

Choosing Theories is Like Choosing Socks¹

David Odden
OSU

1. What are we talking about?

The purpose of **science**, broadly construed, is to gain knowledge of the nature of reality, and render it in an objective form so that it can be understood, evaluated, and acted upon by others. The central activities of science which are designed to increase knowledge are the construction of descriptions, and the construction of theories. Another domain of human endeavor which pertains to knowledge is philosophy and mathematics (these will be collapsed into one domain here labeled ‘philosophy’), which consider ideas generated by the mind without requiring reference to reality. Questions raised in philosophy are decided largely or exclusively on the basis of reason and declarations about the meaning of words — whether or not this is a good thing is outside the scope of this paper.

The distinction between science and philosophy is not necessarily a strict one. Science depends on philosophy, which provides essential methods for systematizing and evaluating statements about reality. Science provides the specialized statement “this *is* the nature of reality, these *are* the principles governing reality”; philosophy provides a methodological framework that mediates between what is, and the understanding of what is (i.e. it explains what notions like “proof” and “evidence” are, and how you decide if you have “proof” or “evidence”). Equally, we could speak of “empirical science” which is the study of reality, vs. “philosophical science” which is the study of ideas and not the details of reality. Physics, to use the contemporary term, used to be called “natural philosophy” (see Newton’s *Principia*).

A **description** in science is a set of statements laying out the known attributes of a group of entities (including conceptual entities such as numbers, ethical systems, etc.) and events.² Example: having encountered a particular collection of eight reindeer, I notice that all members of the group have big elk-like antlers, thick light beige fur, big feet, horse-like faces, are ruminants, live in the Arctic, and stand three feet tall. That description would be wrong if one of these reindeer did not have the particular attributes, e.g. had white fur, were six feet tall, or had gills.

A **theory** is a set of statements giving the attributes of all entities in a conceptually unified set, which includes attributes and entities not yet experienced. A theory of reindeer would be that they have thick fur, big feet, horse-like faces, are ruminants, have low tolerance for heat, and when adult stand between three and four feet tall. Only those reindeer which live near the North Pole are three feet tall as adults (all reindeer which avoid drowning or becoming lunch attain the height of 3 feet as juveniles), only certain reindeer have light beige fur, and while most reindeer have antlers, not all do and it is therefore not an essential attribute. Living in the Arctic (wherever, exactly, that is) is also not an essential attribute of reindeer, even though all non-zoo reindeer do: what is crucial

¹ You can’t let your mom do it for you all your life; also, it’s hard to find one that isn’t full of holes. I thank Mary Bradshaw, John Frampton, Charles Reiss and students in my classes for comments on earlier versions of this paper.

² I will avoid cluttering the discussion by not saying “entities and events” at every point. Events are action relations between entities, and the study of events can be safely subsumed under the study of entities. Complete knowledge of entities requires knowledge of their actions.

to reindeer survival is a particular temperature range that makes them unsuited for life in the African savannah or Central Ohio, and the Arctic is the only place with the required climate. The theory would be wrong if any reindeer did not have these specific attributes (the reader is advised that this is an amateur theory of reindeer given for illustrative purposes only, and it is unimportant whether this theory of reindeer actually happens to be correct). A theory is a special kind of description of observations — actual and potential — about attributes of all members of a definable group. The crucial point is that it is designed to extend beyond observations already made, i.e. it is more than a list of known facts: it includes a **prediction**, a saying in advance of observation what *will* be seen.

Because the concept of “observation” is important, the status of observations in science must be clarified. Etymologically, “observation” derives from a Latin verb translated into English as “observe”, i.e. “watch” which is a variety of visual perception. The importance of selecting the verb “observe”, rather than “see” or “glimpse” in describing what scientists do is that the act of observing is more active, and involves a higher level of cognitive involvement than “seeing”, “looking” or “glimpsing”. Please note that while visual perception is an extremely important sense for comprehending the universe, it is not the only mode of perception useful in gaining knowledge. When we speak of “observations”, we do not limit ourselves to visual evidence, but include all modes of sensory perception.

It is important to consider the connection between cognition and sense organs in observation. The reason why this is important is that all knowledge must be rooted in an observation of the universe, and observations involve sense organs.³ We will take the case of visual observation for exemplification, but the principles extend with small modifications to all modes of perception. Suppose there were an apple a few feet in front of you. Light reflects off of the apple in a particular way. The essential physical principle is that because of the chemical composition, i.e. atomic structure, of the skin of the apple, particular light wavelengths are absorbed, and others are reflected. The reflected light enters the eye, and photons are absorbed by cells in the retina, causing an electrical discharge. These electrical impulses travel up the optic nerve, being passed to the brain, and they result in raw sensations. The brain organises this information, resulting in particular percepts such as color, shape and distance. At higher levels, this information is integrated into information such as “that is an apple”. Thus, the act of observing an apple involves the interaction between the sensory apparatus and the world outside of us, via a chain of events transmitting information in accordance with physical law.

The importance of cognition to observation cannot be understated, especially insofar as observation involves consciously grasping a relation between a fact of the universe and its description. When a person perceives that there is an apple on the table, a significant amount of cognitive processing is required to arrive at that knowledge. The eyes deliver to the brain a set of electrical impulses that constitute the sensory ‘raw material’. Low-level cognition is required to apprehend simple specifics such as color, size and shape. Higher-level cognition is required to arrive at the conclusion that the object satisfies the essential criteria for applehood, and in some instances, conscious reasoning may be needed to decide whether the object is an apple, for example if the size, shape or color are atypical. A crabapple may not be recognised as an instance of “apple” the first time someone sees one because of the atypical size; an Asian pear-apple (which is not an apple) may be mis-

³ Do not be confused about the difference between “is entirely reducible to” and “is rooted / grounded / based in observation” — implicitly, using reasoning to derive a conclusion *about* those observations.

perceived as an apple because its shape is very similar to that of an apple; certain varieties of apple have a shape reminiscent of nectarines and peaches. Thus the statement “I observed an apple” carries with it some cognitive processing. More processing underlies other observations, for example saying that you have observed a Cloning Anemone (*anthopleura elegantissima*) or a Giant Green Anemone (*anthopleura xanthogrammica*) in some location, because to do so, you must know and explicitly apply the scientific criteria for distinguishing the species. In a nutshell, you unquestionably *see* the critter, but may or may not *understand* that it is an *anthopleura elegantissima*, not to mention an animal. As far as science is concerned, the interest is not just with the simple act of perceiving some thing, but with understanding what has been observed — not just seeing that there is a roundish reddish object, but correctly identifying that it is a particular kind of apple.

There is little significance, from the perspective of scientific value, to the distinction which is sometimes made between “direct observation” and “indirect observation”. Written words in a book at one point size, which can easily be seen by the unaided eye, do not suddenly change their essential character when printed at a smaller point size that requires a magnifying lens to read. Because of their size, amoebae and bacteria cannot be “directly” observed (in the intuitive sense), they can only be seen through a microscope. Various instruments make it possible to extend the range of human senses, allowing us to observe more things. A **direct** observation could be defined as one made without the aid of any instrument not a part of the human body. An **indirect** observation would then be one made with the aid of an instrument. But this distinction is of no particular use to understanding the nature of data. Indirect observations have the same validity as direct observations, because both are grounded in the transmission of light (or sound, etc.) from the observed object to the brain, in accordance with physical laws. There is no reason to put higher value on “direct” observations simply because it uses no manufactured instrumentation, thus we will only speak of observations. There *is* a reason to put higher value on more **reliable** observations. Whether or not instrumentation introduces precision or greater possibility of error depends on the nature of the instrument, the object being observed, and the nature of the observer’s sensory apparatus.

Finally, there is the question whether it is important to distinguish observation of an entity or process, and the inference that an entity or process must exist because of its effect on observed entities. An example of this type would be the discovery of the planet Neptune, which (under some historical accounts) was posited theoretically by Le Verrier in 1846 on the basis that the hypothetical existence of such a planet could explain perturbations in the orbit of Uranus, and was observed telescopically later that year.⁴ An observation that some object has a particular temperature, e.g. 200° F, would accordingly be classified not as an observation, but as an inference based on an observed effect on something else, in this case an effect on a column of colored alcohol. Similarly, measurement of sound amplitude and frequency would have to be classified not as observations, but

⁴ Neptune was first seen by Galileo in 1612, but he and others after him believed it to be a star. The perturbations in Uranian orbit were known in the 18th century and were thought to be measurement errors. In 1821, Bouvard suggested that there might be an “outside influence” that could explain the perturbation. In 1845, the existence and location of the planet Neptune was postulated by a Cambridge undergraduate, John Adams, but he did not publish his results, awaiting visual confirmation. Le Verrier published his conclusions in 1846, and soon after at the Berlin Observatory, Neptune was seen telescopically and identified as a planet based on Le Verrier’s calculations. The idea of studying the irregular motion of Uranus was suggested to Le Verrier by his teacher François Arago: it is unknown whether Arago may have had in mind the postulation of a planet.

as inferences based on effects on other entities, namely the instruments used to measure the properties. This dichotomy applied too liberally will lead to the empty conclusion that all knowledge is inferential (there is always some cause-effect chain), which makes the concept of inference definitionally useless: “inferential knowledge” would no longer contrast with anything stronger.

There is a difference between *seeing* an object, and claiming that it *probably* exists because a mathematical computation tells you it is possible and could explain an anomaly. The most important factor which distinguishes these cases is not the tool used for gaining knowledge, or the role of inference, but rather the possibility of error in gathering and interpreting the data. There is next to no chance of being wrong when one directly sees an apple; there is little chance of being wrong when one bases a claim that a planet exists on the fact that one has seen it in a telescope; there is a non-negligible chance of being wrong when one bases the claim solely on a calculation of patterns in light oscillations from a distant star, because your knowledge of underlying physical laws might be in error.⁵

2. What do we require of a theory?

The goal of science is acquiring knowledge of reality, and any component of science is evaluated by that standard. Since more than one theory can exist regarding a portion of reality, we need a basis for evaluating the value of one theory compared to another, and for choosing one theory and rejecting all others, because there is only one reality. We therefore require certain things of theories. The most important requirement of a theory is:

Well-Formedness — statements of a theory must have specific knowable meanings.

This is an inviolable law. Any string of gibberish which does not satisfy this principle is to be instantly rejected. Adherence to this axiom prevents the following kind of tragedy.

- A: I have a new theory of the universe: the universe is a large marshmallow ramiculated on a spline of extended propositions.
B: Professor, that is bull pucky. My middle finger is not composed of marshmallow but it is part of the universe.
A: I don't mean marshmallow in the vulgar sense, I mean it in a sense akin to that applied to energy fields. But it's not the same as an energy field.
B: Well then let me ask, what does 'ramiculated' mean?
A: That's for me to know and you to find out.
B: I see. Then what are 'extended propositions'?

⁵ Hypothesized planets and Newtonian mechanics provides another example. In 1855, Le Verrier — who correctly postulated the existence of Neptune — attempted to explain anomalies in the orbit of Mercury, from the perspective of Newtonian theory, by positing the presence of another planet between the sun and Mercury. Lescarbault claimed to have observed the object on March 26, 1859. Le Verrier then named this putative planet “Vulcan”, and calculated a specific diameter and orbit. Based on analysis of 50 observations, Le Verrier announced a transit of Vulcan across the sun on March 22, 1877, which never happened. An elaborate photographic campaign during the solar eclipse of 1905 failed to detect Vulcan. Mercury's orbital anomaly was explained in 1916 on the basis of relativistic effects of the sun, due to effects of the Sun's mass bending space-time. Thus, Le Verrier was correct only 50% of the time, while Adams, who made no such erroneous prediction, was correct 100% of the time.

A: An example would be the class of abstract entities defining the properties of attributes, integrated over the range of properties of properties.

B: Since a ‘spline’ is a *geometric* object, what does it mean to have a spline of ‘extended propositions’, which I’m guessing is about propositions?

A: Because, once you understand ‘extended propositions’ you will understand that they are undefinable atomic objects, so there is no contradiction.

B: But you just contradicted what you just said about ‘extended propositions’ ten seconds ago.

A: Yes, but that was then and this is now.

At which point, B commits suicide (or homicide, depending on temperament). To prevent death (either physical or mental), we will not discuss meaningless gibberish. The student is warned that due to little-understood failures in the human cognitive apparatus, you may actually encounter meaningless gibberish which is labeled as a ‘theory’, just as one may encounter inorganic refuse mislabeled as ‘art’. Mixtures of gibberish and meaningfulness are possible, and actually quite common. With respect to theory evaluation, the most efficient and ruthless destruction of a theory is, or should be, a demonstration that the statements of the theory lack known meaning. The topic of the interplay between gibberish and theory will be taken up in a separate paper (some year).

Theories, i.e. real theories and not gibberish, are evaluated according to certain principles, which state the desirable attributes of a theory, which can be reduced to the following hierarchy, the most important being first:

Correctness — the observations entailed by the theory correspond to reality.

Comprehensiveness — the theory covers all essential attributes that define the set of entities taken to be in the domain of the theory, and that domain is the greatest possible.

Aesthetics — the statements of the theory are simple.

Violation of Correctness is violation of the Law of Identity, i.e. $X_{\text{theory}} = \sim X_{\text{reality}}$ or $\sim X_{\text{theory}} = X_{\text{reality}}$. Specific instances would be that the theory predicts that object X should exist but we find that X does not actually exist, or that object X should not exist but we find that X does in fact exist;⁶ or the theory could predict that event X should happen and yet X never happens, or vice versa. Comprehensiveness values a theory that, for example, covers both molecules and ions over a theory that just covers molecules, or a theory that covers all human languages over a theory that covers just languages with more than 30 million speakers. Another linguistic example of Comprehensiveness would be the relationship between a phonological theory which states the form of rules for syllable-construction, and a general theory of all types of phonological rule: the former theory is less comprehensive, hence inferior.⁷ Note, however, that a less comprehensive theory

⁶ What might appear to be analogous is a theory which predicts that X *can* exist. Such a theory is not refuted by the observation that the entity is not seen at the moment. As will be discussed later, it is much harder to assess the correctness of a statement which says nothing more than “the existence of entity X would not result in a logical contradiction, given the axioms of this theory”.

⁷ A further use of this principle is that it precludes unreasoned and ad hoc exclusion of specific observations relevant to a domain. If I were presented with a carnivorous reindeer, which refutes my herbivorous theory of reindeer, this principle would prohibit me from excluding that specific reindeer by saying “Well, yes, but I didn’t mean *that* reindeer”.

could actually be correct. It is possible that there really do need to be separate theories of “syllable rules” and “segmental rules”, and that a theory unifying the two would be factually incorrect. The principle Aesthetics weighs against formally stupid theories (typically and like the case of porn, this is hard to define but you know it when you see it, especially after viewing a lot of it). A system encumbered by statements like “*a* can appear anywhere, except if it is preceded by *c* and followed by *d* then there must be a *q* as well” would be rated low in aesthetics.

The Correct Theory maximises the value G (Goodness) in the equation:

$$G = (\text{Truth} \times \text{Value}(\text{Truth})) + (\text{Beauty} \times \text{Value}(\text{Beauty})) + (\text{Ambition} \times \text{Value}(\text{Ambition}))$$

The mystical aspect of theory-evaluation is that the value-coefficients remain unknown (in addition, we remain clueless about how to quantify truth, beauty and ambition, which is the unit for measuring comprehensiveness).

The ultimate goal of science is a theory which completely describes reality, covers all aspects of reality, and is extremely simple in its statements. Since such a theory does not exist, imperfections with respect to empirical adequacy, elegance and scope of theory must be accepted on the road to this theory. There exist, in very many domains of knowledge, multiple theories which claim to be the best description of reality. Unless the theories happen to be the same in essentials and differ only in how they are presented — are strongly equivalent notational variants — at most one of these theories is correct, since there is only one reality. Scientists must therefore determine which theories are to be rejected (and, by process of elimination, which may be accepted), and this leads to the requirement of theory evaluation, with all of its related problems.

Since scientific beliefs are supposed to be rational (based on reason) and should not be arbitrary, the decision to accept one theory over another as a statement of reality must be based on reason, not whim. This is why it is important to consider explicitly what the basis is for picking one theory rather than another. The fundamental philosophical struggle in theory evaluation is determining how the desiderata for a theory should be balanced. Most working scientists, who profess a profound interest in truth to the exclusion of all else, would probably place Correctness at the top of a hierarchy of values.⁸ This would mean that of two theories, some theory A (which matches the facts better but is less elegant) would be superior to theory B (which does not match the facts well but is elegant) and scientists should prefer A over B, as long as A really is closer to correct. If, on the other hand, Aesthetics were taken to be more important, then theory B would be the superior theory. In fact, one type of study of knowledge — philosophy in its manifestations mathematics and formal logic — often places Aesthetics at the top of the hierarchy. Insofar as linguistics has significant roots in mathematics, it is common to find Occam’s Razor invoked on a daily basis, sometimes ahead of Correctness.

⁸ The goal of this essay is to describe the ideal, philosophically based (rational) situation in the conduct of science. It would be an error to claim that all scientists always do what they should do. For that reason, there will be little reference to the history of science herein. Whether or not all scientists, or most scientists, or just scientists in a particular domain of knowledge, or just a very few good scientists, actually conduct their business according to the principles discussed here does not impinge on the validity of the discussion as a statement of the philosophical principles pertaining to science. In a worst case scenario, this essay is a prescription for the rational conduct of science.

The hierarchy of values used in theory comparison should not be random or accepted by administrative fiat or social pressure, but should be selected by reason. The purpose of theory comparison is to identify the best theory. The best theory is the one which best suits the **goal** of having a theory, so what is the purpose of theory? To answer this, we return to, and expand on, the principles stated earlier. Knowledge is the result of the mind conceptualizing reality; it is not a wish that reality should be such and such a way, or a statement that it would be nicer or more convenient if reality were some particular way. It is a statement ‘this is what is’. A theory is a form of knowledge, and is presented as the condensation of reality into concepts. For this reason, **the highest value for a theory must be its accuracy as a representation of reality, and empirical correctness must be the highest evaluative principle.**⁹ The secondary value should be a theory’s efficiency in integrating and differentiating concepts about reality. However, second place is valuable as well.

The distinction between metaphysics (the nature of reality) and epistemology (the nature of knowledge) is important. Knowledge of reality does not yet cover all of reality (a condition, were it to exist, known as *omniscience*), and scientists must recognize this fact. The highest desideratum for human knowledge is *certainty*, that is, the absence of a reason to disbelieve, as to the nature of reality. This implies that neither the intrinsic content of the knowledge which we hold (e.g. “this is a general theory of matter”), nor the shape of that knowledge (e.g. “this is a numerical theory of X”), is as important as the objectively warranted understanding that this knowledge is correct, as a representation (description, summary) of reality.

3. What Is The Truth?

We must now discuss the acquisition of knowledge via “proof” versus “argument”. There is significant variation in the world as to how the term “proof” is used. One trend, following the usage of formal logicians and paid philosophers (and others inspired by them) is that a proof is a formal deductive means of obtaining a result, given unquestioned and often arbitrary premises, using certain methods of symbol-manipulation. The other, adopted by the guy in the street (and other locations), is roughly “convincing demonstration”. This essay will treat certain terms as technical concepts, ceding ground to the paid philosophers, in order to avoid pointless and orthogonal debate over what statements you can “prove” based on different definitions of “prove”. The term “proof” will be considered only in the formal deductive sense alluded to above, without any implication that this interpretation is intrinsically correct except “in the context of symbolic logic and how it’s usually used by certain philosophers”. A proof boils down to saying “Given the meaning of the terms A and B, then if we assume P as an axiom (unquestioned assumption) where P contains A and B, some other statement Q must also be true”.

⁹ What constitutes accuracy needs to be specified in more detail. Contrast two clocks, one which is 10 minutes slow and another stuck at 12:00. The nonfunctioning clock has a higher correspondence to reality than the slow clock for 20 minutes every 12 hours, and is therefore occasionally more accurate. This does not render the nonfunctioning clock more valuable as an instrument of knowledge — it is totally useless for increasing knowledge. The accuracy of a theory is not the best results obtained by the theory, but is the totality of results. While the nonfunctioning clock is dead on at 12:00, it is as inaccurate as is logically possible at 6:00. The inaccuracy of the slow clock is 10 minutes; the precise degree of inaccuracy of the nonfunctioning clock at a given time varies as a function of when the clock is checked, but averages out to 180 minutes.

The principles of symbolic logic will not be discussed at length, but are quite intuitive: they boil down to stating the meanings of certain basic words like “and”, “not” and so on. Propositions have two values, T (“true”) and F (“false”),¹⁰ so a proposition symbolized A (which could stand for any number of arbitrary statements such as “my car is blue”, “snow is white”, “snow is blue” or “reindeer eat lichen”) is either T, or it is F. The negation operator (symbolized as \neg) reverses logical value, so when A is T, $\neg A$ is F and when A is F, $\neg A$ is T. Two propositions can be combined to form another proposition using certain logical connectives. The proposition “A and B” is itself a proposition which asserts that the two component propositions A, B are both T. The statement “A and B” is T, only if the two statements A, B are each T; if A is F, or B is F, or if both A and B are F, then the compound proposition “A and B” is also F. This proposition is often symbolized as “A&B”, “A·B” or “A \wedge B”. It is well known that ordinary language usage of “and” is not identical to logical conjunction. The logical proposition “(Sue had a baby) & (Sue got married)” is identical to the logical proposition “(Sue got married) & (Sue had a baby)”, but the ordinary English statement “Sue got married and she had a baby” has different temporal implications (ergo truth conditions) from “Sue had a baby and got married”.

Similarly, propositions can be combined with the connector “or”, so that “A or B” (symbolized “A \vee B” or “A|B”, known as a disjunction) is T if at least one of these propositions is T. Again, natural language use of “or” does not translate trivially into logical disjunction: the sentence “Give me your wallet or I’ll cut you” is not equivalent to the disjunction “(Give me your wallet) | (I’ll cut you)”, which at best has the semantic interpretation “(You are obliged to give me your wallet) | (I will cut you)”.¹¹ A third connector is “if and only if”, often written as “iff” and symbolized as “ \equiv ” as in “A \equiv B”, which is T only when A and B are both T, or A and B are both F — that is, A and B have the identical value (this one is not strictly necessary, since “A \equiv B” is equivalent to “(A&B) \vee ($\neg A$ & $\neg B$)”).

The last connector is “if-then”, symbolized with “ \supset ”. The statement “A \supset B” is F if A is T and B is F, otherwise the statement is T. Thus, “if it is raining, the ground is wet” would be a true statement if it is indeed raining and the ground is wet, and it would be false if it is raining and the ground is not wet. This statement is also true if it is not raining and the ground is dry; and the statement is true if it is not raining and the ground is wet. In other words, the truth value of the consequent B does not matter if the antecedent A has the value F, in terms of the validity of the relation itself.¹² This being a point which confuses many people, it must be emphasised that the proposition “A \supset B” states a relation between propositions, A and B. It is not a statement about B alone is true:

¹⁰ We will not consider models of logic which have more than two values; the most important other possibility is a three-valued logic which allows T, F, and unknown/undetermined.

¹¹ So, even if you fail to give the speaker your wallet, the ‘obligation’ exists, therefore the first conjunct is still true. The reader is warned that in nearly all contexts, “Give me your wallet or I’ll cut you” is said with the intention of conveying the meaning “I will cut you if and only if you do not give me your wallet”, but furthermore you might get cut no matter what you do.

¹² If you want to say “If A is true then B must be true, and if A is false then B must be false”, that could be expressed with the biconditional “A iff B”. Note that “A \equiv B” is equivalent to “(A \supset B) & (B \supset A)”. The proposition “A \supset B” is equivalent to (B \vee \neg A), hence “if-then” is not a primitive operator.

if you want to know if the ground is wet (is B true), you directly test B itself. When it is not raining, then the fact that the ground does not happen to be wet would not invalidate the relation. When there is no rain, the ground may be wet because someone is running the hose outside; or it can be dry because there is no source of water. The truth of the relationship between rain and wetness is not affected by the possibility that there could be other causes of wetness, and is certainly not affected by dry ground in a drought.

Two other logical operators are the quantifiers “all” symbolized as “ \forall ”, and “there exists” or “some” symbolized “ \exists ”. The proposition $\forall x(A(x))$ says that all entities have the property A, and is false if any entity lacks the property A.¹³ The proposition $\exists x(A(x))$ states that at least one entity has the property A, and is false if there is no entity with this property. A more common use of quantifiers is in combination with if-then, as a way of stating that members of a definable group also have some other property. Thus the proposition $\forall x(A(x) \supset B(x))$ says “all entities which have property A also have property B”, literally “for all entities x , if x is an A, it must also be a B”. An example would be “all humans are mortal”, and the proposition would be false if some entity were found that was human and immortal. The proposition $\exists x(A(x) \& B(x))$ says “there exists some entity which has property A also has property B”, for example, “some humans are bald” which is true because there is an individual, Martin, who is bald: the proposition would be false if no bald humans existed.

A formal proof begins with assumptions, so you can never deduce the statement “Socrates is a mortal” from nothing. Given two assumptions, you can prove the mortality of Socrates. First, you need to assume the truth of the proposition “all humans are mortal”. One may rightly wonder what justifies making that assumption, even though it’s not typically called into question. Second, you need to assume the truth of the proposition “Socrates is a human”. (Again, what exactly justifies making that assumption? But we are not questioning these assumptions, just saying that conventionally, you have to make the assumptions to get anywhere). Given these two assumptions, the third proposition “Socrates is mortal” can be derived, following laws of symbolic logic. Because our purpose is not a full exposition of symbolic logic, we skip over crucial details. We will accept certain laws of inference (whether they are axioms or theorems derived by observing how reasoning works), one of which says “If a proposition of the form $A \supset B$ is true, and if the proposition A is true, then the proposition B is true” — a principle of inference classically called *modus ponens*; or “any proposition which is true of a universally quantified variable remains true when any specific entity is substituted for the variable”, due to a rule known as “universal instantiation”. Our assumptions can be more precisely stated as “For all entities x , if x is a human, then x is a mortal” and “Socrates is a human”. Given the first assumption and by substituting the entity Socrates for x , we derive the statement “if Socrates is a human, then Socrates is a mortal”. Given the rule of inference *modus ponens*, we can combine this proposition and the proposition “Socrates is a human” to formally derive “Socrates is a mortal”.

Symbolic logic is not concerned with establishing the truth of the initial assumptions “All humans are mortal” and “Socrates is a human”. Formal logic simply says that *if* we make two such true assumptions, the third proposition “Socrates is a mortal” is also true. It would be in the domain

¹³ This kind of statement is not likely to be encountered very often, since the only two propositions that hold of all entities are Existence (“all entities — things which exist — exist”) and Identity (“all entities are identical to themselves”).

of empirical science to establish that the assumptions are correct. Alternatively, some assumptions might be initially adopted as arbitrarily true, used as a means to prove other propositions, and then later the initial assumptions might themselves be proven to be true — this approach is common in mathematics. Proofs are rarely applied to determining the mortality of dead Greek philosophers, and are generally reserved for mathematical statements (which might include statements about formal properties of grammatical systems). Thus you can probably prove that the number 7 is between 3 and 10, given such things as the definitions of the numbers themselves and the meaning of “between”, although a full formal proof is daunting.

A typical characterization of a (formal) proof is that it is a sequence of axioms and definitions, plus other statements which derive exclusively by applying a very few laws of inference to axioms, definitions, and other statements already proven to be true. A proof is thus an entity which can be entirely validated within the mind looking just at the form of statements, without requiring appeal to evidence drawn from experience and without depending on the nature of the universe for validation. Do not take this to imply that fundamental concepts such as “and”, “or”, “all”, “exists” and so on could be developed by a mind in lieu of a human being who is conscious of the universe, or that cognition is possible without any experience of the universe. Rather, this implies that *given* the existence of the mind, *given* these fundamental concepts, and *given* the ability to define and to reason, the formal correctness of a formal proof can be determined without further reference to entities outside the mind.

The proposition being proven may be extremely abstract with no clear relation to real objects (e.g. “a sheaf of cohomologies defined on Riemannian space”), or it can actually be quite firmly grounded in real objects and practical questions. An example of the latter is the Four-Color Theorem, which is the claim that for any map, you can always color regions so that no two adjacent regions have the same color, using at most four colors. The practical use of this fact is that if it is true, a map-manufacturer requires fewer colors to be in stock, thus reducing cost while producing maps with a desirable attribute. Though stated in terms of “maps”, “colors” and “regions” which are tangible objects, first-order percepts and simple abstractions over space, in the context of the proof they are given abstract definitions which do not depend on contingent properties of reality, such as the fact that “color” is a percept arising in the brain due to the processing of neural impulses which result from the absorption of photons by certain cells in the human eye; the concept “map” is generalised in such a way that the proof is valid whether the “map” in question is on paper, cardboard, sheepskin, or is a holographic projection — indeed, the “map” itself does not have to exist in any sense, except as a concept in the mind of a person. This conjecture was formally deduced to be true, which means, *given* the laws of formal logic, the definitions of essential terms (“map” and so on), the statement must be true. No experimentation or observation of reality is required to reason to the conclusion that this is a true statement.¹⁴ The interest of “proof” in the formal sense lies in the fact that *if* one starts with valid assumptions and *if* one uses valid rules of inference and *if* one cor-

¹⁴ The proof took over one hundred years to arrive at, depended on a program which required 1200 hours on a supercomputer to do an exhaustive analysis of around 1500 discrete cases, took years to check for errors, and had to be published as a separate book. No human to this date fully “comprehends” the proof, but in principle if someone wanted to dedicate their entire intellect to the proof it could (presumably) be done purely in the mind.

rectly applies the rules of inference, *then* one can be utterly certain of the conclusion. That is a lot of ifs, however.

4. What is a Scientific Fact / Claim / Hypothesis?

Typically, scientific questions do not lend themselves to formal (symbolic) proof. Consider a simple scientific proposition “there are reindeer in Longyearbyen”. Suppose you provide definitions of “reindeer” and “Longyearbyen”.¹⁵ Then you might think it would be a simple matter to prove that indeed there are reindeer in Longyearbyen, by observing at least one reindeer in this location. If you *assume* that there is an entity x in Longyearbyen and x is a reindeer, then you can (formally) prove “there are reindeer in Longyearbyen”, because this proposition (in ordinary English) translates into the statement “there exists an entity x which is in Longyearbyen and which is a reindeer”, which is what we assumed from the start. In otherwords, to logically prove the conclusion, you have to take it as an *assumption* that the conclusion is true. Technically this is a proof, of the form “ $A \supset A$; A ; $\therefore A$ ”, but it is a supremely uninformative proof. Assuming the truth of the claim which is to be proven (usually in the guise of another essentially, or in this case literally, identical claim) is known as “begging the question”, a logical fallacy vile enough to have garnered the Latin appellation *petitio principii*.

Can the crucial component propositions themselves be formally proven, that there is an entity x in Longyearbyen or that x is a reindeer? That would render the proof more interesting. The simplest approach is to look at the entity, apply the criteria for reindeerhood, and determine whether the entity is actually a reindeer. Imagine that you are in Longyearbyen and see a three-foot tall quadruped with thick light beige fur, big feet, a horse-like face, elk-like antlers, eating lichen. This entity seems to be a reindeer; you check your field guide for reindeer and see that the entity which you are staring at looks exactly like the picture, and the picture is identified by renown authorities in mammalian biology as being a reindeer. Are you correct in concluding that the animal must *logically* be a reindeer? No, because it is “logically possible” (that is, imaginable, without violating the laws of symbolic logic) that it simply looks to you like a reindeer, but is in fact another animal. You object “but the only animals in Longyearbyen are birds, mice, foxes and polar bears, and it can’t be a bird, mouse, fox or polar bear”. Nice objection, but *logically* incomplete: how do you know that it can’t *logically* be a polar bear in a clever disguise? How do you know that your assumed list of creatures of Longyearbyen is correct? How do you know it is an animal at all? Perhaps it is a wax-works dummy, a paper maché model, or a very clever holographic projection. How do you know you are seeing anything at all — it might be a dream brought on by skillful brainwashing techniques. How do you know that you are not just a brain in a vat, imagining that the universe exists? The point is that it is always possible to say that “it could logically be something else”, and as long as the entity could *logically* be something else, you cannot *logically* prove this proposition.

¹⁵ A definition must be given in terms of previously defined terms, not undefined terms. Therefore, before defining “Longyearbyen”, one has to define a system of coordinates such as latitude and longitude. Unfortunately, this is usually done by reference to formally undefined locations such as “the Equator”, “the North Pole”, and “the Greenwich Meridian”, leading to an infinite regress of trying to define terms in other terms. Since the enterprise of formally proving the existence of reindeer at Longyearbyen is doomed to failure, we will not pursue this particular aspect of the failure.

This may seem like a bit (or load) of silliness, but it is rooted in an actual ancient philosophical system. Skepticism is a school of philosophy founded by Pyrrho, a Greek philosopher and priest of the -4th century, in reaction to an existing belief by philosophers that one could actually have knowledge. His claim was that all knowledge is uncertain, and one can never know the truth. The Skeptic believes that nothing is true or false, it only appears to be so; similarly, nothing is good or evil, only opinion or law makes it so. Once this is understood, the rational person declines to make choices and lapses into apathy. Action is founded on belief, belief is delusion, so lack of activity is the proper goal for the sage.¹⁶ Skeptic philosophy comes in many varieties. The most radical holds that you can never know anything (even supposedly logically provable propositions, because you can never know that inferences really follow the laws of logic and don't just seem to, etc.), and entirely denies reality. Less noxious beliefs allow that the universe exists but we can't really trust our senses. An actually useful skeptical view is that not every claim on the market or in books should be taken to be correct, and justification is required to hold a claim to be true.

Science obeys the laws of formal logic, but does not (generally) concern itself with the formal proofs of mathematics, even though it is often interested in the useful methods discovered by mathematics. Even in uncommon cases where a scientific theory can be made so explicit that it is amenable to formal proofs — as in physics — all that a proof can do is demonstrate that a conclusion follows from axioms. A proof cannot show whether the axioms are valid, or whether the deduced conclusion corresponds to reality: it simply shows a particular relation between ideas. Sometimes people speak of “scientific proof”, as in giving a scientific proof that a five pound weight will fall to earth if you drop it (this requires a number of calculations involving the mass of the weight, the mass of the earth, equations involving gravity and inertia, also calculations involving density of air). This invokes the other concept of “proof”, one that is sufficiently different from the concept of proof used in mathematics and logic that we will use a different term — “argument” — to characterise the epistemological objects (the form of knowledge) which science creates. **Science seeks reasoned justification grounded in reality, not formal proof that considers only word meaning and symbol manipulation.**

In evaluating a statement like “a five pound weight will fall to earth if you drop it”, we want to know if it describes reality, i.e. “is true”. The bare mind cannot determine the truth of all statements about reality, since that would require the brain to contain, in advance of experience, a complete representation of the universe. In other words, the statement cannot be formally proven given just what we are born with. A scientific claim requires the application of epistemological principles to things we perceive (“color”, “length” and so on being examples) and to concepts derived from percepts, and the scientist's concern with logic (in the broad sense) is with valid means of relating the mind to the external world. That relation is not one-to-one: we make errors, and operate within a particular range of precision (one which can be modified by judicious application of technology).

¹⁶It is perilous to summarize millenia of philosophy in a sentence, but the essence of the Skeptic view is the same as the mainstream traditions of India, rooted in the Upaniṣads via the concept of *māyā* — the ability of god to project the illusion of existence, realised in Hindu and Buddhist tradition in the principle that the universe is illusion. This is almost certainly not a coincidence: Pyrrho spent time in India during Alexander's invasion, where he learned of Indian philosophy at the height of the development of these beliefs in India. Pyrrho is reputed to be the only skeptic philosopher who actually believed his own theory strongly enough to have let it influence his daily life, to the point that his students had to follow him everywhere so that he would not walk off of cliffs. That is the story, though I'm skeptical.

Human sensory organs operate in accordance with the laws of reality, but not with infinite precision. Eyes are structured to react to only a small portion of the electromagnetic spectrum, between 3500 and 7500 ångström, so we cannot see gamma radiation, or even ultraviolet (which some bugs can); the ability to resolve details is a little less than a millimeter; human hearing does not allow signals below about 50 Hz or above 15,000 Hz to be heard (bats, on the other hand, can perceive acoustic signals between 1,000 and 120,000 Hz). While instrumentation extends the human sensory apparatus by translating information which is beyond the range of the sensory organs into information that can be perceived (sometimes by very different modalities, e.g. an apparatus that displays sound wave information visually as a set of symbols known as “numbers”), such instrumentation is still not perfect in its resolution of information, and may be very poor in mapping information about the universe. Sensory information about the universe is fallible.

In addition, raw sensations do not themselves equal knowledge, which arises from processing of sensations by human cognitive faculties. First order cognitive processing is sufficiently reliable for basic survival (e.g. the feeling of heat, a 100 p.s.i. pressure on the skin), but as the raw information derived from our sense organs is processed at successively higher levels, more possibilities for error arise, since a significant amount of information processing involves discarding irrelevant information — or *was* that really irrelevant?

Physical contact between human skin and a piece of metal which is glowing orange will yield a specific set of sensations — highly unpleasant ones which by experience humans associate with the percept “being burnt”. After a little bit of experimentation, humans develop a rudimentary theory which systematizes this knowledge, so we know that objects which glow bright orange are very probably quite hot, and contact between human skin and hot objects nearly always results in a particular unpleasant sensation. You also learn of the additional, non-visual sensation which correlates with this pain, namely the feeling of heat on the skin just moments before the hand grasps the glowing metal (a generally more reliable cue to heat, as 2 year olds learn, but dependent on data which is available only at much greater proximity than the visual evidence). We cannot “mathematically prove” that grasping a piece of metal which is glowing orange will cause pain; and yet it is a rule which people confidently rely on.

How could you be wrong in assuming that a glowing piece of metal will cause pain if touched? You could be wrong about whether the object is glowing. If the light coming from an object is strong compared to the ambient light, the percept of “glowing” will be strong (and more reliable), whereas if the light is not so strong, the percept of “glowing” will be weak. You could therefore be wrong about whether an object is glowing (as opposed to reflecting). You could also be wrong in your theory of glowing objects, because actually not all objects which emit light are correspondingly hot — bioluminescent light sources are not “hot”. If you have never experienced such a light source, you would have no reason to suspect the existence of non-hot glowing objects.

We answer scientific question on the basis of the evidence, and given different kinds of evidence, we would say that the object is definitely hot, or very probably hot, or probably hot, or simply, it might be hot. Such statements are evaluations of the relationship between evidence and fact, where a “proven” statement, is one where all evidence supports the conclusion, and no evidence, even conceptual evidence, points to any alternative. There are a number of ways of expressing the relation between a statement and the evidence for it. Strong predicates such as “know”, “be certain”, “be positive” can be used, or a somewhat weaker evaluation can be used such as “be confident”, or even weaker “think”, “believe”, “infer”, all the way down to “suppose”, “guess”, “con-

jecture” (there are also predicates expressing disbelief). Knowledge is not restricted to a yes/no value, it is a scalar statement — a probability — of our judgement that in reality, the claim is a fact. If you see a piece of metal come out of a blast furnace, the metal is glowing an intense white color, the light surrounding the metal is distorted (“heat shimmer”), and you feel heat coming from the direction of the hunk of metal, you will assign a very high probability that the metal is hot. If a piece of metal comes out of a freezer, gives off no detectable light, and has icicles hanging from it, you would assign a very high probability to the truth of the proposition “the object is cold”. Based on other evidence, e.g. lack of a furnace in the vicinity, you may assign a different probability of truth to the proposition.

There are (at least) three related concepts of “probability”. The classical inductive concept of probability pertains to the relation of evidence to the hypothesis that it bears on, and this evidentiary sense is the most important for our purposes. Weather reports are inductive probability statements, being statements of how strong the evidence is that it will rain, and this concept of probability is being invoked when someone says “there is a very high probability that Swahili and Bambara are genetically related languages”. A second concept, combinatoric probability, relates to a priori possibilities, such as the fact that there is a 50% probability of a specific outcome if there are two possible outcomes such as tossing heads vs. tails, a 16.6% probability of a specific outcome if there are six possible outcomes such as a role of a die, and so on. The third concept is statistical probability, which is related to observational frequency (i.e. “the probability of owning two cars and a house in the U.S. is 1 in 8”, meaning that ownership of two cars and a house occurs at the rate of 1 individual for every 8 surveyed). In referring to probability, we are primarily referring to the classical sense, but all three senses have in common that they are about evidence: the third sense may indicate that there is “something going on” for example there are fewer reindeer or laterals than expected (so maybe there is a polar bear eating the reindeers or a rule deleting laterals), and the second sense is relevant because it helps to define expectations (if there are six combinatorial possibilities, finding one of them manifested 16.6% of the time is evidence for the ‘null hypothesis’).

The combinatoric concept of probability bears on science because it relates to the number of logically possible outcomes of an event. In terms of combinatoric probability, the chances of rolling a six with a single die is 1 in 6, and the chances of rolling two consecutive sixes is 1 in 36 (etc.). This knowledge can be acquired without even possessing a die, and can be compared to an actual frequency of an event. Inferences about the nature of reality can be made by comparing the theoretical and observed frequencies of events. You can role a die a dozen times and observe how often six appears. Statistically based inferences are founded on the presumption that events take place at random, and if the observed frequency does not match the theoretical frequency, you should suspect that the event is not random, hence there is something for the theory to explain. Given 12 tries with a die, each outcome should appear twice. Supposing that you do get exactly 2 each of the 6 logically possible results, this is a perfect match between observed and predicted frequency, and would constitute good evidence that the die is acting at random (so perfect a match that you could well expect that the die is not acting at random but is being controlled by external forces). If you get 3 sixes and 1 four (the remainder being 2 each), observed and predicted frequencies do not match, but they are close enough that you can reasonably maintain the hypothesis of randomness — or, you could consider this to be rather weak evidence that the die is weighted.

If you roll 11 sixes and 1 four, the disparity between the two distributions is very high, and it would not be reasonable to maintain the belief that the die is behaving randomly. Via a mathe-

mathematical analysis of the frequency distribution, this extreme disparity between observed and expected frequency can be expressed as a statistical probability, which states that it is highly unlikely that the distribution is random. This then would constitute evidence that the die probably has a small weight inside it, off center in the direction opposite of the six's face (inference about the specific cause requires knowledge of physical laws — the statistical analysis simply tells you that something is wrong). This can be expressed as a statement such as “the die is probably loaded”, or, given this extremely lopsided distribution, “the die is almost certainly loaded”.¹⁷ Non-mathematical probability (evidentiary probability) is typically expressed via a non-mathematical measure such as “virtually guaranteed”, “highly likely”, “very probably”, “likely”, “possibly”, “unlikely”, and so on. Mathematical probabilities are typically expressed as a percentage (“a 1% probability”) or proportion (“1 chance in 100” or, in statistical practice, “.01” meaning “1 chance in 100”).

No magic probability value defines “truth”. Certain values of statistical probability are at present taken by different sciences in western tradition to constitute the sufficient warrant to rationally assert that a proposition is true; these values land somewhere in the range of “one chance in ten” to “one chance in a hundred”, meaning that if there is only one chance in a hundred that the proposition is false, the proposition is therefore considered to be true. The most frequently occurring cutoff point is .05, i.e. one chance in twenty. But in fact scientists generally are not seriously interested in or able to assign meaningful numeric values to evidentiary weight, and it is better to simply say that in light of the evidence, such-and-such a conclusion is “probably true” or “very probably true” or even “a virtual certainty”.

In the case of the falling 5 pound weight, confidence about the outcome is so high that we speak of it as a scientific certainty. This extraordinary confidence is founded on the fact that the physical principles have been very well studied, and the predictions verified innumerable times. The claim that northern Europe was covered with glaciers 15,000 years ago is well supported with material evidence, but not by any photographic or eyewitness evidence, and in no meaningful sense is the claim “experimentally reproducible”. While the evidence that this claim is true does not enjoy the stunningly high levels of confidence that are attributed to hypotheses involving gravity, the evidence is still strong enough that the Ice Age is considered to be a fact. The statement that humans inhabited the New World 10,000 years ago is sufficiently well established, i.e. the evidence is sufficiently strong to warrant the conclusion that this is a true proposition, and we take this to be a fact. The statement that humans inhabited the New World 30,000 years ago also enjoys support, but there is not as much empirical support for this claim: it is reasonably likely to be true, and it is considered to be a hypothesis and not a scientific fact.

5. What Burden Should the Theoretician Shoulder?

Scientists require rational principles for deciding which theory is closest to reality. It must be emphasized that valid comparison can only take place between actual theories. You may wonder about “possible theory”, as opposed to “actual theory”. The difference is simple: an “actual theory” is a theory which has been articulated, where specific decisions as to the structure of the theory have been made. A “possible theory” is just an idea or direction for looking at a problem, which might eventually be articulated into an actual theory. We will only consider comparison of actual theories,

¹⁷This is indeed why the mathematical theory of probability was invented.

since only actual theories carry with them the intellectual commitments which are avoided by “potential theories”.¹⁸

The first step in evaluating theories is establishing the **burden of proof**. Given two theories, which theory should we initially assume to be correct, and which theory is required to “make its case”? Before engaging in complex experimentation aimed at deciding the superiority of one theory over another, it is important to determine the conceptual relation of the two theories, to determine what could count as an argument for one theory over the other. We need to know what responsibility a theory has, with respect to showing that it is superior to a competing theory.

That responsibility, the burden of proof, is to demonstrate the existence of the entities predicted by the theory to exist, which the alternative theory predicts do *not* exist. Consider the simple case where theory A predicts entities {H, I, J} and theory B predicts entities {H, I, J, K, L}. Theory B has the responsibility to prove that {K, L} actually exist. If these entities are shown to exist, theory A is rejected in favor of theory B, and if these entities cannot be shown to exist, theory A must be accepted over theory B. The proponent of theory B must therefore make the case for that theory, by showing that the entities exist. The proponent of A cannot refute theory B by failing or neglecting to demonstrate the existence of the “extra entities” of B, and you cannot directly show that entities do not exist anywhere (because you cannot look everywhere, everywhen, to do the inspection).¹⁹ If my theory of biology does not include unicorns and my competitor’s theory is the same as mine except that it posits unicorns, how can I refute that theory? If I claim to have searched for unicorns but failed in my efforts, my competitor could simply object that I did not try hard enough, that my skepticism was sensed by the unicorn (who therefore avoided me), that there are unicorns out there, and I should just look harder. This would render scientific theorizing worthless, since you could always posit any theory which claims that all sorts of unobserved entities “must exist”, leaving us with no factual basis for choosing any particular theory. The default assumption is that the entities which have been observed are exactly the entities that exist, so the postulation of new entities (in a general sense of “entity”, including events) requires the postulator to provide the evidence that there are such new entities.

It may be possible to refute a theory by the very act of failing to demonstrate the existence of predicted entities, if the theory is sufficiently comprehensive that the conditions for creation and/or observation of the entity can be objectively reproduced. This is the gold standard which science strives to maintain, and may be successful in the physical sciences. Even then, a single failure to achieve a predicted result would not be taken to refute a theory, since it is unlikely that the experimental controls could be so tight that experimental error could be ruled out. If the science is more observational than experimental, e.g. astronomy versus particle physics, or historical linguistics

¹⁸ Appeal to “possible theories” is not a rarity, even in print. The concept is often invoked when a proponent of a theory addresses counterexamples to the theory, and hints that these counterexamples could be disposed of by a minor modification in the theory, saying something off the cuff like “One can imagine a theory where...”.

¹⁹ One could prove formally that some entity *cannot exist*, meaning that given the appropriate shared assumptions and the definitions of terms, the supposed entities are logically contradictory, for example “hornless unicorns” (presuming that they are entities distinct from horses). In that case, we would simply reject the alternative “set of statements” as not even being a theory, i.e. failing the test of Well-Formedness, meaning that the competition isn’t a theory at all, much less a scientifically viable theory.

tics vs. particle physics, it may be impossible to produce entities on demand. In such a case, the opponent of a theory can do little in good faith to determine whether the entity exists, and refutation would lie simply in the realization that a theory predicts an entity which there is no reason to believe exists, such as the planet Vulcan.

At this point, it is important to clarify the concept “exist”, a fairly primitive term. A particular piece of paper may exist now; if it is burned up it will no longer exist (even though the atoms that the paper was built from continue to exist, dispersed in some fashion, the paper itself will not exist). Before the piece of paper was manufactured, the paper did not exist (although the raw materials that were used in the creation of the paper did exist). Thus existence is dependent on time and place: a thing exists in a particular form, at a time and a place, and not just “generally exists”. Historical sciences, such as history, archaeology and paleontology, are highly concerned with specific things at a specific time. Physics on the other hand is concerned with entities irrespective of the historical period, so the question “do electrons exist?” is framed independent of time, thus is not “did electrons exist in 1873?” (though there is a real question as to whether electrons existed as you turn the cosmic clock back far enough, but the question is not whether they existed at that time, rather, it is whether they existed in conjunction with the other physical conditions that prevailed at that time shortly after the universe “began”).

Sciences aim to give statements of reality which are independent of time and place, so we can think of the quest for entities as being timeless and placeless, meaning not just asking “do they exist now”, but also “did they exist then” and “will they exist in the future” or “over there as well”. A more accurate question would be “do, or *can*, these entities exist?”. The objection to this statement of the question is stylistic: it is clumsy, and while the question “does it exist” is emotively strong, demanding a “yes/no” answer, the question “can it exist” tends to evoke responses like “how can we know what is possible?”. The discussion here about whether entities exist is, more literally, whether the entity did, does, or can exist. It is almost certain that element 102 does not currently exist, but it has existed and we know how to create it. A science may stipulate that it is concerned with a specific time period, in which case the question of entities outside the stipulated time-frame is irrelevant. We will not be concerned with time- and place-dependent sciences. What is important to understand is that science is usually interested in claims which are not tied to a specific time or place.

More problematic in the business of comparing theories, and more common, is that the predictions of one theory do not form a subset of the predictions of another theory. Typically, when a new theory is posited, it adds to the discussion new predicted entities which have not yet been observed, and precludes other entities predicted by the competing theory and which have not yet been observed. The burden of proof is thus dispersed: the novel theory is responsible for showing the existence of its predicted but unobserved entities, and the old theory is responsible for showing the existence of its unobserved, predicted entities.

The question of burden of proof plays a central role in linguistics, under the guise of “constrainedness”. Theories are often rejected on the grounds of being “unconstrained”. A theory is “constrained” to the extent that it restricts the class of possible languages, i.e. the theory predicts the existence of fewer entities (in the form of languages, grammars, or both). It is never literally the case that a theory is completely unconstrained, since such a theory would predict that there are no limits at all on language. Rather, one theory can be said to be more constrained than another, meaning that the entities allowed within one theory are a subset of those which are allowed by the

other theory. The substance of the statement that a theory is more constrained is simply that the more constrained theory predicts a subset of the entities of the other theory. Bear in mind, of course, that a theory which says “nothing exists” is highly constrained, but also very wrong.

The distinction between “prediction” and “observation” being essential to theory-evaluation, it is important to remind ourselves of what a “prediction”, since it has a somewhat specialised meaning as applied to theories. We will take as an example a theory that certain things (animals, rocks, morphemes) in a particular conceptual domain fall into 5 groups that are numbered arbitrarily 1, 2, 3, 4 and 5. Analysis of these groups reveals that all of the things in group 1 have attribute A, they are non-B, and they are also C. Things in group 2 have attributes A, B and C; group 3 is made up of things that are non-A, non-B but have attribute C. In the fourth group the things are all A, B and non-C and finally the last group has all and only objects that are non-A, and have properties B and C. This leads us to posit a theory where the essential characteristics of entities are stated in terms of three properties, A, B and C, which can be present or missing.

The problem is that while all existents can be described in this theory in terms of some arrangement of these properties, there are three more configurations of properties which do not correspond to known existents: things that are A, non-B and non-C; things that are non-A, B, and non-C; and things that are non-A, non-B and non-C. The theory is thus said to *predict* three further existents, so characterized. Any theory which posits a set of attributes or entities predicts that these attributes or entities can be freely combined in any logically imagineable fashion, including stringing conglomerations in an infinite random sequence, consistent with any principles of combination which also define the theory. Since infinite sequences of random combinations of entities do not actually exist,²⁰ rules of ‘combination and derivation’ are extremely important in theory building.

By the logic of linguistic formalism, there are eight possible configurations of the two-valued attributes A, B and C. The description of entities is, according to this theory, a conjunction of three properties. Therefore, an entity which is A, B and non-C is the same as an entity which is B, non-C and A — the order in which attributes are presented does not matter. Also, since these attributes are either-or properties, there is no difference between describing an object as being “A, B and C” or “A, B, C and A”, i.e. repetition of an attribute in a description has no consequences. The theory would encapsulate this observation by the very simple statement that “entities are describable by **conjunctions** of properties A, B, C”. An alternative theory could say “entities involve gluing together other entities of the form A, B, C”, i.e. uses **concatenation**. The alternative theory thus claims that AA is distinct from A, and AB is distinct from BA. Thus the theory is not just the statement that the “things” are A, B, C, but also this important principle of how to combine “things”. Yet a third alternative theory would be that the number of atoms that can be combined is not limited, but the concept of “order” is irrelevant, so that AA could be distinct from A and AAA,

²⁰ An exception could be made for strictly conceptual entities such as numbers; but actually there need be no exception in the mental domain. A particular string of digits, for example 78,457,834, is composed of a finite combination of digits, as are all specific numbers. While the mind can grasp the abstract concept of an infinite sequence of digits, it cannot actually grasp an infinite sequence of digits. Qua concept, the concept “infinite sequence of digits” exists, but is itself finite in size — applying a metaphor to concepts. Specific numbers such as π which are composed of an infinite sequence of numbers are conceptualised either via memorization of a finite subpart of the number such as 3.14159 (or 3.1), or by means of the finite definition “the ratio of the circumference to the diameter”.

but AB would not be distinct from BA, and ABAB would be the same as BBAA (or AABB, BAAB, ABBA, etc.) — i.e. entities involve combinations and not permutations of atoms.

A theory either states particular principles regarding combinations and operations, or it makes no statement because, it claims (implicitly or explicitly), the universe actually operates at random in the relevant respect. The question of whether the universe behaves randomly in any respect is a long-standing philosophical and scientific one which will largely be ignored here. A specific point will be emphasized here: failure on the part of scientists to discern a pattern does not itself mean that reality contains no pattern (thus sanctioning the belief that the universe is random), nor does postulation of a pattern by scientists mean that the universe actually contains that pattern in any form (except that the pattern is a concept in the brains of scientists, who exist). The appearance of randomness usually reflects incomplete knowledge, or computations which are so complex as to be uncalculable in a reasonable amount of time given current technology. In such a case, the scientist is well advised to keep trying to find the underlying system, or to simplify the calculations so as to make the computation tractable. If a particular aspect of the universe is in reality unpredictable, the scientist has no business trying to predict that aspect, and switches attention to other aspects of the universe.²¹

We now arrive at a pseudo-conflict of principles. The universe is assumed to be rule-governed, and any aspect of the universe which is not predicted by the theory is random, not rule-governed. There is no contradiction, but phrasing the issue this way gives the outward appearance of a contradiction, for a useful reason. Because contradictions do not exist (A and $\neg A$ cannot both be true), something has to give. Given that anything not predicted by theory is considered to be random, and the fact that the universe is *not* random, theories have to be improved so that all aspects of the universe can in fact be fully predicted. Only with extraordinary evidence that the universe is actually random would we accept a theory that violates the principle of nonrandomness.

Practically speaking, this amounts to stating more precisely what it means to be “a prediction of a theory”. When a theory is presented, the proponent of the theory cannot protect the theory from criticism by saying “I didn’t specifically say that X could also happen”. If a theory says that the primitive elements are A and B, and the recursive operation of concatenation is the only principle of combination in the theory, you cannot then counter “But I didn’t *say* that ABBABA can exist”. By failing to preclude ABBABA, and by providing a mechanism (unrestricted concatenation) that would *allow* this entity to be formed, the theory does indeed predict this entity. If it is necessary to preclude the entity from the set of entities, some additional element of the theory, such as a constraint *ABBABA, must have the consequence that the entity is not created.

Two caveats are needed at this point. First, it is not necessary for a theory to explicitly exclude non-existents. An explicit exclusion is just one way — not even the best — to accomplish this end. Another way is simply to say “these are the only operations allowed in the system”, that is, a list of entities or operations is taken to be exhaustive.²² Second, as will be discussed later in more

²¹ Scientists may abandon a question temporarily if there is no reasonable expectation that a question could be answered given available technology; such questions are often taken up later, as new technology or philosophical approaches develop. Even at the level of quantum mechanics, relatively little has been ceded to the truly random.

²² A theory which states “here is what exists” typically does not bother to explicitly say “and no other things exist”. Unfortunately, this often leads to problems where entities are discovered which are not on the explicit list, and the defender of the theory objects “But I didn’t say these are the only things that exist”. It may be true that the scientist did

detail, a theory can maintain silence about aspects of reality if the theory explicitly is intended to describe just a portion of reality, in which case the theory would not implicitly claim “all random possibilities exist”, in a domain which the theory explicitly excludes from consideration.

We now return to the hypothetical example set out earlier in this section, where the theory in question posits the properties A, B and C, and where three specific combinations which we can symbolize as [A,-B,-C], [-A,B,-C] and [-A,-B,-C] have not been detected: there is a disparity between observation and prediction. There are three responses to this disparity. First, you may accept the prediction and reject the implication that the three unobserved entities do not exist. In following this tack, you are claiming “these entities *do* actually exist; we simply need to observe them”, and then you conduct the search for the missing entities. This approach — shouldering the burden of proof — is the most highly valued (and also riskiest) response to an empirical lacuna. Second, you can accept the prediction and the statement that the entities do not exist. Since this amounts to saying that the theory does not correspond to reality, this means that the theory itself must be rejected. Third, you can reject the prediction by modifying the theory, so as to preclude the missing entities. Essentially, you are rejecting the theory but immediately proposing an alternative which is the same as the original theory except for a specified change in the theory.

The happiest outcome, in the first case, is actually finding the missing entities. Since a theory is a tool for stating what the universe is like (not just what has been observed), a theory which makes a prediction that is subsequently shown to be correct is a valuable tool for extending knowledge: the best theory is one which makes correct predictions. If the predictions are fully verified, the theory which made the predictions has a dramatically increased probability of being correct.

What response should a scientist have if the prediction is *not* verified? When a prediction inherent in a theory is comprehended, it is not reasonable to demand that the predicted entities actually be found the next day. On the other hand, if legions of scientists spend their entire lives looking for these entities over a period of a millenium, and they still are not found, it would not be reasonable to continue to believe that the entities exist and will be found with just a little bit more experimentation.

A critical question of theory-evaluation is how to decide to abandon the search for predicted entities: there is no easy answer to the question. It is not an issue of elapsed time. If, for example, a single scientist spends one hour a year for twenty years looking for a predicted entity whose nature makes detection extremely difficult, one could rationally say that the theory is viable and that the entity may reasonably be expected to be discovered at some point, because no serious effort would have yet been put into searching for the entity. Nor is it just a matter of the number of scientists looking into a problem: a thousand scientists can search for ten minutes each and find nothing.

Consider an idealised example. Our task is to check for a 1 inch square flat piece of steel on a surface. (Why would we do this? Because some person purported that such an object exists somewhere, and has even offered a reward for finding it). Glancing at the surface is not a sufficiently thorough search to answer the question, though one might happen to see the object, thereby verifying its existence. Failure to see the object does not entail nonexistence. To prove that the ob-

not make the claim that there are no other things on the list, but failure to explicitly address the exhaustivity of the list is a scientific crime, whose penalty is sanctioning the worst possible interpretation of the theory consistent with what little is explicitly said by the theory. Scientists are usually human, and may need punishments for bad behavior.

ject does not exist, we must show that there is so little probability of existence that it would be unreasonable to maintain the hypothesis.

Assume first that the area to search is a square 3 feet on each side, the surface is flat and contains no holes or other objects on it, is painted white, is in a well-illuminated room, and the object is standard unpainted stainless steel. Under that condition, given normal human visual abilities, it would be possible to determine with satisfactory certainty that the object does not exist on the square by looking for a few seconds; or, by systematically running a hand over the square, one could make the same determination; or since the object is steel, using a magnetometer would reveal the presence of the object, and failure to produce the object has to be taken to mean that there is no such object. There are various conditions that could invalidate the search, for example if the light in the room were so bright that vision was impossible, visual evidence would be useless: in that case, the experimenter would find a way to adjust the lighting. The tactile evidence might be thwarted if the area was saturated in a local anesthetic which deadened the nerves; the magnetometer could be rendered useless by the presence of a strong electromagnetic field in the area. These are all possible interfering factors, but individually they are unlikely and the combination of all three is very improbable — moreover, the scientist can test for these factors, and make appropriate modifications to the testing procedure.

On the other hand, if the area to be searched is a square two miles on each side, the “just look” technique for searching would not yield a way to determine that a negative result is due to non-existence, rather than an inadequate search. If the object is made of glass, a magnetometer would be pointless. If the surface is multi-colored and the object is transparent, visual clues may be extremely unreliable. The essential point is this: if the entity exists, its nature, and the tools available for conducting the investigation (including the intellectual resources available for the task) largely determine the difficulty of the investigation. This is the standard for evaluating the probability of existence of an unobserved. As a greater portion of the effort required to find the object is invested in the search, the probability of having already found the entity increases, and if the entity has not been found, the probability that the entity does exist decreases.

The major factor with respect to how “hard” scientists have to search (relevant to the question of how soon we should conclude that a predicted entity doesn’t exist and the theory is wrong) is the degree of control which can be exerted in the search. An experimental discipline like physics has a very high degree of control over the setup, and given the rigorous nature of theoretical physics (i.e. minimal use of undefined concepts), the main struggle is getting the experimental setup in place. If a particle decay is predicted to take place at 10,000 TeV, the main stumbling block would be the lack of equipment to achieve such energies. However, if the experimental setup is not way beyond our technological grasp, the assumption is that soon after looking for a thing, we should find it. Some predictions may depend on events which are completely uncontrollable at present — e.g. the exact position of the Earth, Sun, Moon and some distant star; or, the relevant conditions may arise naturally only once every million years as planets line themselves up correctly with respect to some distant star. Under such conditions, not observing a predicted entity or event is of little consequence — that is, until the requisite experimental conditions occur.

Other sciences, such as linguistics or biology, are more limited to events which happen beyond the control of humans. A theory of bird migrations would depend nearly exclusively on observing the actual migrations of birds. The entire field of archaeology is limited to non-experimental observations of events which happened (although the methods of instrumentation used

by an archaeologist to learn facts about artifacts may be exactly the same as those used in physics and chemistry). Even in experimental areas of linguistics, one of the fundamental limitations on theory testing is that we are limited to working on actually existing languages: unlike the particle physicist or chemist, we cannot create artificial languages and implant the grammars of these languages in the brains of experimental subjects, to test the results.

6. How Can One Person Test Another Person's Theory?

A fundamental characteristic of science is that it is *objective*. The requirement of objectivity was first mooted as the principle of Well-Formedness of theories: “the statements making up theory must have meanings which are unambiguously knowable”. This is a consequence of the axiomatic principle that all aspects of science must be objective; theories must be stated objectively, experiments must be conducted so that their validity can be judged objectively, and theory testing must follow objective principles.

A value, state, or thing may be considered objectively or subjectively. To say that an experience or condition has an objective value means that it can be sensibly experienced by all observers,²³ and thus the object or condition has reality independent of the specific mind that experiences it. A subjective experience or condition is one which is particular to the individual. Suppose I am stabbed in the hand with a needle. A certain objectively stateable sequence of events takes place: the needle pierces my skin and comes close enough to a nerve, which causes an electro-chemical reaction to take place, and a signal is transmitted to the brain. In the brain, other electro-chemical events take place. With appropriate technology, these facts could be studied. In addition, a subjective event takes place, namely I experience pain. I can describe the pain, I can liken it to that which I believe you must have felt at some other moment (e.g. when you had a tooth pulled without anesthetic), I can attempt to replicate the sensation of pain within you by stabbing you in the hand, I can express an opinion about how horrid the pain was on a scale of one to ten. The causal events leading to the pain can be studied objectively, and the physiological and psychological consequences of those events can be studied objectively, but the actual sensation of pain will always be subjective.

To take another example, when one sees a red object, due to the nature of human anatomy and the laws of physics, the brain receives certain signals — sensations in the brain. The causal principles are objective, but the mental sensation is itself subjective, and for this reason it is pointless to ask whether the sensation of red that I experience is “the same” as the sensation of red which you feel looking at the same object. The objective causal principles behind reflection of light from red objects, the absorption of light by the rods and cones in the retina, and the transmission of the signal to the brain, are all facts that anyone can study. The electro-chemical events in my brain can be studied; my report of the sensation can be studied. You can compare these events and determine whether my brain reacts the same way at seeing a red object as does the brain of another person; or

²³ It may be necessary to further restrict this to observers in possession of the appropriate biological faculty for apprehending the object or condition. An apple is objectively seeable, but only by those who are able to see. A simple logical proposition such as “Either Tom is a human, or Tom is not a human” can in principle be objectively comprehended and evaluated by any human who has the ability to comprehend and reason: however, a person in a coma lacks such an ability, as does a newborn infant. The existence of exemplars of the species *homo sapiens* who happen to lack the ability to perceive that which is objectively there does not invalidate the objective reality of the entity or the concept of objectivity, in my personal opinion.

you can compare some psychological state such as my ability to determine whether two objects have the same or different colors. However, you cannot study whether my sensation of red is the same as your sensation of red.

Since the nature of the universe is independent of your experiencing it, and since the purpose of science is knowing reality, science must be objective. A personal and subjective view of the universe can only be an accurate characterisation for the experiencer, and cannot be known to anyone else. All knowledge is rooted in subjective experiences (such as seeing something), but we need to progress past the roots. The objective aspect of knowledge comes when the creator of the knowledge begins to abstract away from ineffable experiences, and expresses their knowledge in an objective fashion, focusing not on the ineffable sensations but seeking objective equivalents — “red like this book”, “long and thin like this pen, only longer”, “about as heavy as this rock”.

A further consequence of making science objective is that the same knowledge can be created, evaluated and held by more than one person. A scientist can study the universe for years, and arrive at a purely subjective, ineffable comprehension of the topic, but this knowledge will die with the scientist unless it can be made objective and communicated in some form. By a trade of knowledge between scientists, perhaps one conducting an experiment and the other replicating the results, both scientists may increase their individual knowledge more than either would have acting alone.²⁴ In order to express knowledge objectively, concepts with fixed and objectively knowable values are required. The most obvious examples of fixed and knowable concepts are numbers: anyone who knows the shapes of digits and grasps basic numeric concepts knows the value of “36”, “51” or “3,297” (still, it is important to specify the convention of communication, since “36” typically pronounced “three six” could be a hexadecimal number with the decimal value “54” pronounced “fifty four”). As you venture further from mathematical concepts, it becomes more difficult to express concepts objectively, because there is no exhaustive, transparent and objective metalanguage for describing entities and concepts. This motivates many scientists to reduce their theories, as much as possible, to mathematical equations, because of the objective nature of numbers. Whether the mathematisation of science always renders knowledge more objective is debatable, insofar as it is only the numeric part of the equation that is totally objective. The equation “ $\text{blort} = 2 \times \text{flad} + \text{bleep} \div 7$ ” is in no way made “objective” just because it employs numbers and mathematical operators. Mathematical models of reality crucially depend on non-numeric constructs, and if the non-numeric terms in an equation or theory cannot be known objectively, the remainder of the equation or theory is correspondingly uninterpretable.

The lack of a concise, precise, and objective calculus of description does not preclude meaningful scientific activity. Often knowledge is stated using words whose meanings are reasonably well known. Take the statement “alcohol is combustible”. It may eventually be realised that the meaning of some term is not known objectively, typically because the term covers two or more concepts, and it is not stated whether the theory holds only under one interpretation of the term. Conflicting statements about reality may arise when the entity being referred to by one individual is not the same entity as the one referred to by another individual: such a concept is peculiar to the individual holding the concept, hence is not objective. The term “alcohol” is a broad concept cov-

²⁴ Unless the scientists spend more time squabbling pointlessly over terms, techniques and the meaning of results than they would have spent each doing the entire work independently.

ering a number of more precise concepts,²⁵ such as ethanol, methanol, pentanol, isopropyl alcohol, butanol, and tyrosol. When the ambiguity becomes known (which is itself a type of knowledge, hence scientific progress), the scientific statement must be refined to resolve the conflict. As it turns out, not all alcohols are “combustible” — the term “combustible” is being used here informally to mean “it’ll catch fire if you stick a lit match into it”, knowing that eventually this concept must also be refined. Tyrosol is not combustible; at some point the statement is rephrased so as to make clear what alcohols are combustible (and, what exactly “combustible” means).

Definitions play a very important role in science, but ironically it is hard to give a satisfactory definition of “definition”. Simply saying that the “definition” of a word is “what the word means” does not help, because then we must ask what it means for a word to “mean” something, and we can get into an infinite regress of defining meaning in terms of “sense”, defined in terms of “semantic interpretations” and so on. We will therefore leave “definition” as an undefined term. However, unlike art or smut, we usually know it when we see it.

It is desirable separate the objective definition of a term, and the detailed knowledge one has of the referent. Take as a simple example the word “dog”. An ostensive definition of the term is:



but this leaves us unclear whether that means that



or



objectively constitute instances of the term “dog”.²⁶

²⁵ Technically, a hydrocarbon group and one or more -OH groups, so in principle the term refers to an unbounded class of molecules.

²⁶ The entity on the right is an instance of the concept “child”.

Dictionary definitions of the term are as follows. The Cambridge International dictionary is supremely uninformative: “A common four-legged animal, esp. kept by people as a pet or to hunt or guard things”. Webster’s dictionary says “Canid, *especially* : a highly variable domestic mammal (*Canis familiaris*) closely related to the common wolf (*Canis lupus*)”, and the American Heritage Dictionary says “A domesticated carnivorous mammal (*Canis familiaris*) related to the foxes and wolves and raised in a wide variety of breeds”. These latter two definitions beg the question, because we lack a definition of *Canis familiaris*, which turns out to be defined as “domestic dog”. An anatomical and biological characterization of “dog” could be essayed, stating properties of “dogs”. The question now becomes, what is the difference between a *definition* of the term “dog”, and the scientific knowledge of the characteristics of dogs? If new research reveals that the liver of dogs secretes a previously undetected enzyme, what changes is our knowledge about dogs, not the definition of the word “dog”.

Perhaps the problem lies in the fact that “dog” is such a self-evident term and is not a crucial concept in scientific theory-formation, so no definition is needed. Definitions serve scientific precision, not ordinary communication. Let us turn to a linguistic concept. It has been posited that there is an entity termed “syllable” in language: we need to know what this means. Webster’s Dictionary defines “syllable” as follows (note that this definition actually resolves in the negative the long-standing question of whether there are onsets, rhymes and coda):

A unit of spoken language that is next bigger than a speech sound and consists of one or more vowel sounds alone or of a syllabic consonant alone or of either with one or more consonant sounds preceding or following.

and “syllabic” as applied to consonants is defined as:

Constituting a syllable or the nucleus of a syllable: not accompanied in the same syllable by a vowel

Which is circular. The Cambridge International Dictionary defines “syllable” as:

A single unit of speech, either a whole word or one of the parts into which a word is separated when it is spoken or divided when it is printed. It usually contains a vowel.

According to the American Heritage Dictionary, a syllable is:

A unit of spoken language consisting of a single uninterrupted sound formed by a vowel, diphthong, or syllabic consonant alone, or by any of these sounds preceded, followed, or surrounded by one or more consonants.

As linguists recognise (and budding linguists will soon recognise), these are not useful definitions for “syllable”, being a hodgepodge of definition, diagnosis, exemplification and circularity. We begin with a stipulative definition of syllable, that is to say, this definition cannot be questioned. What can be questioned is whether this definition corresponds to a real entity. The term “syllable” is defined as

a specific constituent in prosodic representations, conventionally symbolised as σ , which immediately dominates skeletal positions, and which is immediately dominated by the node “foot”.

Elsewhere, the terms “foot” and “skeletal position” will be defined. What we learn from this definition is how it relates to these entities. Now consider a competing definition:

a specific constituent in prosodic representations, conventionally symbolised as σ , which immediately dominates the nodes “onset” and “rhyme”, and which is immediately dominated by the node “foot”.

Again, the definition of “onset” and “rhyme” must be defined elsewhere.

These statements are clearly different, but how do we characterise this difference? The statements differ in their claims regarding the properties of syllables, so under the first definition there is no constituent “onset” or “rhyme” intervening between the syllable node and the segments. Again, we have to ask, what is the difference between the definition of the term “syllable”, and our specific knowledge of the characteristics of the object referred to by the term. No scientific questions actually depend on the definition of the syllable: all issues surrounding the syllable pertain to its characteristics.

The point of this exercise in definitions is to make clear some of the issues involved in making knowledge objective. Theoretical statements should be made using terms with objectively known meanings, and using unambiguous arrangements of words.²⁷ While imprecision does not invalidate a statement, it decreases the degree of objectivity of the statement, and thus reduces the utility of the statement as a tool for gaining knowledge, because we cannot be certain which of two or more possibilities — those corresponding to the ambiguity — is reputed to be correct.

7. How Much Must We Do At Once?

We have spoken in terms of constructing a theory about the nature of the universe, but clearly no one theory comes anywhere near being an adequate statement about all aspects of the universe. Rather than solve all problems simultaneously, science divides the problem of the nature of reality into smaller problems and solves the smaller problems. Therein lies a major problem for theory evaluation. The universe does not come with naturally occurring domain delimitations. The conceptual division of the universe results from reasoning. There are an infinite number of imaginable ways to partition the universe. This has a major consequence for theory evaluation, that when theories describe different aspects of reality, the scientific success of the theories cannot be meaning-

²⁷ Consider the statement: “A syllable is not aligned on the left with a vowel”. Leaving aside what a syllable is, and the unexplicated notion “aligned”, this statement could be intended to mean “there exists a syllable which is not aligned on the left with a vowel”, or “no syllable can be aligned on the left with a vowel”. The plain English expression is ambiguous, and it is scientifically unacceptable to leave the interpretation of the statement to the reader. Nevertheless, this exact statement of a grammatical principle was proposed in an unpublished paper. The author intended the “no syllable can...” interpretation, a fact only guessable from knowing the author’s intent.

fully compared. Quantum mechanics is not ‘better’ than generative phonology because it enumerates more things: the domains of explanation are incommensurate.

The nature of humans is so complex that no single theory can hope to make sense of humans in general. A less ambitious goal would be to develop separate theories of human cognition, respiration, digestion and so on; but even developing a theory of cognition is an immense task. Therefore, scientists first attack more tractable questions, such as “what is the nature of vision?”, “what is the nature of audition?”, “what is the nature of tactile perception?”, “what is the nature of language?” and so on. But as we know, language is still a huge domain, and the most progress has been made by further subdivisions of reality, e.g. “synchronic phonology”, “diachronic syntax”, “socio-pragmatics”, “acoustic phonetics”.

One of the most prominent and unresolvable problems of theory evaluation arises from the fact that competing theories often do not describe the same aspects of reality. Let us take the case of phonology; we wish to articulate a “theory of phonology”. Therefore we must know what exactly “phonology” refers to. As discussed earlier in the course should clarify, it is insufficient to say that “phonology is about language sound”, because that characterisation fails to distinguish phonology from phonetics. One reasonable response could be to say “then maybe there should be no distinction: the two should be unified”. Let us assume, temporarily, that this is good advice, and see which rocks it leads us to. It is also well known that phonetics involves at least three sub-components: articulation, acoustics, and perception. A full understanding of speech articulation requires full understanding of articulatory anatomy, and there is no reason to think that the principles governing the musculature of the tongue are significantly different from those governing any human musculature. But at this point, there is no compelling and principled reason to distinguish the need to understand human musculature from the need to understand that of other mammals, and once we’ve folded mammalian musculature into the equation, we may as well look at reptilian musculature and amphibian synapses. Therefore there seems to be no major reason of principle to exclude the entirety of biology from the study of phonology, and once we’ve gone that far, we may as well just include the entire universe (especially since acoustics is part of phonetics, thus phonology by assumption, but acoustics is a subdomain of physics, which is ultimately the study of everything).

This returns us to the problem that developing a theory of everything is an unreasonable goal, and no progress could be made in increasing knowledge if we are prevented from breaking the problem into smaller pieces. Therefore the principle which justifies limiting the domain of phonology is that if it is *not* limited, the domain of study becomes intractably huge, and we cannot reach definitive answers to any questions. A unified theory of phonetics and phonology would be a conceptually better theory than separate theories of phonetics and phonology, but only if the unified theory is equal (or superior) to the separate theories with respect to accounting for reality. If the conceptual unification of two domains comes at the price of a decrease in knowledge (for example if an explanatory mechanism available under the two-theories approach is lost in the unified approach), then the unified theory has less scientific value than the two theories. Domain unification, i.e. theory simplification, is subordinate to the goal of maintaining or increasing knowledge of reality.

That said, the question arises as to exactly which aspects of the universe *are* most coherently explained by corraling them into the domain “phonology” as opposed to “phonetics” or “syntax” or “physics”. This question is very hard to answer, and unfortunately is very rarely addressed in a systematic fashion. To illustrate the problem, we consider a real domain-related question which

comes up in phonology. Are properties of underlying representations part of a phonological grammar (and thus are the responsibility of a theory of phonological grammars)? No words in English begin with the cluster /zv/ in underlying forms. If the theory of phonological grammar is not required to account for properties of underlying forms, then properties of underlying forms cannot be used to refute or support some specific theory of phonology; if such properties are in the domain of phonology, then these facts are of relevance.

One extreme position with respect to this question is to deny any scientific import at all to the claim. Even if you grant the descriptive statement that no words of English underlyingly begin with /zv/, this truth could just be a truth of no profundity. The observation could be true, if underlying forms are a selection of random combinations of sounds. There is no requirement that there actually be an underlying form /slæwp/, and the fact that there is no such word could be considered to be an accidental gap arising from the fact that only a subset of the combinatorically possible words of English actually are words. The probability of a complete lack of underlying initial /zv/ arising at random is extremely low if we assume no principles, whereas if we assume a theory which contains statements that predict such a gap, the probability of the gap is very high (perhaps, guaranteed). It would therefore be scientifically irresponsible to reject any attempt at explaining this gap, leaving it to random distribution, unless it can be shown that the gap does truly lack any explanatory basis.

All this rules out is the approach of deciding not to explain the fact at all, i.e. ignoring reality entirely. There are at least two theories which do explain the fact; the pattern may be due to the presence of a principle in the mental grammar of English which precludes such words, or, it may be due to other factors which are not part of the theory of grammar, but which are still part of the theory of the universe. There are actually quite a number of grammar-external reasons to expect English to have no such words. First, the properties of English, meaning Modern English, are largely determined by historical antecedent. A child learning a language does not invent the language at random, based on no experience, but rather constructs a grammar based on data available to the child, that is, how people around her speak. By this chain of data-custodianship, even though grammars are only imperfectly reproduced in the process of language acquisition, there is a great probability that properties of the language at one time will be true at later stages, even millenia later. The fact that Modern English has no words of the form /#zv.../ can be explained on the basis of the fact that there were no such words in Middle English with that structure (which can be explained by reference to their lack in Old English, itself a function of properties of Proto-Germanic and so on). Furthermore, no historical changes would have contributed such words from other sources, for example there was no apocope process which would have turned earlier *ziva* into *zva*.

Our goal is not to exhaustively explore the nature of evidence in phonology, or to resolve this specific question. Rather, it is to clarify that while systematic observations require explanation, they do not require a specific theory, such as the theory of formal phonological grammars, but rather simply need to be explained by *some* theory. If — and this is a very large if — *if* the “outside of grammar” theory of the lack of /zv/ is actually as plausible as the “part of grammar” account, you may not be able to rely on empirical considerations in deciding which is the better theory, in which case you must resort to aesthetic principles. In such a case, the “outside of grammar” theory has an aesthetic advantage, because it does not impute to the theory of grammar unnecessary and arbitrary principles which can be fully explained by mechanisms already available outside of the theory (if indeed those mechanism *are* effective at predicting the observation).

Just as scientists have little objective basis for a quantitative assessment of evidence and only use general expressions like “this stands a good chance of being right”, we also do not know how to mathematically calculate the goodness of fit between theory and observation with any reliability. While the factors outlined here do give a hint of why underlying /zv/ is not likely to exist, the probability is not null. We have not stated all of the factors which govern the history of the language (upon which we base our grammar-external explanation). Since we cannot compute the actual probability of a word of the form /#zv.../ arising through an explicable historical mechanism, we cannot actually see to what extent the lack of such words matches the grammar-external explanation. It is therefore very important to distinguish between apparent redundancy between domains with actual empirical consequences, versus gratuitous principle-repetition. It has been suggested that certain facts of phonology are in some sense an expression of Boyle’s Law: if that is really true, phonology should not contain such principles. On the other hand, phonology might contain principles that are evocative of but not the same as the physical principle Boyle’s Law. A grammatical treatment could predict “physically unnatural” grammars as well as precluding physically natural but grammatically impossible rules. Finding (or failing to find) the distinguishing cases will tell you the status of certain laws of phonology.

This motivates a “modular” approach to theories: greater progress may result in stating and testing theories if the theories are about relatively small conceptual domains, where there is less probability that an alternative theory can be constructed which redefines the domain. Rather than constructing and attempting to test a huge theory of synchronic phonological grammars, you would be more successful if one were to construct smaller theories, which form parts of a grand theory. Thus one could address questions about derivations: whether grammars are composed of partially ordered or strictly ordered sets of rules, whether ordering relations are reflexive (and under what conditions); one can separately address questions about representation: whether phonological representations are built on value-attribute pairs with two values (more? fewer?), whether there is a one-to-one mapping between segments and atoms, etc.

At the same time, the process of breaking down a larger problem into bits can be carried to an absurd extreme, so you could pursue a theory of rules of palatalization of velar fricatives preceded by high front vowels, claiming that eventually such a theory will be unified with the millions of other theories of palatalization that would have presumably been developed at the same time. In the second section, reference was made to the principle of Comprehensiveness, which requires a theory to be a complete account of reality. Comprehensiveness values a theory which provides a statement about all rules in phonology, and not just rules of a particular type such as “syllable rules”, “segmental rules”, “assimilation rules”. As revealed in the names, it is assumed that there is an underlying unity to the concept “rule”, and that while there may be some differences between the entities, there will be more similarities than differences. This would be expressed in a general theory of rules by stating a formal theory of rules, i.e. what elements combine to form rules, what are the metatheoretic rules of rule construction, what aspect of rule form is not fixed in advance of knowledge of the function of the rule (that is, what data it applies to). The similarities in different kinds of rules is thus expressed in a formal theory of rule. A single theory of phonological rules would therefore be preferable to a series of unrelated theories of rules for specific domains, and if the ideal of a unified theory of rules is unattainable, the retreat from that goal should be minimal.

8. Summary

The goal of science is gaining knowledge of reality. A theory is the conceptual tool for integrating that knowledge in the mind, so as to extend knowledge beyond actual observations or reports of observations. To be a useful epistemological tool, a theory must be precise: it must lead to propositions with knowable meanings and truth values. Secondly, to be held in the mind, its form must be compact.

Evaluation of a theory as a representation of reality cannot only use formal proof, since few propositions can be shown to be formally true, and the basic assumptions required to prove most scientific statements are not self-evident. Scientific argumentation proceeds on the basis of statements about reality whose truth value is external to formal logic, and is rooted in human sensory apparatus and the principle that the universe operates by rules, not at random. Because humans are not omniscient, knowledge is grounded in statements of probability that a given statement is true, which constitutes an evaluation of the quality of the evidence supporting a proposition.

In evaluating a theory as a synopsis of reality, highest philosophical priority is given to determining the burden of proof, and highest scientific priority is given to establishing that the predictions which are the burden of a theory are correct. Failure to demonstrate the existence of predicted entities suffices to refute a theory, with the provision that suitable testing of the theory has been conducted. What exactly constitutes a “suitable effort” to demonstrate predictions is determined by the nature of the entity — the extent to which it must be observed without human intervention, the efficacy of human efforts to bring about the conditions favoring presence of the entity, and the difficulty of distinguishing the sought entity from similar entities.