painful emotions, he looks back at the sorry mixture of metaphysics and mathematics which was so common in the last century and at the beginning of this. The analysis of today is indeed a transparent science built up on the simple notion of number, its truths are the most solidly established in the whole range of human knowledge. It is, however, not to be overlooked that the price paid for this clearness is appalling; it is total separation from the world of our senses. (p. 406)

This is a remarkable passage. Pierpont knows what is going to happen when mathematics comes to be conceived of mainly—or only—as being about "rigorous" formalism. Mathematical ideas—he uses the unfortunate term "intuition," which misleadingly suggests vagueness and lack of rigor—not only will be downplayed but will be seen as the enemy, a form of mathematical evil to be fought and overcome. He can't help himself. He has been converted and comes down on the side of "rigor," but he sees the cost and it is "appalling."

What Weierstrass Accomplished

Weierstrass's accomplishment was enormous. He achieved the remetaphorization of a major part of mathematics. With Dedekind, Cantor, and others, he played a major role in constructing the metaphorical worldview of most contemporary mathematicians. His work was pivotal in getting the following collection of metaphors accepted as the norm:

- Spaces Are Sets of Points.
- Points on a Line Are Numbers.
- Points in an n-dimensional Space Are n-tuples of Numbers
- Functions Are Ordered Pairs of Numbers.
- Continuity for a Line Is Numerical Gaplessness
- Continuity for a Function Is Preservation of Closeness

Where would contemporary mathematics be without these metaphors?

Weierstrass was also responsible for the demise of the respectability of the geometric paradigm. As Pierpont understood, that was a shame. The problem was not that the geometric paradigm was vague. The problem was that its limits were not properly understood. It is a wonderful tool for understanding naturally continuous, everywhere differentiable functions—functions that can be conceptualized as curves in a plane. Its advantage is that *for such functions* it permits visualization and spatial understanding to inform arithmetic and vice versa.

The existence of the monster functions shows that the geometric paradigm is not a fully general intellectual tool for studying *all* functions. It does not follow that it should be scrapped; all that follows is that it needs to be supplemented and its limits well understood.

One of the wonderful effects of the Weierstrass revolution was that it made mathematicians break down the relevant properties of functions and spaces and study them independently: continuity, differentiability, connectedness, compactness, and so on. Weierstrass's preservation of closeness concept (mistakenly called "continuity") is at the heart of modern topology. The grand opening up of twentieth-century mathematics had everything to do with Weierstrass's remetaphorization of analysis.

Continuity and Its Opposite, Discreteness

The German mathematician Hermann Weyl (1885–1955) noted in his classic work *The Continuum* (1987, p. 24):

> We must point out that, in spite of Dedekind, Cantor, and Weierstrass, the great task which has been facing us since the Pythagorean discovery of the irrationals remains today as unfinished as ever; that is, the continuity given to us immediately by intuition (in the flow of time and in motion) has yet to be grasped mathematically as a totality of discrete "stages" in accordance with that part of its content which can be conceptualized in an "exact" way.

Why should it, as Weyl says, be a "task" of mathematics to "grasp . . . the continuity given to us immediately by intuition (in the flow of time and motion) . . . as a totality of discrete 'stages'"? Why does mathematics have to understand the continuous in terms of the discrete?

Each attempt to understand the continuous in terms of the discrete is necessarily metaphorical—an attempt to understand one kind of thing in terms of another kind of thing. Indeed, it is an attempt to understand one kind of thing—the naturally continuous continuum—in terms of its very opposite—the discrete. We find it strange that it should be seen as a central task of mathematics to provide a metaphorical characterization of the continuum in terms of its opposite. Any such metaphor is bound to miss aspects of what the continuum is, and miss quite a bit.

If "the great task" is to provide absolute, literal foundations for mathematics, then the attempt to conceptualize the continuous in terms of the discrete is self-defeating. First, such foundations cannot be literal; they can only be metaphorical. Second, as Weyl himself says, only "part of its content" can be conceptualized discretely. The rest must be left out. If Weyl is right, the task cannot be accomplished.

We believe there is a greater task: understanding mathematical ideas.