

## Underdetermination, With a Glancing Consideration of Occam's Razor

### Introduction:

Underdetermination is a worry both for philosophers of science, especially of the realist persuasion, and for scientists, especially those more foundationally thoughtful ones also of the realist bent. What is the worry? Crudely put: Could it be the case that science is wrongly crediting a highly empirically successful theory with being true or even approximately true, when in fact it is wildly false? Does the fact that the history of science is littered with the corpses of failed theories indicate that we are making progress towards truth or that our current favored theory is as likely to be wrong as those past, rejected ones? – Note that the emphasis is on THEORY. A theory is an artifact of language, and most (realist minded) folks would like to think a correct theory is one that correctly describes the world. But, nothing is so simple. The problems are subtle and ugly. – In this (poorly written) paper, I am taking special consideration of problems that would be pertinent not just to human beings, but also to robots and machines which we would like to design to do science on our behalf. Often, issues which we would like to dismiss for our human selves, cannot be so easily dismissed for a robot. – We start by thinking about the linguistic/logical basis of theories, language being only means by which science can express its opinions about the world.

### 1. Syntax vs. Semantics.

Let's start with a basic divide that is all-important to logic and philosophy. The language with which we use to construct theories has two fundamental aspects: syntax and semantics. Syntax is the pure *form* of the language, is a set of symbols and a set of rules for how to put those symbols together, *without regard to meaning*. A close approximation to syntax would be algebraic statements, such as:  $x * y = z$ . The variables  $x, y, z$  are themselves not supposed to have any particular meaning but are intended to be a place-holder for some particular meaning (a number) that you might plug-in later. We can transform that sentence to an infinite number of equivalent statements ( $y = z/x$ , etc.) all without ever needing to know the particular meaning of  $x, y, z$ . [If you need another analogy for syntax, you could think of playing cards. Each game defines different rules for how the cards are to be put together, how to transition from one step to the next, and what makes for the end of the game. No one ever asks: 'But what does the 3 of clubs *mean*?' ]– Pure syntax is more abstract than algebra; it is strictly a specification of meaningless symbols (for those recalling a logic class taken years ago: constants, predicates, functions, connectives), rules for concatenating those symbols, and rules for transforming those symbols into new strings of symbols. – The great power of syntax lies in its generality, in particular for proofs. If you've done your syntactic job correctly, then *ANY* assignment of meanings, no matter how far-fetched, will work just the same. (*Viz.* any assignment of meanings making the premiss-set true, will guarantee the truth of the conclusion)

Semantics is about the meanings that would be assigned to those otherwise meaningless symbols. In math, the meanings are usually numbers. In science, the meanings are (at least *prima facie*) referring to things (or purported things) in the world. That is, scientists aim for their theories to be about things in the world. Constants are supposed to name objects. Predicates are supposed to describe properties or relationships about/among objects. Functions allow a more general, if indirect, way to refer to objects. Connectives allow us to describe more complex circumstances. Semantics is just this assignment of meaning to otherwise empty items of notation. The most important semantic element is TRUTH.

By folks with lots of work to do, the relationship between syntax and semantics is taken for granted. But, for those sensitive to foundational issues, puzzles cannot be evaded. What exactly is *meaning*? What is *truth*? How are meanings *attached* to the notation? How do they *stick*? What is the criteria for correct sticking/attachment? Insofar as we want to describe the world, is there more than one (but

mutually incompatible) way to describe the same thing? If so, which way is to be preferred? Are there limitations to description/meaning? Do these limits bear on our capacity to do science?

## 2. A Few Naïve (and false) Assumptions

One naïve idea is that we use our *ideas* of the world and other mental content to serve as semantic content. The notion is that determining the meaning of a sentence like 'each planet moves in an ellipse with the sun in one focus,' is just to insert our pure mental images, associations, etc. about planets for the symbol string <planet> etc., for each term of the sentence. –But this naïve notion commits a mistake: it assumes thought and language are two different things. We first think the thought, then find some notation to express it. Wittgenstein mocked a French politician who once said: French is the superior language because: "The French language uses its words in that order in which we think them" – as if Chinese speakers first think their thoughts in French then must clumsily translate from French to their Chinese expressions. – The notion that thought is distinct from language mostly amounts to thought being a 'private language' (a kind of pure mentalist (or sensationalist/phenomenological language)) which we then convert to a public language in order to communicate with others, etc. Wittgenstein famously attacked this notion. Here's an example:

258. Let us imagine the following case. I want to keep a diary about the recurrence of a certain sensation. To this end I associate it with the sign "S" and write this sign in a calendar for every day on which I have the sensation.—I will remark first of all that a definition of the sign cannot be formulated.—But still I can give myself a kind of ostensive definition.—How? Can I point to the sensation? Not in the ordinary sense. But I speak, or write the sign down, and at the same time I concentrate my attention on the sensation—and so, as it were, point to it inwardly.—But what is this ceremony for? for that is all it seems to be! A definition surely serves to establish the meaning of a sign.—Well, that is done precisely by the concentrating of my attention; for in this way I impress on myself the connexion between the sign and the sensation.—But "I impress it on myself" can only mean: this process brings it about that I remember the connexion *right* in the future. But in the present case I have no criterion of correctness. One would like to say: whatever is going to seem right to me is right. And that only means that here we can't talk about 'right'.

One could suggest that we establish correctness via some kind of internal dictionary which we consult to translate from the private language to the public one.

265. Let us imagine a table (something like a dictionary) that exists only in our imagination. A dictionary can be used to justify the translation of a word X by a word Y. But are we also to call it a justification if such a table is to be looked up only in the imagination?—"Well, yes; then it is a subjective justification."—But justification consists in appealing to something independent.—"But surely I can appeal from one memory to another. For example, I don't know if I have remembered the time of departure of a train right and to check it I call to mind how a page of the time-table looked. Isn't it the same here?"—No; for this process has got to produce a memory which is actually *correct*. If the mental image of the time-table could not itself be *tested* for correctness, how could it confirm the correctness of the first memory? (As if someone were to buy several copies of the morning paper to assure himself that what it said was true.) Looking up a table in the imagination is no more looking up a table than the image of the result of an imagined experiment is the result of

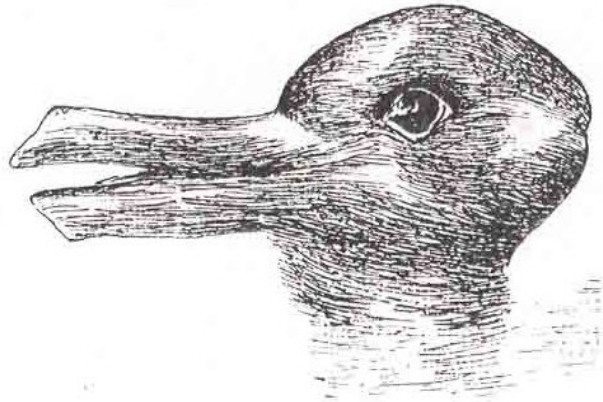
an experiment.

293.....Suppose everyone had a box with something in it: we call it a "beetle". No one can look into anyone else's box, and everyone says he knows what a beetle is only by looking at *his* beetle.—Here it would be quite possible for everyone to have something different in his box. One might even imagine such a thing constantly changing.—But suppose the word "beetle" had a use in these people's language?—If so it would not be used as the name of a thing. The thing in the box has no place in the language-game at all; not even as a *something*: for the box might even be empty.—No, one can 'divide through' by the thing in the box; it cancels out, whatever it is. That is to say: if we construe the grammar of the expression of sensation on the model of 'object and designation' the object drops out of consideration as irrelevant.

So, the naïve idea that we simply fill-in the empty notation (syntax) with thought/sensations/mentalist content (semantics) proves elusive. There's just no way to get that ethereal mentalist stuff to 'stick.' Yet, if this is not how it is done, then how *is* meaning attached to the syntax? If the language we use to compose theories about the world requires mental content of some sort, then we see the threat of a kind of regress: Rather than items of the world serving as semantic content, we instead substitute our idea of those worldly objects, because we had some confidence about them. Yet, on closer inspection, we see they are not so secure. So, if we nevertheless keep insisting on mentalist content, then we should need some, yet still more intimately possessed mentalist content to substitute for the less intimately possessed mentalist content. And so on, *ad infinitum*. (Wittgenstein instead argued that meaning is generated from the way we *use* language, as a community and individually, but I find this answer, to this problem, unsatisfactory, especially when we think about building a robot (how can the robot be brought to the point where it is a facile language user, which doesn't *already* involve deep knowledge about the world?))

This may all sound arcane, for *we* certainly know the meanings of the words we use, even if it proves difficult to say just how this meaning process works. Anti-realists will agree that we know the meanings of our words, too. The trouble comes in saying what those meanings are *about*, actually. The anti-realist wouldn't even mind that mentalist content (as we described) should serve as semantic content, because for him, it's not important that language really be about the world (as the realist conceives it). He will say that our sentences that seem to be about quarks, neutrons, black holes, etc., are in fact not about anything in the world which might bear those names, but are instead about something else, all of which boils down to a pragmatic exercise to help us get things done to our satisfaction. There may or may not be something in the world which corresponds to our language-use, but that (according to the anti-realist) is irrelevant to how language works and what is the purpose of language.

Another, related, naïve (and false) idea is that we can unproblematically and directly read-off from our perceptual 'data' the world as it actually is. There is a rock on the table; we observe this fact and can then report: "I observe the rock on the table."



Consider a last example from Wittgenstein:

This image produces one and the same optical sensation, yet what is it you are looking at? A duck or a rabbit? One not familiar with rabbits would only ever see a duck. – It would seem that, to see anything here at all, we must *first* have the concept of duck and/or rabbit firmly in place. – This should raise the worry: given one and the same set of sensations, might one person experience the world in one way, but a different person, with a different set of concepts, experience something very different? As one philosopher put it: How do we know our concepts are 'cutting the world at its joints' or not? This is a much deeper worry than the idea that we see green where a color-blind person would see gray. – Scientists have theories and then run experiments, etc., and they regard the empirical outcomes of these experiments as evidence for/against a theory. It is assumed here that what the scientist sees can only be seen in one way, can be seen directly and almost without dispute. But, if it is, in fact, wrong to assume this, then it seriously undermines the role that evidence is supposed to play with respect to a theory. One particular fear: if the theory is providing the very conceptual framework through which we experience the 'evidence,' then the 'evidence' is no longer confirming, since it is itself determined by and dependent on the very theory it is supposed to be validating.

One might try to be more careful about framing the role of perception and suggest: First, we have the perception; then we interpret that perception *as* a duck or *as* a rabbit. This introduces an old idea which has also been thoroughly vanquished: sense data. This thesis goes back to 'modern' philosophy and stemmed from the Newtonian inspired corpuscular theory of perception: light (in the form of tiny corpuscles) enters our eyes and stimulates our nerves, etc. It, as it were, illuminates a screen which we view from 'the inside' (our mind's eye view). This idea was proposed to deal with the problem of differentiating perceptual qualities (eg. shape, quantity, motion, etc.(Locke calls 'primary qualities')) that are supposed to belong to the thing being perceived vs. those qualities (eg. color, taste, odors, etc. (called 'secondary qualities')) which are really due to the nature of the 'screen' itself that is presenting the perception. Thus, our unified perception of ice cream is a jumble mix of qualities, some (shape, temperature, chemical composition, etc.) belonging to the object in our hands and some belonging in fact to us (taste, sensation of cold, odor, etc.). –The sense-data account attempted to argue: We are given a set of basic sensations, and from these, we interpret what we experience, ultimately deciding which qualities belong to the thing perceived and which not.

20<sup>th</sup> century philosophy ultimately defeated this and similar forms of the sense-data account. I won't review the history or the arguments, but just give a general flavor of the objections. First, the corpuscularian account of perception is simply false; it is not what we know of our actual perceptual experience. When we look at the duck-rabbit picture, we just *\*see\** a duck or just *\*see\** a rabbit. We

don't have the experience of studying over something primary (bits of sense data, like pieces of a puzzle), applying some interpretive procedure (putting the puzzle-pieces together into a coherent picture), and inferring a result. We do sometimes have the experience of looking at, eg., a blurry photo and trying to decipher what's in it. But, this is not what goes on in ordinary perception or when shifting back and forth between seeing a duck and a rabbit in the duck-rabbit picture. Moreover, we should ask the deeper question (whose answer I won't pursue here): *how is it possible to see an object according to an interpretation?* -Another important problem is this: There's nothing we can identify (of our supposed sense-data, or our more phenomenological or 'pure' perceptual experience) which isn't also, at times, an explicandum (thing-to-be-explained). That is, the role of sense-data was supposed to be to serve as an explicans (thing-that-explains) for something else, like the duck-rabbit picture. Eg. (the perception of) a scar may sometimes be an explanation for something else (why the clockmaker was forced to retire) yet at other times be the thing to be explained (as in pictures of scars in the classroom of a medical school). Hence, it becomes difficult to locate anything like sense-data that is immediately knowable, unquestionable and foundational for all our empirical knowledge. – There are further points to make, but suffice to say that the sense-data thesis is RIP dead. Consequently (and for other reasons unrelated to sense-data), the old worries which sense-data was introduced to settle, now re-emerge with new, more terrifying talons.

So, if our 'experiences' are direct, in the sense that we simply have them whole and uncut, but if what we are 'experiencing' may not be the world as it truly is, but is merely the product of our conceptual apparatus together with something else (nervous system output) beyond our ken, then how are our 'experiences' related to the world? – For the issue of underdetermination, you may begin to appreciate how items like 'evidence,' 'experience,' 'sensory systems,' 'perception,' etc. are not only *not* going to help you out of the hole, but may be deepening and worsening that hole. – As you can imagine, if you are building a robot, you can't simply pop TV cameras on its head, and it will then know what it's looking at. One would presume, you would have to tell the robot: 'when the camera has output x, y, z, then you are seeing a *rabbit*.' But, if you are doing this, then you could equally well tell the robot *anything* at all! What's the point of having eyes if you then have to be told what it is you are looking at in the first place?

### 3. Observation & Conceptual Schemes

So, to this point, we still don't have a satisfactory answer to the question of how semantics fleshes out syntax in a way that warrants our theories being about the world. Though, assuming it gets there somehow and affixes in some way, we see a new problem with the relationship between the world and our ideas about the world. The preceding indicated (a) our talk of the world may not be *about* the world at all, in fact, (b) the way we experience the world seems to require our first having a settled theory about it, and (c) what we seem to perceive of a thing includes aspects it doesn't in fact have and/or excludes things it may in fact possess. --The naïve notion of theory-world relation is that we have a theory in one place, well removed from the world of which it is about, and a world 'out there' (the target of the theory), whilst we make theory-neutral observations (seeing the world directly as it really is) about the fit between the theory and the world, in order to test the veracity of that theory.

But, this picture breaks down when we consider our actual epistemic position with respect to the world. It's not like we have some privileged god's-eye view (aka 'Archimedean viewpoint') from which to peer out onto the world, seeing it as it truly is, collecting 'facts' and putting them in our baskets for later theory-appraisal. We are stuck inside our heads, so to speak, and our view of the world is very 'theory-infected'. Imagine some aborigine people, never having seen a game of tennis, being brought to a match. They would have the identical visual experience we would, but they would certainly not see the same thing, would not know how to pull-together/cohere what would be, for them, near

unintelligibility. We would see tennis facts; they would not. --Seeing the same  $x$  involves sharing theories/knowledge of  $x$ . Coming to know the world requires primarily a search for intelligibility (a mode of conceptual organization), and only secondarily a search for objects/facts (which require a conceptual scheme to know them *as* objects/facts). – Consider the observation of a footprint in the sand; just calling it a 'footprint' implies volumes. Quoting Norwood Hanson: 'As Wisdom would say: every perception involves an aetiology and a prognosis.' Seeing that something is a footprint brings to the observation a whole array of theoretically determined possibilities, without which we couldn't see it as a footprint in the first place.

As noted earlier, scientific knowledge must be formed of language. Language has a versatility that pictures or audio recordings lack (i.e. pictures lack audible aspects, but a sentence can express a synthesis of aspects). Truth/falsity can only apply to assertions in language, and not to pictures or models. And, somehow, our theories and knowledge provide a conceptual mold into which our perceptions coagulate, enable disparate perceptions to cohere around a single axis, enable us to expect 'if  $x$  were done,  $y$  would follow.' Again from Hanson: No amount of shuffling pictures of antelopes will yield: 'antelopes are ungulates...' –In other words: no amount of pictures or experiences of particular things or bits of the world (or even summaries of these) could ever reach beyond themselves, ever result in an explanation/theory that ties these particulars together. Eg., merely collecting a large number of instances of 'sunlight reflecting off beveled mirrors produces spectra' will fail to add up to an explanation for this phenomenon. The reason for this phenomenon is not given by summarizing the instances and reporting: 'beveled mirrors do this.' The theory of light refraction cannot be milked out of observations of instances of beveled mirrors producing spectra. –Science is a way of thinking about the world, of forming conceptions that are bigger than and not included in the particulars about which science aims to explain. The paradigm observer is not one who reports what normal observers see, but who sees in familiar objects what no one has seen before.

Hanson asks us to consider (a) what is an observation of a fact? and (b) what is an in-principle inexpressible fact? A picture of the sun at dawn is not a picture of a fact. A heliocentrist and a geocentrist would see different 'facts' in the same picture (recall the duck-rabbit case above). – What is an in-principle inexpressible fact? By this I don't mean 'very complicated' or 'not yet ascertainable' – but a fact that, under any condition, would be inexpressible. Hanson asks us to consider how a different way of expressing things would produce a shift in the way we conceive of them: We (in English) use the adjectival idiom to say things like 'the sun is yellow,' 'the grass is green,' 'sugar is sweet,' 'bears are furry,' and this way of speaking conveys that properties like yellowness, greenness and sweetness inhere (passively) in the sun, grass, sugar. But, Russians will say 'the sun yellows,' 'the grass greens,' 'sugar sweetens' conveying by verbal idiom the idea that properties are more like actions by the objects. Saying 'the grass greens' implies that it radiates greenness, like x-rays fluorescence.

Try 'the sun rounds,' 'St.John's hall rectangulates,' 'sugar cubes.' Activity is suggested here. Would one who saw the round sun see the sun rounding? The college hall *is* rectangular [says the person who insists on the adjectival idiom as the correct one]. Would this fact be apprehended by a man for whom the hall recangulates – holding itself in a rectangular form against gravity, wind, cold and damp? (Hanson, 34)

These examples indicate how even a slight shift in language produces a substantial shift in concept and, so, a shift in the observation. But, it also indicates: if a distinction cannot be made in language, it cannot be made conceptually nor observationally.

Hanson offers us the example of Galileo's struggle to formulate the concept of velocity. Galileo's initial

(and erroneous) principle: the velocity of a falling body increases in proportion to the distance it has fallen from its starting point. – Galileo sought not a merely a descriptive/predictive formula. He sought more: an *explanation* of these data; they must be intelligently systematized; to reason back to a more fundamental principle from which the 'accidents' of time, etc will follow. – Though, Galileo sought not the *cause* of acceleration, that was Descartes' program.

Galileo initial principle, however, was in error: the principle he adopted (the velocity of a falling body increases in proportion to the distance it has fallen from its starting point) could never lead to a law of falling bodies as he formulated it. It leads to an entirely different law expressible only as an exponential function, a formulation which he could never have managed via the mathematics at his disposal at the time. The correct statement of the principle, as he eventually and with enormous difficulty, having to overcome the inherent limitations of the conceptual system within which he was working, arrived at: velocities of a free-falling body are proportional to *times*, not distances. As Duhem later showed: to transform *distance fallen is proportional to the square of the time* into the law *velocity is proportional to the time* requires the idea of instantaneous velocity, a concept expressible only in the notation of the fluxion or derivative. To detect the non-equivalence of points of time with points of space would have required the concepts of integral calculus, utterly unavailable to Galileo. – Given the alternatives between velocity being proportional to time vs. space, Galileo *et al* chose space, since it coincided naturally with geometry, the only formal system available to scientists of the era. In geometry, time has little significance or prominence; a time coordinate would have been as meaningless as a 'fragrance' or 'beauty' coordinate. Also, compounding the problem, differences in times were difficult to detect and measure. At distances less than 50', differences in elapsed times were not as likely to capture attention as would have differences in distances. To the early Galileo, focusing on time would have seemed a needless complication, especially when the alternative parameter of distance was so much more obviously compatible, though ultimately unworkable. Galileo eventually broke through these conceptual limits, but not until much later, 1604.

Ironically, indirectly, the focus on distance (space) drove Galileo to cleave from an earlier conceptual trap, Aristotelian impetus theory which is time-dependent (successive actions of impetus occur in time). Geometrizing motion allowed Galileo to ignore impetus (seen as the cause of increased acceleration). Impetus theorists sought to explain why motion continued, but for Galileo the point of this explanation became pointless; motion (under the geometric conceptual apparatus) became an unanalyzable brute fact. He thus chose spatial orientation vs. the causal/time-dependent one in adopting the geometric formulation of the problem. The early move took Galileo out of one trap, but then placed him into another.

What drove Galileo to geometry in the first place? He first tried to build physics within the old Aristotelian system, built on the idea of impetus, but this project failed. Then, he tentatively substituted for internal motive power the notion of repeated external attractions or shocks, producing each new effect. This shift began his march out of the wilderness of contradiction besetting his predecessors; they sought a constant cause (impetus, gravity, etc.) to produce a variable effect (velocities of falling body). By allowing increase in acceleration when a body was under the action of a constant cause, impetus theorists admitted creation *ex nihilo* [constant cause should yield constant effect], and even in a consistent theory this could not be represented geometrically. – Under the newer conceptual system, Galileo could treat velocity as a basic, defining property of moving bodies; it no longer required being sustained by something more fundamental. Galileo's geometric proof has the sum of instantaneous velocities acquired at each point of space fallen, which can be plotted on triangles as a linear function. But, velocity is also the sum of instantaneous velocities acquired at each moment, yet this cannot be plotted geometrically. Triangular representation only allows uniform increase in relation to time.

Observation in terms of what he *saw* (as opposed to what he endured) reinforced the distance interpretation. For Galileo, velocity was observed in the fall of an object from point A to point B; when these are fitted together as legs of triangle, there was no 'logical space' for a time parameter in that geometric conceptualization.

It took many years before Galileo was able to claw his way out from the confines of a geometric conceptual apparatus and MacGyver it to suit the correct principle. Seeing the hole punctured in the limitations of that old system, Newton set about fashioning a new one, introducing the Calculus. – But this now solicits the question: If Galileo's moves from the Aristotelian conceptual apparatus to the geometric conceptual apparatus and from that to the beginning of a new conceptual apparatus, perfected by Newton, were so extremely difficult and required such historic genius, what does that illustrate about the conceptual systems within which we are doomed to operate? If each system so greatly limits thinking, so tightly controls the ways of looking at the world and dictates such a narrow bandwidth of possibilities for formulating scientific problems and solutions, what does that say about the feasibility of doing science? What if the limits of our conceptual system are not recognized? What if some other conceptual apparatus had arisen *instead* of the one we currently have? The question poised by Hansen was about whether there are in-principle inexpressible facts. Clearly, whatever facts are beyond the reach of our conceptual system to express are in-principle inexpressible facts, and having seen an example of such limitations, we can appreciate all those facts which (we'd like to think) are available to us, but vastly beyond the reach of even a genius such as Galileo. It leads us to wonder: What if the conceptual system within which we are now laboring under is similarly limited, and what if our ability to detect such a limitation is beyond the capacity of our greatest genius? Who can say whether *any* conceptual system humans could concoct is equal to the task of science? – These are not idle questions. We can inspect the historical example of Galileo and all those great scientists who achieved similar triumphs and can know the epistemic situation is a genuine one, yet we cannot say that our position is any better.

#### 4. Underdetermination

Underdetermination is short for: Underdetermination of the theory by the evidence. I believe what I have discussed so far illustrates the many layers of problems that precede and lead up to proper underdetermination worries. In other words, the situation is worse than a straightforward presentation of underdetermination would lead one to believe.

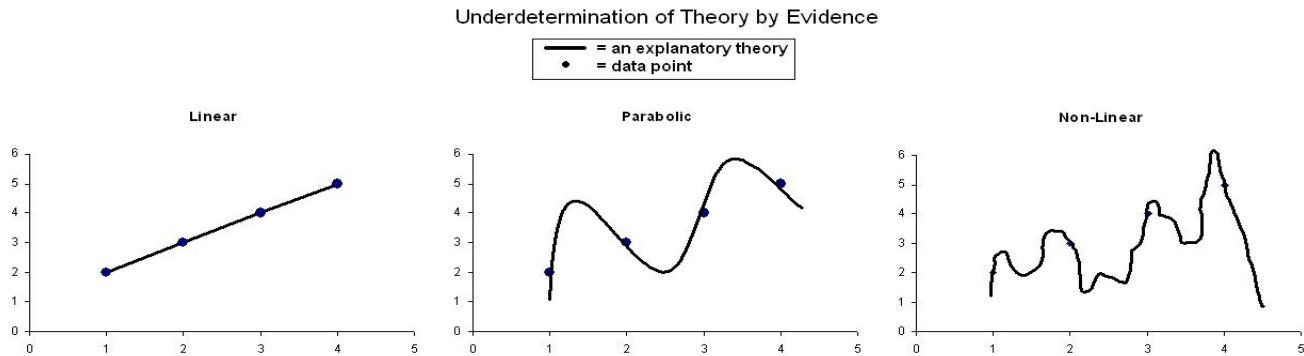
Underdetermination proper is the concern that exactly the same body of evidence will equally support a given theory T and some other theory T' which is incompatible with T. In fact, the history of science shows how very typical this situation is. Failed theories are the norm, not the exception. Successor theories replace the old ones. All the evidence that was accumulated to support Aristotle's physics equally supports all the successor theories as well. Who can say where or whether this process of theory succession will end?

One related argument to underdetermination, called pessimistic meta-induction, asks us to consider the history of science, not as exhibiting increasingly successful theories, but as exhibiting a high likelihood of any given theory to fail and be replaced by a new theory. If every scientific theory to date has failed and been replaced, then the odds are that the current theory will fail and be replaced as well. Thomas Kuhn, a historian of science, took this and the conceptual scheme arguments above and wove together a dismal picture of science as not progressing in any direction, especially not towards truth, just passing through regular cycles of theory-revolution-new-theory. He regarded the conceptual scheme of each new theory as being so comprehensive that even where two theories have the same terms ('gravity,' 'space,' etc.) they are incommensurate, and no rational discourse can take place comparing the two

theories, choosing one over the other. All that can happen is that a theory may be displayed and advocates win over 'converts' via non-rational appeals.

A more recent philosopher, Larry Laudan, has written some powerful papers/books covering the history of science and studying the interesting relationships among theories. In particular, he gives the lie to the idea that successor theories retain important elements of predecessor theories. The naïve notion that our current theories are carrying forward the true bits of predecessor theories, and that these true bits are what account for the empirical success of now-failed theories, is shown false.

The simple idea of underdetermination may be illustrated by this philosophy 101 example:



Each line describes a different theory, and only a moment's reflection will suffice to conclude that an infinite number of different theories would equally fit these same data points, or any collection of data points.

The more robust proofs of underdetermination require us to think of logically possible worlds. Imagine a book containing an infinite number of all possible statements. Now, imagine assigning truth-values to all these different statements. Eg. “The US civil war was won by the Union” could be assigned TRUE, while “Bill Clinton was president of the US” could be assigned FALSE, etc. for each and every possible statement. This book with its assignment of truth-values now describes one possible world. We can generate a second book with a different assignment of truth-values, and so on. For an infinite number of statements, the entire set of combinatoric assignments of truth-values produces an infinite number of described worlds. Of course, only one of these books would correctly describe the actual world. Unfortunately, we don't know which one of these books is that correct book. Thus, we could think of all of these books as being potential theories of the actual world. Waiving the issues raised above, about the problem of evidence and observation, etc., even if we could unproblematically check that observable parts of the actual world fail to correspond to any of these books (thus eliminating many), we cannot narrow the collection to that one, true book/theory.

Sometimes, it's hard to get our heads around the SIZE of this collection of books/possible-worlds/theories. Even limiting these books in different ways (by complexity, by causal consistency, etc.) it still does not indicate anything less than an infinite remainder. Importantly, if we limit the collection of theories by agreement with all the evidence that was or ever will be collected by us, it is not obvious that the collection has been limited to less than an infinite number. In other words: there are an infinite number of equally empirically successful theories (i.e. theories that fit the evidence) which are incompatible with each other (i.e. they can't both be true or even both be approximately true) and even which are equally complex (i.e. describe the world to the same degree of 'simplicity'). How to tell which theory is the correct one? The underdetermination argument will answer: WE CANNOT

TELL!

Now, I give an example of a different approach to underdetermination that feeds off these ideas, but also introduces the probability of our finding a true or approximately true theory from among that infinite group.

Let us assume that appropriate notions of truth and of approximate truth of theories have been defined, and that true and approximately true theories exist (otherwise the whole discussion about scientific realism would not make sense).

Let  $D_1$  be a finite set of data.

Let  $T_1$  be the set of theories such that  $T_1 := \{T, T \text{ is relevant for and consistent with } D_1\}$ ;

we assume that  $T_1 \neq \emptyset$ .

Let us now introduce a partition of  $T_1$  into two subsets: those theories which are true or approximately true, and those which are radically false, by which I mean that they are not even approximately true. Thus

$T_{1AT} := \{T \in T_1, T \text{ is true or approximately true}\}$

$T_{1RF} := \{T \in T_1, T \text{ is radically false}\}$

with  $T_1 = T_{1AT} \cup T_{1RF}$

In order to precisely articulate the different magnitudes of the theory sets in question, I need the mathematical concept of a measure on a space of theories. A measure is a generalization of the familiar concept of volume which is defined for the three dimensional Euclidean space. In other words, a measure states how big a subset of a space is, for more general spaces than just three-dimensional Euclidean space. By means of a measure on the theory space  $T_1$ , we can express the idea about the differing relative size of the theory sets  $T_{1AT}$  and  $T_{1RF}$ . According to our supposition, the measure  $\mu$  of  $T_{1RF}$ , i.e.  $\mu(T_{1RF})$ , will be much larger than  $\mu(T_{1AT})$ . So underdetermination in this form tells us that

$\mu(T_{1AT}) \ll \mu(T_{1RF})$ .

**Argument 1** (Transient underdetermination)

$T_1 = T_{1AT} \cup T_{1RF}$  and  $T_{1AT} \cap T_{1RF} = \emptyset$

$\mu(T_{1AT}) \ll \mu(T_{1RF})$ .

Therefore for any  $T \in T_1$ , it is very probable that  $T \in T_{1RF}$ .

Realists have, for some time, been fond of countering or diminishing underdetermination worries with appeal to the 'miracle argument' originally coined by Putnam. In general, the argument looks to the enormous success of our current, mature theories and states that realism is the only view which does not make a miracle out of the success of our theories. In other words, it would be a cosmic coincidence if our theories were so unbelievably successful but yet radically wrong. One special case of this is the 'novel-use' sorts of theoretic predictive success. An example: Fresnel's false wave-thru-ether theory of light was disputed by Poisson (corpuscular theory proponent), who by intended reductio calculated it would predict a bright spot in the shadow of a revolving disc; in fact, experiment showed prediction correct (the phenomenon thus sardonically named the 'Poisson bright spot')! – Of course, while this is a good example of novel-use or 'fruitfulness' of a theory to predict a hitherto unknown phenomenon, it still does not vanquish underdetermination, which has the numbers on its side (in the sheer mountain of possible worlds). Also, you will note Fresnel's ether theory of light is a false theory, though realist offer a different interpretation of history which gives wiggle-room to this particular problem.

Here is a formalization of the Miracle Argument:

**Argument 2** (Miracle Argument)

$T_1 = T_{1AT} \cup T_{1RF}$  and  $T_{1AT} \cap T_{1RF} = \emptyset$

$\exists T^* \in T_1$  such that  $T^*$  makes the novel prediction  $N$

For any  $T \in T_{1RF}$ , it is very improbable (or even impossible) to make prediction  $N$ .

Therefore, it is very probable (or even certain) that  $T^* \in T_{1AT}$ . i.e.  $\mu(T_{1AT}) \gg \mu(T_{1RF})$ .

The underdetermination theory, however, can easily deal with this version of the Miracle Argument:

Let us now investigate the potential effects that transient underdetermination has in a situation where a theory  $T^*$  that has been adapted to the data set  $D_1$  is capable of predicting use-novel data  $N$ . After the data  $N$  have been produced, scientists try to invent theories that are adapted to the new data set  $D_2 := D_1 \cup N$ .

Let  $D_2 = D_1 \cup N$  is a finite set of data.

Let  $T_2$  be the set of theories such that  $T_2 := \{T, T \text{ is relevant for and consistent with } D_2\}$ .

Again, we introduce a partition of  $T_2$  into two subsets: those theories that are true or approximately true, and those that are radically false. Thus

$T_{2AT} := \{T \in T_2, T \text{ is true or approximately true}\}$

$T_{2RF} := \{T \in T_2, T \text{ is radically false}\}$

with  $T_2 = T_{2AT} \cup T_{2RF}$ .

Remember, again, that the radically false theories contained in  $T_{2RF}$  are also relevant for and consistent with the data  $D_2$ . At this point, we can bring in transient underdetermination again. As in the case of the sets  $T_{1AT}$  and  $T_{1RF}$ , transient underdetermination tells us about the relative sizes of the sets  $T_{2AT}$  and  $T_{2RF}$ . It tells us that

$$\mu(T_{1AT}) \ll \mu(T_{1RF}).$$

Thus, most of the theories that manage to be relevant for and consistent with the old data  $D_1$  and the novel data  $N$  are *not* even approximately true.

Thus, regardless of how well our theories can predict, regardless of how well our theories perform, regardless what technologies and fruits they bear, our best scientific theories are far more likely radically false than even approximately true.

## 5. A Quick Point Regarding Occam's Razor.

Well, I ran out of time to write much about Occam's Razor. The idea of Occam's Razor is that, given the choice between two competing and otherwise empirically equivalent theories, we ought to choose the simpler one, typically understood as meaning: the one that makes less assumptions. From the discussion above, you can clearly see that it will not make the problem of underdetermination go away. But, even in the absence of that particular worry, philosophers have still struggled in futility to see in Occam's Razor anything more than a pragmatic or tactical virtue. In short, there's no compelling reason to believe that the world is more simple than complex. Why should the world care to bend to the dictates of a principle of theory evaluation? Isn't it in fact question-begging to assume so? (The simpler theory is the more likely true one. Why? Because the world is simple, as the theory describes.) – But, the generation of an infinite number of equally complex/simple, but incompatible, theories is not just an abstract idea. Concrete examples can be given: Consider:

Let us call Newton's theory (mechanics and gravitation)  $TN$ , and  $TN(v)$  the theory  $TN$  plus the postulate that the center of gravity of the solar system has constant absolute velocity  $v$ . By

Newton's own account, he claims empirical adequacy for  $TN(0)$  [zero velocity]; and also that, if  $TN(0)$  is empirically adequate, then so are all the theories  $TN(v)$ . –We see that all the theories  $TN(v)$  are empirically equivalent exactly if *all the motions in a model of  $TN(v)$  are isomorphic to motions in a model of  $TN(v + w)$* , for all constant velocities  $v$  and  $w$ . (van Fraassen)

Every piece of evidence we could ever collect, past, present, and future would be equally compatible with any of these Newtonian theories, yet each one describes a different universe, while being all exactly the same degree of complexity in their respective descriptions. It misses the point of the example to say that Newtonian theory was invalidated and replaced. The point is that complexity does not settle which one is correct or incorrect.

#### 6. Final Comment:

Reading this (poorly written) little paper, you might conclude that Alex is an anti-realist and/or anti-science skeptic. But, this is not the case. I am a realist, and I am of the opinion that science is indeed able to move towards truth, despite the serious challenge of underdetermination. However, as a student of philosophy, I know the only way to make my case is first to articulate the very strongest opposition position and then to see whether I can fashion an argument that is still viable in even this worst case scenario. We don't achieve anything by ignoring or wishing-away the opposing arguments.