Truth in Science

by

Tabish Ismail

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Master of Cognitive Science

in

Cognitive Science

Carleton University
Ottawa, Ontario

Your file Votre référence
ISBN: 978-0-494-94332-8

Our file Notre référence
ISBN: 978-0-494-94332-8

Abstract

It is argued that the best theory of the evidential data at hand is "true". In particular, I argue that human cognition is genetically endowed with a science-forming faculty (Chomsky, 1980), that the concept of truth is built into this faculty, and finally that the notion of "truth" in science is equated with "best theory of the data". In making the argument, it is shown that the standard theory of truth (Tarski, 1933; 1944) and the coherence theory of truth (Davidson, 1983) do not adequately account for how humans actually use the concept of scientific truth in cognition, while the Inference to Best Theory theory of truth does. Specifically, scientific propositions once thought to be true, but now considered false, can nevertheless be thought to have been true. Having embedded truth in human cognition, we attempt to derive these considerations from rational principles of communication laid out in Grice (1975).

Acknowledgements

This thesis consists of research spanning about four years of dedication and commitment regarding truth, science, and the world. During this time, I have gained innumerable amounts of helpful and insightful feedback from friends, colleagues, and other professionals without which this thesis would not have taken the shape it has. I thank everyone who has inspired and encouraged me on this project, in particular: Dr. Ana Arregui, Dr. Eros Corazza, Dr. Raj Singh, and Dr. Robert West. I have greatly benefitted from conversations with: Dr. Lev Blumenfeld, Dr. Andrew Carnie, Dr. Noam Chomsky, Dr. Patrick Davidson, Dr. Deepthi Kamawar, Dr. Claude Lamontagne, Dr. Ian Mackay, Dr. Eric Mathieu, Dr. Mark MacLeod, Dr. David Matheson, Dr. Jeff Mielke, Dr. John Perry, Dr. Laura Sabourin, Dr. Andres Salanova, Dr. Paul Thagard, Dr. Eta Schneiderman, Dr. Daniel Siddiqi, Dr. Ida Toivonen, Dr. Mario Zadrag, and Dr. Tania Zamuner. I personally thank and am indebted to Raj Singh, my thesis supervisor, for everything he has taught me, for guiding me in the proper direction, and for being there in times of angst and confusion.

I thank all members of the Logic, Language, and Information Lab for keeping me on my toes. I thank all my friends for keeping me off my toes. And I thank my family for giving me toes. The world is my biggest inspiration: I thank you, World, for giving me something to talk about.

Table of Contents

Abstract ii

Acknowledgements ii

Table of Contents iv

0 Preface 1

1 Chapter: Introduction 2

2 Chapter: Science 4

2.1 What it is and What it Does 4

2.2 Univeral Grammar and the Science-Forming Faculty 7

3 Chapter: Truth in Science 11

3.1 Scientific Propositions 11

3.2 Puzzle of Truth and Scientific Propositions 12

4 Chapter: Characterizing Truth in Science 14

4.1 Prediction of the Standard Theory 14

4.2 Prediction of the Contextual Theory 17

5 Chapter: Evaluating Truth in Science 24

5.1 Deriving Inference to Best Theory 25

5.2 Explanatory Criteria Satisfaction 28

5.3 Evaluating Scientific Propositions in Inference to Best Theory 30

6 Chapter: Putting Inference to Best Theory into Practice 37

6.1 Theory Evaluation of Commensurable Theories 37

6.2 Theory Evaluation of Incommensurable Theories 38

7 Chapter: Conclusion 41

Bibliography or References 43

Preface

"[A *Universalis Characteristica* requires] a kind of *general algebra* in which all truths of reason would be reduced to a kind of calculus... for the characters and the words themselves would direct the mind, and the errors — excepting those of fact — would only be calculation mistakes"
– Leibniz (1714).

"Truth consists in the agreement of cognition with its object"
– Kant (1801).

"The proper course is to conceive and begin to construct an ideal language, incorporating the formal devices, the sentences of which will be clear, determinate in truth value, and certifiably free from metaphysical implications; the foundations of science will now be philosophically secure, since the statements of the scientist will be expressible (though not necessarily expressed) within this ideal language" – Grice (1975).

1 Chapter: Introduction

Suppose that we have a scientific theory of some domain, say of physics, at some point in time. Suppose further that this theory satisfies all the relevant criteria for being the best explanation of the data (e.g., it is the simplest theory that captures all the data). Does this best theory of the data then count as being "true"? I will argue in this thesis that it does. Specifically, I will argue that human cognition is endowed with what we might call a "science-forming capacity" (Chomsky, 1980), a capacity that allows us to generate intricate theories of sometimes complex data across a range of domains. This endowment comes with a notion of "truth" which I will argue is to be characterized as one that identifies truth with "the best theory of all the available data", where "the best theory" is given by considerations such as simplicity and breadth of coverage. I will argue for this novel conception of truth in science by showing how it poses a major challenge to classical theories of truth by capturing the intuition that scientific theories aim at truth, are treated as true at various points in time, and yet can also later be shown to be false (when new evidence comes in or a better theory is offered). Indeed scientists themselves simultaneously present their theories as true, yet acknowledge that their theories may turn out to be false. Making this case will involve illustrating the puzzle of truth in science coming from the observation that theories are constantly revised, and showing how one particular theory of truth, Inference to Best Theory (IBT), solves the puzzle sufficiently and adequately.

One challenge raised by this approach is to try to make sense of the criteria that govern which theory is "best". Two considerations seem fundamental to rational choice between theories: (i) Quantity: If theory X_i captures more data than theory X_k, *ceteris* *paribus*, X_i is preferred to X_k, and (ii) Manner: If theory X_i and X_k capture the same data, but X_i is *simpler* than X_k, *ceteris paribus*, X_i is preferred to X_k. Quantity directs agents to select theories with greater empirical coverage, while Manner directs agents to select simple theories over complex ones given some fixed range of empirical coverage (this is often invoked as "Ockham's Razor"). Together these principles derive a preference ranking over candidate theories, with the best theory being the simplest one that captures all the data. We identify this best theory with the "true" theory of the data. Second, we try to derive Quantity and Manner from similar principles of rational communication formulated in Grice (1975). Because we are considering science as a cognitive faculty, principles of cognitive science such as Gricean principles of communication become immediately relevant. We show that this connection might give principled foundations to principles such as "Ockham's Razor," which otherwise seem unmotivated (compelling though it may be).

This thesis is a contribution to cognitive science, epistemology, philosophy of science, philosophy of language, and foundations of pragmatics.

The scope of this thesis is limited to characterizing truth from the point of view of science only. The relationship between scientific truth and truth in general is undoubtedly interesting; however, characterizing any one of them in precise detail would still be needed to characterize the relationship between them both. And since there has been vast amounts of literature, from antiquity onwards, dedicated to the topic of truth in general, this thesis will offer a novel contribution taking a different route, adding to the scarce literature on scientific truth in specific.

Although this thesis aims to magnify truth from the lens of science, and cognitive science in particular, there are important practical consequences of characterizing scientific truth. For example, in policy theory, the role that committees like the Natural Sciences and Engineering Research Council of Canada (NSERC) play in constructing optimal governmental policies is central. Scientific theories need to be evaluated with tremendous care and scrutiny as many policies are constructed on the basis of scientific theories. A simple example illustrates this point: if scientists and engineers believe that a building will not collapse based on theories of physics, mechanics, and mathematics, it is absolutely crucial that these theories explaining why the building will not collapse are actually true, for human lives are at stake. Human lives, not just human cognition, presuppose the truth of scientific theories.

The presentation of this thesis is as follows: first, we'll outline our domain of inquiry, science; second, we'll articulate the puzzle of truth in science that this thesis aims to solve; third, we'll characterize two distinct predictions for solving the puzzle of truth in science showing that only one of them sufficiently solves it; and lastly, we'll focus on *how* this theory of truth is able to sufficiently solve the puzzle of truth in science, and will derive the principles of the theory from formal considerations coming from Gricean pragmatics.

2 Chapter: Science

"Science is the search for truth" – Linus Pauling (1998).

This section will survey the landscape of the scientific domain. First, we'll discuss what science is as a system and what science does as an activity. And second, we will introduce the notion of the science faculty, an innate structure of human cognition that forms theories and explanations based on evidential and epistemic considerations.

2.1 What it is and What it Does

Science can be construed in at least two distinct ways. On the one hand, science is a system of methodology that is based on evidence, which is to say that science is a belief that depends on action.[1] But, science as a system has to be put to use, for after all, it is simply just a belief if not put into practice. And so, on the other hand, science is also an activity, it is a process of knowledge acquisition, something scientists do in order to acquire knowledge. In other words, science is also an action that depends on a system of methodology, or more simply, a belief. In this way, it is in particular a cognitive act, an act that depends on cognition.

This thesis is dedicated to characterizing truth in science as a system of methodology or belief. However, the following points in this section will briefly characterize truth in science as a cognitive act, i.e., an action that depends on cognition.

---

[1] We use the term "action" to mean "something that produces an observable effect", where "something" doesn't necessarily need to refer to a human agent.

When we say that an act depends on cognition, it is a specific type of cognition that seems to be most relevant. With respect to the concept of truth, actions partly depend on truth in cognition, like true (or false) beliefs. For example, if it is true to believe that eating poison is innocuous, then acting out that belief should cause no harm. But, if it is false to believe so, then not acting out that belief should cause no harm. In general, the following point can be made: truth in human cognition partly directs human action, i.e., humans have a tendency to act according to what is believed to be true and avoid what is believed to be false. This has obvious survival advantages.

But, there still remains a fundamental question pertaining to the concept of truth of science as a cognitive activity: *when are we convinced that a belief in science is actually true*? Two immediate answers surface. We are never convinced or we're sometimes convinced, but depending on the circumstances.

The first answer is highly unsatisfying for at least the following reason: if we are never convinced that a belief in science is actually true, then why do we go about our daily lives acting as if it were true? It seems to be a fact about the world the people *do* at least on occasion act is if they believe that certain scientific propositions are true. We *talk* about when the Big Bang might have occurred (which presupposes that it did), we allow surgeons to open our chests and work on our bodies, we sit in airplanes, we invest large sums of money into bodies such as NSERC (and believe this is a wise investment), and so on. Our actions thus indicate belief in the truth of certain propositions; not only do we talk as if we believe certain scientific propositions to be true, we put our money where our mouth is, risking our lives on these beliefs and investing our money in the hope of gaining more of these beliefs. So, if scientific beliefs are never true, then we are forced to solve the question of why we act as we do in such a truth-less world.

The second answer, however, provides a more satisfying account than the first whilst vindicating the notion that science seeks the truth for the following reason: if we are sometimes convinced that a belief in science is actually true, as seems to be the case, but are so-convinced in a way that depends on the circumstances, then much of our actions do appear to be partly directed by truth in cognition. Even right now, we act as if scientific beliefs are somehow true, but true depending on various factors. So, if scientific beliefs are sometimes taken to be true, then we are faced with solving the problem of what role truth plays in our scientific beliefs and how to reconcile the fact that scientific theories change. This thesis aims to provide such a solution by characterizing the notion of "truth" in science as a cognitive notion with certain properties that can be examined empirically.

2.2 Universal Grammar and the Science-Forming Faculty

Humans are born with certain capacities that are genetically encoded in human biology. This is why children are able to learn natural languages and, say rocks and fish are not. No matter how many environmental stimuli serve as the input to the rock's and fish's development, they are just not genetically endowed with the capacity to acquire natural language. We call this innate capacity to acquire natural language, the language faculty. Human cognition has several components that work together to generate some outcome. One component of human cognition is the language faculty.[2] As Chomsky (1988) points out, when "presented with data, the child's language faculty forms a language, a computational system of some kind that provides structured representations of linguistic expressions that determine their sound and meaning". Furthermore, "the grammar of a particular language is an account of the state of the language faculty after it has been presented with data of experience; universal grammar (UG) is an account of the initial state of the language faculty before any experience" (*ibid*, p. 60-61). This process can be schematically captured as follows: evidence from the world → the language faculty filter → a particular language is formed → structured representations of expressions that map sound to meaning. UG is an account of what the language faculty filters in and out of cognition.

In parallel to the Chomskyan thesis of the language faculty, this thesis will argue that human biology is endowed with an innate science-forming faculty[3]. This science-forming faculty is another distinct component of human cognition.[4] Informally, we define the science faculty as the innate capacity to form scientific theories about the world depending on the evidential data presented to the science-forming faculty. Further, we define scientific theories about the world formed by the science faculty as a computational system of some kind that generates structured representations of scientific expressions that map observed phenomena to truth conditions.[5]

---

[2] Other components include the memory faculty, the attention faculty, the learning faculty, etc.

[3] We note that questions of modularity (e.g., (Chomsky, 1980), (Fodor 1983)) immediately arise with respect to the place of the science faculty within the mind. We leave this matter for future work.

[4] However, it may turn out to be highly intertwined with the language faculty.

[5] Following tradition, from (Frege, 1878) to (Davidson, 1967), meaning can be identified with truth-conditions, cf. (Lycan, 2000).

Humans evidently have the capacity to construct, understand, and appreciate scientific theories, while rocks and fish do not. And just like the language faculty, this can be explained by the fact that humans are born with these capacities, genetically encoded in human biology. Gopnik (2009) has argued for the existence of such a science-forming faculty in humans by studying the ways children think about the world during each stage of cognitive development. Gopnik has conducted numerous experiments over the last few decades that have supported the following idea: the ways in which scientists learn about the world via the scientific method are strikingly similar to the ways in which children learn about the world via the stages of child development. As Gopnik writes,

> "Plato and other philosophers asked 'how can we know so much about the world?' The scientific answer is that methods of experimentation and statistical analyses seem to be programmed into our brains even when we are tiny babies. Very young children unconsciously use these techniques to change their causal maps of the world. Those programs allow babies, and so the rest of us, to find the truth" (*ibid*, p. 105).

We can depict the process of child development as follows: evidence from the world → the science faculty filter → a particular theory is formed → structured representations of expressions that map observed phenomena to truth-conditions. Accordingly, a theory of "Universal Methodology" (UM) should also be able to account for what the science faculty is like prior to any experience.

Furthermore, as noted by Chomsky, where "complex intellectual structures are developed in an essentially uniform way on the basis of limited evidence, we have hopes of finding something significant about human nature, since it is natural to account for the fact on the basis of assumptions about the initial state of the mind" (1980, p. 249). He suggests scientific creation and understanding as a potential case study. The history of science has shown that scientists often develop rich explanatory theories that are vastly underdetermined by the evidence available, and non-scientists are often able to appreciate the act in retrospect. This would make sense if there were a species-invariant science-forming capacity with certain specified characteristics.

This thesis builds on Gopnik's and Chomsky's proposal that a component of human cognition that is genetically encoded into our biology is the ability to form a computational system of science. An indispensable part of this science faculty is the concept of truth. The purpose of this thesis is to characterize truth from the point of view of this science-forming faculty. As noted, I will argue that truth in science is to be equated with "best theory of the data at hand". Two uncontroversial "best theory" considerations are breadth of coverage (Quantity) and simplicity of theory (Manner). I will derive these considerations from Gricean principles of rational cooperative action.

3 Chapter: Truth in Science

> "In apprehending a scientific truth we pass, as a rule, through various degrees of certitude" – Gottlob Frege (1878).

This section will introduce the problem of truth in science. The problem poses a substantial obstacle for any theory of truth, so no mention of any particular theory will be necessary. First, we'll discuss what a proposition in science is before moving on towards what the puzzle of truth in science is all about. Generally, the problem of truth in science is the problem of how to adequately account for the truth of scientific propositions that undergo epistemic revision due to shifts in evidential contexts of evaluation.

3.1 Scientific Propositions

Traditionally, a proposition is something that is capable of being true or false. A true proposition is one for which the content of that proposition corresponds to what is the case in the actual world, and a false proposition is one for which this is not the case. Now, let us suppose that a scientific proposition is one that is presupposed as being true in some scientific domain. There are at least two strong motivations that support this supposition. Firstly, natural language users talk and act as if propositions in science are true, at least the propositions that have survived scrutiny by the scientific community and experimental replicability. And secondly, to deny this is to deny one of the most fundamental goals of science, for certainly, science does not attempt to assert falsehoods about the world, nor is science mere 'conjecture'. The force of scientific speech acts is *assertive*: the scientist asserts that their propositions are true, and that this truth is supported by evidence.

Next, a theory is to be construed as a set of propositions that is closed under entailment. In specific, a scientific theory, X_i, is a set of scientific propositions closed under logical consequence. A conjunctive proposition, P, is the intersection of all propositions, $\{p_1, \ldots, p_n\}$, in P. A scientific theory, X_i, can be represented with the conjunctive proposition, such that $X_i =$ P. Hence, a scientific theory is a set of scientific propositions that is closed under both conjunction and entailment.[6] In specifying any theory X_i, we can identify it with its basic inventory: a set of primitive symbols defining the language, a basic ontology, and a set of axioms. We will call the set of *theorems* proved by X_i its *range*; and we will call the number of symbols used in its basic inventory its *complexity*. For example, if X_i is a scientific theory containing ten axioms, while X_k is a scientific theory containing those ten axioms plus an additional axiom, and X_i and X_k have the same set of primitive symbols and ontology, then X_i is less complex than X_k. When this is the case, we will sometimes say that X_i is *simpler* than X_k. If the set of theorems derivable from X_k is a strict superset of the set of theorems derivable from X_i, we will say that X_k has a greater range than X_i.

3.2 Puzzle of Truth and Scientific Propositions

Despite scientific propositions aiming for truth about the world, it is a historical fact that there exist scientific propositions that have turned out to be false given some new evidence about the world.[7] And so, the puzzle of scientific propositions can be put as follows: *were these now false scientific propositions also false back when they were*

[6] Cf. (Alchourron, Gardenfors, & Makinson, 1985; Partee et al., 1990; Buss, 1998).

[7] See (Kuhn, 1970) and (Thagard, 1992) for fairly comprehensive, though not exhaustive, list of examples in the history of science, and (Lakatos, 1976) in the history of mathematics. Most cases have to do with epistemic shifts due to evidential updates, e.g. scientific discoveries and revolutions.

considered to be true? And if so, was it rational to treat these falsehoods as being true? And if not, then what are the logical laws that govern propositions that undergo truth-value shift, and how do these laws operate? We will turn now to discuss two prominent theories of truth, and will argue that neither one on its own provides a satisfying solution to the puzzle. This will motivate the development of a revised theory of truth, what is called the Inference to Best Theory theory of truth.

4 Chapter: Characterizing Truth in Science

> "We divide all truths that require justification into two kinds, those for which the proof can be carried out purely by means of logic and those for which it must be supported by facts of experience" – Frege (1878).

There are at least two distinct theories of truth, the standard theory and the contextual theory. We will discuss each in turn as follows: first, we'll outline what each theory of truth is all about in terms of its history and content, and then we'll discuss in particular what that theory of truth predicts with respect to our puzzle regarding the truth of scientific propositions.

4.1 Prediction of the Standard Theory

The standard theory of truth can be traced back to (Frege, 1878; 1892), but modern formulations of Frege's theory can found in (Tarski, 1933; 1944).[8] Under Tarski's formulation, the truth of all propositions lies in their correspondence or matching to what is actually the case in the world via *truth conditions*.[9]

1) Equivalence of form (T): ... is true if and only if ...,

is used as a schemata with sentence X on the right side (in the metalanguage), and the name of X (denoted by *X* here) on the left side, in order to derive:[10]

---

[8] Frege's theory of truth is to treat the truth of all propositions as a truth function, i.e. truth is a function of propositions, where a *truth function* is a function from the set of truth-values to a truth-value. In classical logic, Frege's truth functions obey the principle of bivalence which states that it is necessarily the case that for any proposition q, either q is true or q is false. In this way, a truth function is just a way of representing truth as a property of the world, while falsity is the absence of this property.

[9] A truth condition is a specification of what the world needs to be like in order to satisfy the truth function.

[10] Tarski et al. (1953) notes that to avoid paradox, the metalanguage must be richer than the object language. Here, we assume that some fragment of English is the object language (that consisting of scientific statements in some domain including vocabulary from that scientific domain), and that the

2) Convention T: X is true iff X.

This criterion consisted of what Tarski called "the semantic conception of truth".[11] For modern extensions of Tarski's program, especially for natural language semantics, see Montague (1974).

To help focus the discussion, we will take Frege's and Tarski's theories of truth to be the canonical standard theory of truth. However, other candidates include correspondence theories of truth (e.g., (Russell, 1905; 1940)) and deflationism (e.g., (Horwich, 1990)).[12] In any case, what is important to remember for any version of the standard theory of truth is the following: a proposition, X, is either true or false depending on how the actual world really is. In this way, the canonical standard theory of truth takes truth to be an observer-*independent* phenomenon, i.e., truth only depends on how the world is, not on how the cognitive agent perceives it to be. It will be shown that this theory is necessary, but not sufficient for explaining how humans actually employ the concept of scientific truth in cognition.

Let's now move into what the canonical standard theory of truth predicts for propositions in science. Taking a paradigm example of a scientific proposition in physics, let X be the conjunctive proposition of Newton's theory of Universal Gravitation. By substituting X into Convention T, X is true iff propositions in Newton's theory are the case in the actual world. Note that it is a historical fact that X was considered to be a true proposition. As it so happens, it has been shown that propositions in Newton's theory of Universal Gravitation turn out to not be the case in the actual world, and so by modus

---

metalanguage includes English as well as vocabulary from scientific domains, as well as set theory, first-order logic, etc.

[11] Cf. (Field, 1976) for a re-analysis of (Tarski, 1933).

[12] Cf. (Glanzberg, 2006) for a compendium of truth theories.

tollens, it is predicted that X is now false given the canonical standard theory of truth. But what is important for present concerns is the following question pertaining to our original puzzle: was the conjunctive proposition of Newton's theory, X, which is now false also false back when it was considered to be true? Again, the canonical standard theory of truth predicts that X was also false back when it was considered to be true. This is so because X, the truth conditional description of the actual world right now is inconsistent with X.

This result predicted by the canonical standard theory of truth can be generalized to the class of scientific propositions: for any scientific proposition, X, if X is now considered to be false, then X was also false back when it was once considered to be true. This generalization of the canonical standard theory only captures the fact that scientific propositions can always be shown to be false.[13] However, this is a major problem for the following reasons.[14] First, although scientific propositions still aim for being true propositions, they might never hit the target. This is especially puzzling because if the canonical standard theory predicts that scientific propositions aim for possibly unreachable truth, then we are hard pressed to answer why we ever act as if some scientific propositions *are* true. For instance, copious amounts of time, natural resources, and money are invested into scientific research that is deemed to be true, e.g., the science underlying applications in the industrial sectors, civil infrastructures, environmental sciences, medical and pharmacological sciences, practical engineering, and general technological implementations. And more importantly, time, resources, and money aren't

---

[13] In tune with the traditional philosophy of science (Shapere, 1966), and hence, falsificationism (Popper, 1959).

[14] However, cf. (Sellars, 1963) for another argument against correspondence (and "the myth of the Given").

the only things that depend on the truth of scientific propositions, human *lives* are at risk. Scientific truth dictates the survival of humanity and development of the world at large.

And second, scientific propositions aren't even treated as being true at points in time when they were once considered to be true. We are thus forced to account for why we thought and acted as if these propositions *were* true. Moreover, if the past is any guide to the present and future, we might be led to conclude that most of the scientific propositions we currently assume to be true will also turn out to be shown to be false. So why, then, do we act as if they are indeed true? And why does that strike us as the rational thing to do?

It will be shown that an alternative theory, the contextual theory of truth, overcomes these two major drawbacks of the canonical standard theory whilst retaining the possibility that propositions in science might be shown to be false, but only given a relevant evidential context of evaluation. However, a version of this contextual theory on its own will give rise to new difficulties. My own proposal will aim to overcome these problems.

4.2 Prediction of the Contextual Theory

We'll now discuss a highly influential alternative theory of truth: the contextual theory. Historically, the contextual theory of truth can be traced back to multiple authors at different times for different reasons.[15] As such, it will be best to focus on two of the

---

[15] For e.g., the contextual approach has been used to solve other problems, like pronoun resolution (Reinhart, 1983), indexicality (Schlenker, 2010), demonstratives (Kaplan, 1978), predicates of personal taste (Lasersohn, 2005), disagreement (Kolbel, 2006), actuality (Lewis, 1970), and knowledge (Nozick, 1981; Lewis, 1996), that are taken to be customary in contemporary circles. Note: Lewis (2001) hints at a notion contextual truth.

most compatible versions of the contextual theory. First is the coherence theory (Davidson, 1983)[16], and second is the Inference to Best Theory theory, along the lines of (Peirce, 1898; Quine, 1936; 1951; Harman, 1965; Thagard, 1975; Putnam, 1983; Wright, 1994; Lipton, 2004).

Davidson (1983) is a defense of a coherence theory of truth.[17] In it, Davidson provides an argument that "purports to show that coherence yields correspondence" (Lepore 1986, p. 307). This is important because the particular version of contextual theory that this paper will adhere to will follow Davidson on this point. What a coherentist says is something like the following,

3) Coherence T: *X* is true at c iff c coheres with X,

where *X* is the name of a proposition, X is the truth condition, and c is the context of evaluation containing sets of all of the available facts or evidence. Treating coherence as a context dependent phenomenon, one can formally capture coherence between propositions via entailment relations, such as that 'c coheres with X' is to be identified with 'c is entailed by X'.[18] Following (Williamson, 1996; 2001), the following equivalence can also be made: evidence = knowledge. And so, the context parameter, c, is the context of evaluation containing sets of everything that is known to be the case at that context.

A crucial difference between Tarski's Convention T and Coherence T is the context parameter, c. But, it is important to note that Coherence T is just Convention T plus the additional context parameter. Thus, the coherence theory yields the standard

---

[16] Cf. (Bradley, 1914) for early defense of coherentism.
[17] Davidson's coherence theory extends to knowledge as well, but for the time being, we focus on the coherence theory of truth.
[18] Cf. (Lehrer, 2000) for coherence as (traditional) logical entailment.

theory with the addition of the context parameter. And in this way, the coherence theory of truth takes truth to be an observer-*dependent* phenomenon, i.e. truth depends on both how the world is and on how the cognitive agent perceives it to be. This observer-dependence is crucial in overcoming the difficulties laid out for the standard theory above. Firstly, theory-bound propositions do aim for truth and can successfully hit the target – if they cohere with the evidence at hand. As a result, scientific propositions can now be treated as being true at points in time when they were once thought to be true – having been shown to be false *now* does not entail falsehood at prior points in time because the shift in truth-value may have been induced by a change in the evidential basis. And thirdly, since the context-dependence of propositional truth-values can now explain why we think and act as if they are true (or were true in the past), this also helps to explain why it is (or was) rational for us to do so – it is rational to believe true propositions.

However, we will argue that although this move to a context-dependent notion of truth is necessary, it is still not sufficient for explaining how humans actually employ the concept of scientific truth in cognition. To see this, suppose X_i and X_k are two versions of Newton's theory of Universal Gravitation, where X_i is the set containing all propositions in *Principia*, and X_k contains X_i plus an additional axiom, *Water is H_2O*.[19] It would be predicted that X_i and X_k are both "coherent" theories since they both provide explanations for the same evidential context of evaluation, viz. the evidential domain of all physics. However, they are treated differentially: given X_i, X_k would not be a serious candidate theory. We clearly have some criteria which enter into the choice between otherwise

[19] Any true proposition, or set of true propositions, consistent with physics would make the same point.

"coherent" theories. Without considering and clarifying the nature of these criteria, the coherence theory is at risk of becoming a variant of post-modern conceptions of truth relativism.[20] Not only does truth depend on the facts of the actual world and observer-dependent perceptions of the world, truth also depends on a ranking of our descriptions of these perceptions (i.e., on a ranking of candidate theories of the data). To overcome this problem, we'll need to discuss another contextual element[21] as construed in Inference to Best Theory (IBT) (Kuipers, 2004; Douven, 2005). In the philosophy of science literature, it has been argued that Inference to Best Explanation is better construed as Inference to Best Theory, and efforts are made to make this explicit in formal detail. However, we adopt IBT as a working model with respect to its description of how the science-forming faculty actually operates, not focusing just on the general abstract mechanics of IBT that is independent of cognition.

First, some formal considerations are needed. We are going to treat the context parameter, c, as representing all of the evidence and knowledge available at c across all relevant scientific domains. For instance, in the natural sciences, c represents all of the evidence and knowledge in physics, chemistry, and biology. But, in order to restrict our domain of scientific discourse, we take a subset of c, call this c_e, as representing all of the

---

[20] In some formulations, post-modern truth says that every proposition is equally true (as any other) regardless of evidential and doxastic considerations. This type of 'anything-goes' relativism will not be adopted as it reduces the concept of truth to triviality, subjectivism, and similar deflationary views, for all propositions are said to be construed as true (or false) and equally so.

[21] IBT can be viewed as a contextual theory that employs a complex contextual parameter that is both sensitive to the context and a partial ordering relation on the context. Cf. (Kratzer, 1981) for application in modality, (Kaplan, 1978) for demonstratives, and (Garcia-Carpintero & Macia, 2006; Chalmers, 2010) for general possible world semantics.

evidence and knowledge available at c_e within some particular scientific domain.[22] For instance, c_e can represent all of the evidence and knowledge in the domain of, say, physics. Furthermore, a candidate scientific theory, X_i, is just a set of propositions whose conjunction entails all of the available evidence within a particular scientific domain, c_e.

It is commonly assumed in philosophy of science that discovery of theories often follow a form of "inference to the best theory". This inference is of course not a deductive inference from the data, but some kind of abductive inference to some set of propositions that entail the data at hand. What needs to be explicated is the notion "best." By what criteria are candidate theories ranked for "goodness"? Some criteria that have been proposed are: scope, simplicity, mechanism, precision, unification, and fruitfulness. A candidate scientific theory, X_i, is better than another, X_k, if X_i explains more types of scientific evidence than X_k, if X_i explains some scientific evidence more simply than X_k, if X_i explains more of the underlying mechanisms of some scientific evidence than X_k, if X_i explains some scientific evidence with more precision than X_k, if X_i discloses more new scientific evidence than X_k, and if X_i explains more unification underlying scientific evidence than X_k. Here we will focus on two uncontroversial criteria, and will use them in our revised definition of truth.

First, take two candidate theories X_i and X_k of some set of data, c_e. Recall our definitions of the *range* of a theory and the *complexity* of a theory from section 3.1. We will say that X_i is to be preferred to X_k if X_i's range entails a larger subset of c_e than X_k's range. Second, given some fixed range, we will say that if X_i and X_k have the same range

[22] *All* of the evidence follows from the Principle of Total Evidence; cf. (Carnap, 1947) for original inductive formulation, and (Grice, 1975) for pragmatic formulation, in particular, Grice's second sub-maxim of Quality, "say what you have evidence for saying". See Section 5.1 where we make the connection to Gricean principles more explicit.

(i.e., prove the same set of theorems), X_i is to be preferred if X_i is *simpler* than X_k (this is just a formalization of Ockham's Razor). From these two uncontroversial principles it follows that the best theory of the data is the simplest theory that entails all of the data:

Definition: *Best Theory*: We will say that the *best theory* of data c_e is the simplest theory that entails c_e.

We will now use this as the basis of our revised theory of truth:[23]

4) Best Theory T: "X_i" is true at c_e iff X_i is the best theory of c_e.

Let's now move into what the IBT theory of truth predicts for propositions in science. Continuing with our paradigm example in the history of physics, let X_k be the conjunctive proposition of Newton's theory of Universal Gravitation, let X_i be the conjunction of propositions of Einstein's theory of General Relativity, and let c_e be the set of all evidence and knowledge in physics up to now (especially evidence and knowledge regarding the physics of gravity). By substituting X_k into Best Theory T, X_k is true at c_e iff X_k is the best theory of the actual world given c_e, where X_k is the best theory satisfying the most explanatory criteria than any other competing candidate. Of course, it is a historical fact that X_k was considered to be true in candidate theory X_k given a different evidential basis $c_{e'}$. This is so because it was the best theory of the evidence available *then*. But, as it so happens every once in awhile in the history of science, new evidence has come along updating $c_{e'}$ to c_e, and another competing candidate theory, Einstein's theory of General Relativity, X_i, has shown that X_k is not the best theory given

[23] Other candidates of the contextual theory of truth include pragmatism (e.g. James, 1907; Dummett, 1959; Rorty, 1979), revisionism (e.g. Kremer, 2006; Gupta, 2011), verificationism (e.g. the later-Wittgenstein; Quine, 1951), identity theories (e.g. Lewis, 1966), and prosentential theories (e.g. Grover, Camp & Belnap, 1975; Brandom, 1994). We leave comparison for future work.

c_e. And so, it is predicted that X_k is now false at c_e since Einstein's candidate scientific theory X_i is the best theory given c_e. That is, like the coherence theory of truth, but unlike the standard theory of truth, Best Theory T allows the truth of propositions to vary depending on the context of evaluation; and like the standard theory, but unlike the coherence theory, it provides *objective* criteria in the definition of truth, namely, the range and complexity of a theory.

This result predicted by the IBT theory of truth can be generalized to the class of scientific propositions: let X_i^p be any scientific proposition of scientific theory X_i, and X_k^p be any scientific proposition of another scientific theory X_k, where X_i and X_k both seek to explain c_e. If X_i^p is now falsified because X_k is actually the better theory of c_e than X_i, then X_i^p can still be said to be true, but only when this better theory, X_k, was not a live candidate competing with X_i in explaining for c_e. This generalization of the canonical contextual theory now captures the following points: i) that scientific propositions of X_i can always be shown to be false at c_e if new evidence comes along that X_i cannot capture, or if some alternative theory X_k turns out to be a better theory of some scientific phenomena than X_i; ii) that scientific propositions aim for being true propositions while successfully hitting the target, but only with respect to X_i given c_e; and iii) that scientific propositions now shown to be false can nevertheless be treated as being true at points in time when they were once considered to be true. It has now been shown that IBT theory of truth is the only theory that overcomes the major drawbacks of the canonical standard theory and the coherence theory when it comes to solving the puzzle of truth in science, i.e. IBT is necessary and sufficient for explaining how humans actually employ the concept of scientific truth in cognition.

5 Chapter: Evaluating Truth in Science

> "Truth is that concordance of an abstract statement with the ideal limit towards which endless investigation would tend to bring scientific belief, which concordance the abstract statement may possess by virtue of the confession of its inaccuracy and one-sidedness, and this confession is an essential ingredient of truth" – Peirce (1901).

We have proposed IBT as a theory of truth in science in response to challenges faced by prominent alternatives. The broader goal is to make sense of the so-called "science-forming capacity" (Chomsky 1980; Gopnik, 2009), and we have focused here on the role that *truth* plays in this capacity. By emphasizing the cognitive aspects of scientific practice, the *agents* that develop and use scientific theories gain immediate importance. In this section, we begin (in section 5.1) by trying to *derive* the best theory principles of maximal scope and minimal complexity from Gricean principles of rational communication. If scientific theories are *assertions*[24], then we might expect the norms governing assertion to enter into how we judge the quality of any given scientific speech act. Second, we will briefly speculate on how various factors that have been thought to be relevant to theory-selection different from scope and simplicity might nevertheless be reducible to some combination of the two (Section 5.2). Third, we show how IBT operates in practice, with respect to historical cases of theory-change (Section 5.3). In particular, we discuss IBT in relation to notions of commensurability and incommensurability of theories. We argue that IBT commits us to the idea that theories

[24] Cf. (Stalnaker, 1978; 1999) for explication of assertions.

are *never* incommensurable, and moreover that this follows from rationality principles demanding that an agent account for *all* the data at hand.

5.1 Deriving Inference to Best Theory

Part of the motivation for IBT as a theory of truth is that it makes much of human action come out as rational. Under the standard theory, human action would appear to be irrational insofar as it commits humans to believing in falsehoods. Under the coherence theory, human action would also appear to be irrational insofar as the choice between theories is (somewhat) arbitrary. We have reduced the concept of "truth" in science to best theory considerations – specifically, maximal range and minimal complexity. What makes these principles rational? Here we will try to show how these principles can be derived from pragmatic principles of rational communication, in the fashion of (Grice, 1957; 1975; 1989). Since theories are here understood as sets of propositions that are asserted by some scientist, rational norms of assertion enter into evaluations of theories. We argue here that these norms suffice to derive the principles of range and simplicity. Specifically, they follow as special cases of Grice's *Maxim of Quantity* and *Maxim of Manner*, respectively.

First, consider the Maxim of Quantity (MQ). Here we follow the formulation in Fox (2007):

5) Maxim of Quantity: If S_1 and S_2 are both relevant, and the speaker knows that both are true, and S_1 is more informative than S_2, the speaker must use S_1.

MQ is a general principle that guides speakers to convey more information than less, whenever it is relevant and they have evidence that supports its truth. For example, MQ guides speakers to use *A and B* instead of *A or B* whenever the former is known to be true (and both are relevant).

We suggest that MQ is the governing force behind the explanatory criterion of scope. Suppose that a scientific proposition is *relevant* if it speaks to the truth or falsity of some subset of data in c_e. Under this conception of relevance, MQ would guide a scientist to utter the most informative relevant proposition that they know to be true. Here, "informative" is defined in terms of logical entailment. Thus, p is more informative than q if p entails q, and q does not entail p. Putting these pieces together, the best assertion a scientist could make, with evidence c_e defining what is relevant, is that proposition which entails all the data in c_e. This means that a theory X_i is better than another theory X_j if X_i's range is a greater subset of c_e than X_j's range.

The pressure for maximal range is constrained by a pressure for minimizing complexity. Historically, the explanatory criterion of simplicity has been especially important. One of its first formulations can be traced back to William of Ockham (c. Ockham, 1322). He articulated a fundamental pragmatic law of parsimony, sometimes called "Ockham's Razor", that says something like the following: do not posit plurality if it's not necessary. In other words, simpler theories are preferred to more complex theories if the increased complexity is redundant, i.e., if it does not explain anything that could not already be explained by the simpler theory.

While the preference for simpler theories clearly exists, what isn't clear is what the motivation for such a preference is. In particular, why should it be rational to prefer simpler theories to more complex theories? As far as we know, a principled rationale for the principle has yet to be formulated. Here we submit that the principle can be derived as a special case of Grice's Maxim of Manner (MM). Here we adopt a recent formulation from (Katzir, 2007; Fox, 2008; Schlenker, 2008):

6) Maxim of Manner: If sentence S_1 *and* S_2 conveys the same information as S_1 itself, do not assert S_1 *and* S_2.

MM is an efficiency principle that punishes redundancy. For example, it explains why it is odd to say *John has a German Shepherd and he has a dog;* since the same information would be conveyed by the simpler *John has a German Shepherd*, under MM this simpler variant should have been used.[25]

We suggest that this efficiency principle guiding linguistic pragmatics is behind the preference for simpler theories. If two scientific theories X_i and X_j both entail the evidence c_e, and X_j is equivalent to X_i conjoined with additional propositions (as axioms), then X_i is to be preferred to X_j. For example, the theory of Newtonian Mechanics, X_i, compared with X_i *and water is* H_2O, is preferred because the axiom *water is* H_2O is redundant.

Summarizing, then, if two scientific theories provide explanations for the same information in c, but one is simpler than another, then the science faculty forces the scientist to prefer the simpler one. This guarantees that the scientific theory with the simpler explanation for some fixed evidential context is used for IBT computation. We have tried to show how the scope and simplicity principles can be derived from rational principles of human communication.

---

[25] A more general statement of MM might go something like this (e.g., Fox 2008): *If p is a propositions that contains proposition q as a constituent, and p and q convey the same information, then do not assert q*. This formulation would generalize beyond conjunctive forms, so as to rule out sentences like *John was born in Paris or France*, which is equivalent to the second disjunct *John was born in France*. See Fox (2007), Chierchia et al. (2008), Singh (2008a, b). We leave for future work a fuller investigation of the proper formulation of MM and its relation to Ockham's Razor.

5.2 Explanatory Criteria Satisfaction

Since last section (5.1) discussed and derived the criteria of scope and simplicity from pragmatic principles of rational communication, we can now elaborate and discuss some more of the satisfaction requirements for other explanatory criteria that allow for one scientific theory, X_i, to be more explanatory than another, X_k, given a fixed evidential context of evaluation, c_e. We re-list some of these other relevant criteria that have been proposed as governing the choice between theories as follows: mechanism, precision, unification, and fruitfulness. This section will speculate on how these criteria can be reduced to scope, simplicity, or a combination of the two. We'll start with discussing each criterion first, and then describe their satisfaction requirements before explicating how such a reduction is possible. Future research will need to sort out the details of such a derivation.

The underlying mechanism, included in the explanation of some scientific theory accounting for evidence c_e, refers to some general explanation that explains for the particular evidence in c_e. For instance, before Mendeleev's chemical theory, which still serves as the basis for modern chemistry even today, there was no candidate theory that provided a basis for the underlying mechanisms of chemical reactions. By postulating the periodic table of elements into Mendeleev's theory, his theory of chemistry now provided such a basis for the underlying mechanism behind c_e. Generally, in order to satisfy the mechanism criteria, a scientific theory must explain c_e with more general (rather than particular) explanations than another candidate scientific theory. Mechanism may be reduced to the simplicity criterion in the following way: the more general a theory of some evidence is, and hence, the greater its underlying mechanism, the simpler that theory is.[26]

Precision refers to the level of description and accuracy of a theory that is required to capture c_e. For instance, the level of description in which classical theories of logic (e.g. Aristotle's) account for c_e is less precise than the level of description in which mathematical theories of logic (e.g. Quine's) account for c_e. By postulating formal tools and methodologies used in mathematical theories of logic, the level of description and accuracy required to account for c_e provided a more detailed account of c_e than other candidate theories. Generally, in order to satisfy the precision criteria, a scientific theory must explain c_e with more detailed descriptions and accuracy than another candidate theory. Precision may be reduced to the scope criterion in the following way: the more descriptive and accurate a theory of some evidence is, and hence, the greater its precision, the greater that theory's range is.

Unification refers to the capability of a scientific theory to make generalizations that underlie the evidence in c_e in a way that relates each piece of evidence to one another. For instance, string theory is able to make generalizations that underlie the evidence in c_e in a way that relates subsets of c_e, call them c_e^i and c_e^k, to one another. So goes string theory: it is able to explain the evidence already explained in quantum mechanical theory, c_e^i, and the evidence already explained in General Relativity theory, c_e^k, under one unified theory, i.e. the evidential subsets, c_e^i and c_e^k, are related in string theory that explains the superset c_e. In order to satisfy the unification criteria, a scientific theory must have a greater capability to make generalizations that underlie the evidence

[26] If an explanation is more general, the explanation requires less number of propositions; if an explanation is more particular, the explanation requires more number of propositions.

in c_e in a way that relates each subset to one another than another candidate theory. Unification may be reduced to a combination of both the scope and simplicity criteria in the following way: the more unified a theory of some evidence is, the simpler is that theory's precision.[27]

Fruitfulness refers to the capability of a scientific theory to explain (or predict) new evidence, call this new evidential context c_e^k, other than what is already contained in c_e^i. For instance, Kepler's theory of planetary motion explained the evidence in c_e^i, but Newton's theory, which built and improved upon Kepler's theory, was able to explain even more new evidence in c_e^i, call this updated evidential context c_e^k, than Kepler's theory. Le Verrier's mathematical calculations using Newton's theory theoretically predicted the existence of another planet in our solar system, Neptune. Neptune was actually discovered and empirically confirmed, but only after its existence was mathematically predicted from Newton's theory. And so, Newton's theory had more capability to explain new evidence other than what was already contained in c_e^i than Kepler's theory. In order to satisfy the fruitfulness criteria, a scientific theory must have the capability to explain new evidence, updating c_e^i to c_e^k, than another candidate theory that can explain the evidence in c_e^i. Fruitfulness can also be reduced to a combination of scope and simplicity in the following way: the more fruitful a theory of some evidence is, the greater range is covered by that theory's underlying mechanism.[28]

5.3 Evaluating Scientific Propositions in Inference to Best Theory

[27] Where simplicity takes precedence over precision, and hence, scope.
[28] Where scope takes precedence over underlying mechanism, and hence, simplicity.

We are now in position to articulate how IBT works in practice. We'll explicate the general framework without inserting specific examples into IBT first as more generalizations are needed before putting it to the test.

Theories are constructed to explain the evidence at hand for at least two different motivations: i) there are theories that are constructed to explain regularities in the evidential context, c_e. For instance, in the natural sciences, especially physics, theories aim to account for data in the physical world, data that are typically characterized as universals. For instance, as we've already seen, the Newton-Einstein case shows how Newton's and Einstein's theory of gravity were constructed to explain the observed regularities regarding gravity in the physical world, call these evidential regularities, c_e. Both Newton's and Einstein's theories aimed to explain the same evidence, c_e. We call these types of theories, commensurable theories. ii) There are theories that are constructed to explain evidential irregularities that are not included in the evidential context, c_e.[29] These irregularities, also referred to as evidential anomalies, are pieces of evidence that are somehow in conflict with commensurable theories. In this way, the evidential context c_e^i must be updated to c_e^k to include the evidential anomalies in the set of all evidence and knowledge within a given domain of science. For instance, distinct astronomical theories have recently been put forth to account for Pluto's planet-hood status. Call the theory constructed to account for the astronomical evidence c_e^i before 2006, X_i, and call the theory constructed to account for the updated evidential context c_e^k after 2006, X_k. X_i aims to explain the evidence c_e^i, while X_k aims to explain the evidence

---

[29] This may arise because of various factors, e.g., the anomalous data point may be considered as an exception (an outlier), or the anomaly may have just been recently observed, thus updating the evidential context.

c_e^k, i.e. X_i and X_k are constructed to explain for different evidential sets. We call these types of theories, incommensurable theories.

Evaluating commensurable theories is fairly straightforward as they both compete to provide the better explanation of the same evidence at hand, c_e. These theories account for the same evidential context with respect to scope. However, evaluating incommensurable theories is a little different. These theories compete to provide the better explanation of two distinct evidential context sets, c_e^i and c_e^k. In this way, they aren't even talking about the same thing, for they do not account for the same evidential context in terms of scope, i.e. hence, they are incommensurable with respect to the evidence they are seeking to account for. This complicates theory evaluation because incommensurable theories can't even be evaluated against one another. However, a way around this complication is to make incommensurable theories commensurable. This can be done as follows: first, a common evidential denominator must ground the basis which incommensurable theories aim to explain, and second, a prediction must be generated to account for this common evidence set by the theory that originally did not explain it. Specifically, the first step is accomplished by assessing which evidential context has the greatest range over c, and the second step is accomplished by generating a proposition to account for the evidential anomalies, if possible.

A simple example will illustrate how to evaluate incommensurable theories. Suppose that the evidential context is such that $c_e = \{1, 2, 3, 4, 5, 6\}$. Further suppose scientific theory X_1 explains $\{1, 2, 3\}$, X_2 explains $\{1, 2, 3, 4\}$, X_3 explains $\{1, 2, 3, 4, 5\}$, and so does X_4. Take the following ordering to hold in terms of the simplicity of each scientific theory: $X_1 > X_2 > X_3 > X_4$. Now, firstly, we assess which theory accounts for the evidence with the greatest range over c_e, namely, X_3 and X_4. We use {1, 2, 3, 4, 5} as the common evidential denominator. However, X_1 doesn't account for {4, 5} and X_2 doesn't account for {5}. And so, secondly, X_1 must be able to generate propositions concerning {4, 5}, likewise, X_2 for {5}. If this is not possible, then only X_3 and X_4 are evaluable. Finally, suppose X_3 and X_4 are the only theories that account for the same evidential set, then since $X_3 > X_4$, X_3 is true because it is the best theory of the available data in c_e following IBT computation. Let's now sketch a general algorithm for IBT as follows,

1. Determine the type of competing scientific theories,
2. If commensurable, go to step 3; else:
    a. Evaluate which evidential context set is the one with the greatest range,
    b. Use this evidential context to generate a prediction of the theory that originally did not explain for it,
    c. If no prediction can be generated, the competing scientific theories are incommensurable (stop the algorithm); if a prediction can be generated, go to step 3,
3. Compute theory evaluation following the explanatory criteria satisfaction requirements (as stipulated in 5.1 and 5.2.): a candidate scientific theory, X_i, is better than another, X_k, given the evidence c_e,
    a. Scope: if $range(X_i) > range(X_k)$.
    b. Simplicity: if $| X_i | < | X_k |$.[30]
4. Determine which theory is the better one, then infer that the best theory contains scientific propositions that are true with respect to the evidence it aims to explain (compute Best Theory T):
    a. If X_i is the best theory of c_e, then infer that X_i is true at c_e.

This formulation assumes that theories are sometimes incommensurable, and tries to take that into account. In particular, it rests on the assumption that if a theory cannot derive certain data, the theory can be ignored. For example, X_1 did not account for {4,5}, and if no general extension of a theory X_i could be found, we argued that X_1 should then

[30] Note that mechanism, precision, unification, and fruitfulness should also be checked for satisfaction, we focus on scope and simplicity since they can derive the rest, however.

be dropped from consideration. This is symptomatic of actual scientific practice, i.e., the scientist will typically ignore evidence that does not fit in the working theory of choice (calling them exceptions or outliers). In this way, actual scientific practice is sometimes not rational when it comes to accounting for *all* the known evidence. However, we'll now argue that incommensurability can be licensed by considerations of rationality, for it is not rational to *ignore* any evidential data. We argue that there is *always* at least a trivial extension of any theory, namely, stipulation of the data points it has no general statement for. For example, one extension of X_1 is $X_{1'} = X_1$ *and 4 and 5.* The algorithm for theory comparison should now consider $X_{1'}$ as a live candidate together with X_4 and X_5, with the simplest being preferred. Similar remarks apply, *mutatis mutandis*, to the other theories that were contenders but were dropped for "incommensurability". A revised algorithm is as follows,

1. Determine the type of competing scientific theories,
2. If commensurable, go to step 3; else:
    a. Evaluate which evidential context set is the one with the greatest range,
    b. Use this evidential context as a common denominator and append any data that is unaccounted for by a theory as axioms,
3. Compute theory evaluation following the explanatory criteria satisfaction requirements (as stipulated in 5.1 and 5.2.): a candidate scientific theory, X_i, is better than another, X_k, given the evidence c_e,
    a. Scope: if $\text{range}(X_i) > \text{range}(X_k)$.
    b. Simplicity: if $|X_i| < |X_k|$.[31]
4. Determine which theory is the better one, then infer that the best theory contains scientific propositions that are true with respect to the evidence it aims to explain (compute Best Theory T):
    a. If X_i is the best theory of c_e, then infer that X_i is true at c_e.

---

[31] Note that mechanism, precision, unification, and fruitfulness should also be checked for satisfaction, we focus on scope and simplicity since they can derive the rest, however.

We have argued that considerations of rationality demand that a theory be responsible for *all the data* at hand; if we have no general statement covering these data, they must be stipulated as basic axioms of the theory.

Future research needs to focus on empirical considerations that result from IBT as a theory of truth in cognition. The formal mechanics of IBT have stipulated necessary conditions for truth-assignments of scientific theories and propositions, but what is more importantly lacking are the sufficient conditions that underlie the principles of scientific truth. For instance, a fundamental question that needs to be empirically confirmed concerns the priority rankings (or orderings) of explanatory criteria satisfaction. For example, consider two theories, X_i and X_k. X_i offers the most simple explanation for the evidence c_e that has the least scope. X_k offers the least simple explanation but for the evidence $c_{e'}$ that has the greatest scope. Which theory is actually preferred by scientists? Or another example, X_i offers the most fruitful explanation of c_e, but X_k offers the most unified explanation of c_e. Which theory is actually preferred by scientists? That is, which theory is actually perceived as "true"? There are many of these sorts of questions that arise pertaining to the rank orderings of explanatory criteria that can only be found out empirically.[32]

In addition to priority rankings, there are also issues of the explanatory criteria themselves. We've mentioned only six criteria, but this is just a preliminary, non-exhaustive list. Operationalizing these criteria are wanting, and that is mostly of empirical concern, perhaps leading to the discovery of more basic criteria than the ones discussed here.

---

[32] In fact, there are at least 720 possible rank orderings, i.e. 6!, 6 = | explanatory criteria |.

And finally, sensitivity to the formal syntax of a language when we compare the complexity of theories is of great empirical import. For example, suppose we have two theories, X_i and X_k. Suppose they both share the same set of primitive symbols, {A, B, C, ...}, but differ in terms of their set of connectives in the following way: X_i has the set $\{\neg, \wedge\}$, and X_k has $\{\neg, \wedge, \vee\}$. Logically, $(A \vee B)$ is syntactically equivalent to $\neg(\neg A \wedge \neg B)$ as they are interdefinable. But, as a matter of empirical concern, which formula is actually preferred in the process of scientific interpretation? That is, which theory-bound formal syntax is preferred, X_i or X_k? Empirical experiments are needed to verify these kinds of syntactic processing issues which we leave open to future cognitive science research.

6 Chapter: Putting Inference to Best Theory into Practice

We can now illustrate how IBT actually works in practice using the general algorithm in the previous section as a template to input specific examples. We'll discuss two examples, one employing commensurable theory evaluation, and another employing incommensurable theories.

6.1 Theory Evaluation of Commensurable Theories

An example of commensurable theory evaluation in IBT comes from the history of chemistry, the chemical revolution. We adopt some preliminary data from (Thagard, 1992) pertaining to the chemical revolution as follows: let $c_e = \{E_1, \ldots, E_8\}$,

$c_e = \{$ E_1 = In combustion, heat and light are given off,
E_2 = Inflammability is transmittable from one body to another,
E_3 = Combustion only occurs in the presence of pure air,
E_4 = Increase in weight of a burned body is exactly equal to weight of air absorbed,
E_5 = Metals undergo calcination,
E_6 = In calcination, bodies increase weight,
E_7 = In calcination, volume of air decreases,
E_8 = In reduction, effervescence appears $\}$,

let X_i be the conjunctive proposition of Lavoisier's theory of combustion,

$X_i = \{$ P_1 = Pure air contains oxygen principle,
P_2 = Pure air contains matter of fire and heat,
P_3 = In combustion, oxygen from the air combines with the burning body,
P_4 = Oxygen has weight,
P_5 = In calcination, metals add oxygen to become calxes,
P_6 = In reduction, oxygen is given off $\}$,

and let X_k be the conjunctive proposition of Priestley's theory of phlogiston,

$X_k = \{$ P_1 = Combustible bodies contain phlogiston,
P_2 = Combustible bodies contain matter of heat,
P_3 = In combustion, phlogiston is given off,
P_4 = Phlogiston can pass from one body to another,

P_5 = Metals contain phlogiston,
P_6 = In calcination, phlogiston is given off }.

We now need to compute theory evaluation in order to figure out if X_i or X_k is the better theory given c_e. We'll focus on the most relevant explanatory criterion that highlights the difference between X_i and X_k, the scope criterion. Recall that in order to satisfy the scope criterion, a scientific theory must explain more evidence in c_e than another candidate theory. And so, using our example, this translates into something like the following: in order to satisfy the scope criterion, X_i must explain more evidence in c_e than X_k (or vice versa). As it turns out, X_i explains more of the total evidence in c_e than X_k does,

X_i explanations = { explains((P1, P2, P3), E1),
explains((P1, P3), E3),
explains((P1, P3, P4), E4),
explains((P1, P5), E5),
explains((P1, P4, P5), E6),
explains((P1, P5), E7),
explains((P1, P6), E8) },
X_k explanations = { explains((P1, P2, P3), E1),
explains((P1, P3, P4), E2),
explains((P5, P6), E5) }.

Assuming that scope is the only relevant explanatory criterion in this case of competing chemical theories, we can then infer that X_i is the best theory given c_e, not X_k, as X_i explains more of the evidence in c_e than does X_k. Therefore, we infer that X_i is true given c_e.

6.2 Theory Evaluation of Incommensurable Theories

For an example of incommensurable theory evaluation, let's just focus on Inference to Best Evidence in particular as it is stipulated in our general algorithm for IBT. We employ an example taken from the history of astronomy. Let X_k be the conjunctive proposition of a scientific theory of planetary astronomy where "Pluto is a planet" is just one proposition in the conjunctive set, let X_i be the conjunction of propositions of another scientific theory of planetary astronomy where "Pluto is not a planet" is just one proposition in the conjunctive set, let c_e^i be the set of all evidence and knowledge in planetary astronomy at time t_i, and let c_e^k be the set of all evidence and knowledge in planetary astronomy at time t_k, where $t_i>t_k$. Furthermore, suppose X_i aims to explain c_e^i and X_k aims to explain c_e^k.

First, we need to compute evidence evaluation. And so, computing Inference to Best Evidence, we can conclude that c_e^k must be the better evidential context set because it represents the evidence at time t_k, where $t_i>t_k$, i.e. c_e^k is the most recent updated evidence of the world. Now, we set c_e^k as the common evidential context of evaluation that X_i and X_k aim to explain, otherwise X_i and X_k would be incommensurable.

Next, we need to generate a prediction for X_i as X_i originally aimed to explain c_e^i, not c_e^k. The main difference between c_e^i and c_e^k is the following point: c_e^k contains more evidence of dwarf planets in the Kuiper Belt than c_e^i does, one in particular, Eris has been measured as being equal to or bigger than Pluto itself. A new theory, X_k, had to be constructed to account for the new evidence in c_e^k, and it turns out that in X_k, Pluto not being a planet provides a better explanation for c_e^k rather than keeping Pluto as a planet in addition to a host of other dwarf planets in the Kuiper Belt similar to Pluto itself. And so, given what the new evidence in c_e^k says, X_i turns out to predict in contrast to X_k that Pluto not being a planet would not provide a better explanation for c_e^k seeing as in X_i, it is the case that Pluto is a planet. X_i further predicts that these other dwarf planets similar to Pluto found in the Kuiper Belt should be classified as fully-fledged planets. X_i given c_e^k

predicts that it is the case that our solar system has, say, 100 planets, while X_k given c_e^k says that is it the case that our solar system has 8 planets.

Lastly, we just need to compute theory evaluation between X_i and X_k given c_e^k. In this instance, the most relevant explanatory criterion is simplicity. Recall that in order to satisfy the simplicity criterion, a scientific theory must explain the evidence in c_e with less explanations than another candidate theory. And so, using our example, X_i must explain the evidence in c_e^k with fewer explanations than X_k (or vice versa). It turns out that X_k requires fewer explanations to account for the evidence in c_e^k than X_i, i.e. explaining why 8 planets are planets requires fewer statements than explaining why 100 planets are planets. Assuming that simplicity is the only relevant explanatory criterion in this case of competing astronomical theories, we can infer that X_k is the best theory given c_e^k, not X_i, as X_k explains the evidence in c_e^k using less explanations than X_i does. Therefore, we infer that X_k is true given c_e.

7 Chapter: Conclusion

It has been shown that the standard theory and the coherence theory of truth are necessary, but not sufficient to account for how the embedded concept of truth in the science-forming faculty of human cognition is actually used to direct human action. Furthermore, it has also been shown that Inference to Best Theory theory is both necessary and sufficient for adequately explaining this cognitive phenomenon. To summarize: the standard theory guarantees that "true" scientific propositions correspond to the actual world, but is observer-independent; the coherence theory guarantees that "true" theory-bound propositions correspond both to the actual world and the observer's perception of the world, but doesn't allow for an ordering of observer-dependent perceptions in terms of which ones are "better" than others; IBT guarantees that "true" theory-bound propositions correspond to the world and the observer's theory of the world in such a way that the satisfaction of explanatory criteria (e.g., scope and simplicity) of what makes a theory the "best" is to be identified with "true". Concomitantly, IBT is able to rationalize how humans really employ the concept of scientific truth in cognition and in action through pragmatic principles of communication. Many empirical issues concerning IBT remain to be addressed, such as the extent to which simplicity principles are sensitive to the formal syntax of the language in which a theory is expressed; the relation between theory and evidence in domains where "empirical" evidence seems difficult to adduce, such as logic and mathematics; and what the limits of the science-forming capacity are. Much of Universal Grammar entails that certain languages are humanly attainable while others are not. Likewise, a theory of the innate science-forming capacity, call it "Universal Methodology", would seem to entail that certain theories are humanly expressible while others are not. What is the class of such theories? And how does their truth, defined here as "best theory" of some given evidence, relate to pre-theoretic notions of Truth? While these questions are difficult, our proposal to investigate scientific truth as an aspect of human cognition leads to a clear path for addressing these questions, namely the empirical and theoretical methods of the cognitive sciences.

References

Alchourron, C. E., Gardenfors, P. & Makinson, D. (1985). "On the Logic of Theory Change: Partial Meet Contraction and Revision Functions". *The Journal of Symbolic Logic*, 50;2, p. 510-30.

Bradley, F. (1914). *Essays on Truth and Reality*. Clarendon Press: Oxford, England.

Brandom, Robert (1994). *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Harvard University Press: Cambridge, USA.

Buss, Samuel (1998). "An Introduction to Proof Theory". In *Handbook of Proof Theory* (Buss, S.), Elsevier: Amsterdam, Netherlands.

Carnap, Rudolf (1947). "On the Application of Inductive Logic". *Philosophy and Phenomenological Research*, 8;1, p. 133-48.

Cappelen, H. & Hawthorne, J. (2009). *Relativism and Monadic Truth*. Oxford University Press: NY, USA.

Chalmers, David (2010). *The Character of Consciousness*. Oxford University Press: NY, USA.

Chierchia, G., Fox, D. & Spector, B. (2008). "The Grammatical View of Scalar Implicatures and the Relationship between Semantics and Pragmatics". In *Handbook of Semantics* (Portner, P., Maienborn, C. & von Heusinger, K.), Mouton de Gruyter: NY, USA.

Chomsky, Noam (1980). *Rules and Representations*. Columbia University Press: NY, USA.

Chomsky, Noam (1988). *Language and Problems of Knowledge: The Managua Lectures*. MIT Press: Cambridge, USA.

Davidson, Donald (1967). "Truth and Meaning," *Synthese*, 17, p. 304-23.

Davidson, Donald (1983). "A Coherence Theory of Truth". Reprinted in *Truth And Interpretation: Perspectives on the Philosophy of Donald Davidson* (LePore, 1986). Basil Blackwell: Oxford, England.

Douven, Igor (2005). "Empirical Equivalence, Explanatory Force, and Inference to Best Theory". *Poznan Studies in the Philosophy of the Sciences and the Humanities*, 83;1, p. 281-309.

Dummett, Michael (1978). *Truth and Other Enigmas*. Harvard University Press: Cambridge, USA.

Fodor, Jerry (1983). *The Modularity of Mind*. MIT Press: Cambridge, USA.

Frege, Gottlob (1878). "Begriffsschrift". Reprinted in *From Frege to Godel: A Source Book in Mathematical Logic* (van Heijenoort, 1967). Harvard University Press: Cambridge, USA.

Frege, Gottlob (1892)."Über Sinn und Bedeutung", in *Zeitschrift für Philosophie und Philosophische Kritik*, 100, p. 25-50.

Field, Hartry (1976). "Tarski's Theory of Truth". Reprinted in *Truth and the Absence of Fact* (Field, 2001). Oxford University Press: NY, USA.

von Fintel, K. & Heim, I. (2005). *Intensional Semantics*. MIT Course Notes.

Fox, Danny (2007). "Free Choice Disjunction and the Theory of Scalar Implicatures". In *Presupposition and Implicature in Computational Pragmatics* (Sauerland, U. & Stateva, P.), Palgrave-Macmillan: NY, USA.

Fox, Danny (2008). "Two Short Notes on Schenker's Theory of Presupposition Projection". *Theoretical Linguistics*, 34;3, p. 237-52.

Garcia-Carpintero, M. & Macia, J. (2006). *Two-Dimensional Semantics*. Oxford University Press: NY, USA.

Glanzberg, Michael (2006). "Truth". Stanford Encyclopedia of Philosophy Entry.

Gopnik, Alison (2009). *The Philosophical Baby: What the Children's Minds Tell Us About Truth, Love, and the Meaning of Life*. Picador: NY, USA.

Grice, Paul (1957). "Meaning". *Philosophical Review*, 66;3, p. 213-23.

Grice, Paul (1975). "Logic and Conversation", in *The Logic of Grammar* (Davidson, D. & Harman, G.), Dickenson: CA, USA.

Grice, Paul (1989). *Studies in the Way of Words*. Harvard University Press: Cambridge, USA.

Grover, D., Camp, J. & Belnap, N. (1975). "A Prosentential Theory of Truth". *Philosophical Studies*, 27;2, p. 73-125.

Gupta, Anil (2011). *Truth, Meaning, Experience*. Oxford University Press: NY, USA.

Harman, Gilbert (1965). "Inference to the Best Explanation". *Philosophical Review*, 74;1, p. 88-95.

Heim, I. & Kratzer, A. (1998). *Semantics in Generative Grammar*. Blackwell Publishing: Oxford, England.

Horwich, Paul (1990). *Truth*. Oxford University Press: NY, USA.

James, William (1907). "Pragmatism's Conception of Truth". Reprinted in *Truth* (Blackburn & Simmons, 1999). Oxford University Press: NY, USA.

Kant, Immanuel (1801), *The Jäsche Logic*. In *Lectures on Logic* (1992) (trans. by Young, J. M.), Cambridge University Press: Cambridge, UK.

Kaplan, David (1978). "On the Logic of Demonstratives". *Journal of Philosophical Logic*, 8, p. 81-98.

Katzir, Roni (2007). "Structurally-defined Alternatives". *Linguistics and Philosophy*, 30;6, p. 669-90.

Kratzer, Angelika (1981). "The Notional Category of Modality". Reprinted in *Formal Semantics: The Essential Readings* (Portner & Partee, 2002). Blackwell Publishing: Oxford, England.

Kremer, Phillip (2006). "The Revision Theory of Truth". Stanford Encyclopedia of Philosophy Entry.

Kolbel, Max (2004). "Faultless Disagreement". *Proceedings of Aristotelian Society*, 104, p. 53-73.

Kuhn, Thomas (1962). *The Structure of Scientific Revolutions*. University of Chicago Press: Chicago, USA.

Kuipers, Theo (2004). "Inference to Best Theory, Rather than Inference to Best Explanation: Kinds of abduction and induction". In *Induction and Deduction in the Sciences*, Proceedings of the ESF-workshop *Induction and Deduction in the Sciences* (Stadler, F.), Kluwer Academic Publishers: Dordrecht, Netherlands.

Lakatos, Imre (1976). *Proofs and Refutations: The Logic of Mathematical Discovery*. Cambridge University Press: Cambridge, England.

Lasersohn, Peter (2005). "Context Dependence, Disagreement, and Predicates of Personal Taste". *Linguistics and Philosophy*, 28, p. 643-86.

Lehrer, Keith (2000). *The Theory of Knowledge*. Westview Press: NY, USA.

Leibniz, Gottfried (1714). "Letter to Nicolas Remond, 10 January 1714". In Loemker (1969), p. 654.

Lewis, David Kellogg (1970). "Anselm and *Actuality*". *Nous*, 4, p. 175-188.

Lewis, David Kellogg (1996). "Elusive Knowledge". *Australasian Journal of Philosophy*, 74;4, p. 549-567.

Lewis, David Kellogg (2001). "Truthmaking and Difference-Making". *Nous*, 35;4, p. 602-15.

Lipton, Peter (2004). *Inference to Best Explanation*. Routledge/Taylor: NY, USA.

Lycan, William (2000). *Philosophy of Language*. Routledge: NY, USA.

Montague, Richard (1974). *Formal Philosophy*. Yale University Press: NY, USA.

Nozick, Robert (1981). *Philosophical Explanations*. Harvard University Press: Cambridge, USA.

Ockham, William (c. 1322). Summa Logicae. In *Ockham's Theory of Propositions: Part II of the Summa logicae* (trans. by Freddoso, A. & Schuurman, H.) University of Notre Dame Press: Indiana, USA.

Partee, B., ter Meulen, A. & Wall, R. (1990). *Mathematical Methods in Linguistics*. Kluwer Academic Publishing: Norwell, USA.

Pauling, Linus (1998). *Linus Pauling On Peace – A Scientist Speaks Out on Humanism and World Survival*. Rising Star Press: NY, USA.

Peirce, Charles S. (1898). *Reasoning and the Logic of Things*. Harvard University Press: Cambridge, USA.

Peirce, Charles S. (1901). "Truth and Falsity and Error". In *Dictionary of Philosophy and Psychology* (James Mark Baldwin). Macmillan: NY, USA.

Popper, Karl (1959). *The Logic of Scientific Discovery*. Routledge: NY, USA.

Putnam, Hilary (1983). "On Truth." In *How Many Questions? Essays in Honor of Sidney Morgenbesser* (Cauman, L., Levi, I., et al.), Hackett: Indiana, USA.

Quine, Willard van Orman (1936). "Truth by Convention". Reprinted in *Philosophy of Mathematics* (Benacerraf & Putnam, 1964). Prentice-Hall Inc.: Toronto, Canada.

Quine, Willard van Orman (1951). "Two Dogmas of Empiricism". *The Philosophical Review*, 60, p. 20–43.

Reinhart, Tanya (1983). *Anaphora and Semantic Interpretation*. Croom Helm: London, England.

Rorty, Richard (1979). *Philosophy and the Mirror of Nature*. Princeton University Press: Princeton, USA.

Russell, Bertrand. (1905). "On Denoting", *Mind*, 14, p. 479–493.

Russell, Bertrand. (1940). *An Inquiry into Meaning and Truth*. Penguin University Books: Baltimore, USA.

Sellars, Wilfrid (1963). *Science, Perception, and Reality*. Routledge: London, England.

Shapere, Dudley (1966). "Meaning and Scientific Change". In *Mind and Cosmos: Essays in Contemporary Science and Philosophy* (Colodny, R.). University of Pittsburgh Press: NY, USA.

Schlenker, Philippe (2008). "*Be Articulate:* A Pragmatic Theory of Presupposition Projection". *Theoretical Linguistics* 34:3, p. 157-212.

Schlenker, Philippe (2010). "Indexicality". Chapter in *Handbook of Formal Philosophy* (Hansson & Hendricks, 2010).

Singh, Raj (2008a). "On the Interpretation of Disjunction: Asymmetric, Incremental, and Eager for Consistency", *Linguistics and Philosophy*, 31;2, p. 245-60.

Singh, Raj (2008b). *Modularity and Locality in Interpretation*. PhD dissertation, MIT.

Stalnaker, Robert (1999). *Content and Context*. Oxford University Press: NY, USA.

Tarski, Alfred (1933/1944). "The Semantic Conception of Truth and The Foundations of Semantics". Reprinted in *Truth* (Blackburn & Simmons, 1999). Oxford University Press: NY, USA.

Tarski, A., Mostowski, A. & Robinson, R. (1953). *Undecidable Theories*. North-Holland.

Thagard, Paul (1975). "The Best Explanation: Criteria for Theory Choice". *Journal of Philosophy*, 75;2, p. 76-92.

Thagard, Paul (1992). *Conceptual Revolutions*. Princeton University Press: Princeton, USA.

Williamson, Timothy (1997). "Knowledge as Evidence". *Mind*, 106, p. 717-41.

Williamson, Timothy (2000). *Knowledge and its Limits*. Oxford University Press: NY, USA.

Wright, Crispin (1994). *Truth and Objectivity*. Harvard University Press: Cambridge, USA.