Globethics Repository



Turing Machines and Semantic Symbol Processing

This page was generated automatically upon download from the Globethics Repository. More information on Globethics see https://www.globethics.net. Data and content policy of Globethics Repository see https://repository.globethics.net/pages/policy.

Item Type	Article
Authors	Yee, Richard
Publisher	Saint Anselm College
Rights	With permission of the license/copyright holder
Download date	2026-11-11 22:28:20
Link to Item	http://hdl.handle.net/20.500.12424/172805

Volume V, No. 1 Spring 1993

LYCEUM

What Wisdom is According to Heraclitus the Obscure
Marie I. George

Truth, Adequacy and Being in Spinoza's ***Ethics***
Lance Byron Richey

Turing Machines and Semantic Symbol Processing:
Why Real Computers Don't Mind Chinese Emperors
Richard Yee

Leibniz on Innate Ideas
Shaun A. Champagne

A Publication of the
Saint Anselm Philosophy Club

LYCEUM

EDITORIAL STAFF

Matthew Hayes David Dumaresq

Shaun Champagne

BOARD OF EDITORIAL CONSULTANTS

William E. Carroll James O'Rourke

Kevin M. Staley Robert Anderson

Paul O'Reilly David Banach

Kelly S. Spoerl

PHILOSOPHY CLUB OFFICERS

David Banach	Moderator, Managing Editor
Matthew Hayes	President
David Dumaresq	Vice-President
Maryalice Baker	Secretary
Shaun Champagne	Treasurer

All manuscripts for review should be addressed to the **LYCEUM,** Saint Anselm College #1698, 100 Saint Anselm Drive, Manchester NH 03102-1323. The **LYCEUM** welcomes submissions on disk in any IBM or Macintosh word processing format, but a printed copy is still required. Footnotes should follow the Turabian style sheet.

The **LYCEUM** would like to thank the Student Government of Saint Anselm College for their support.

LYCEUM

Table of Contents

What Wisdom is According to Heraclitus the Obscure 1
Marie I. George

Truth, Adequacy and Being in Spinoza's ***Ethics*** 21
Lance Byron Richey

Turing Machines and Semantics Symbol Processing:
Why Real Computers Don't Mind Chinese Emperors 37
Richard Yee

Student Article:

Leibniz on Innate Ideas 61
Shaun A. Champagne

What Wisdom is According to Heraclitus the Obscure

Marie I. George

Heraclitus of Ephesus, one of the earliest Greek philosophers recorded in history, lived from the 6th to 5th century BC. Few of his writings have come down to us: somewhat more than one hundred fragments, a number of which are of dubious authenticity.[1] This dearth of extent writings is not the only obstacle to learning from the sage of Ephesus. A second problem arises when people become familiar with Heraclitus' teachings in the form in which other authors represent them, rather than in their original form. Aristotle, for instance, mentions Heraclitus in a number of places, most often to ascribe to him a form of relativism, and several other philosophers do the same[2], the end result of which is that people tend to think that he is a skeptic and relativist. Moreover, the more popular and oft-cited fragments such as DK 91a: "it is not possible to step twice into the same river" seem to confirm this belief.

Doubtless there are fragments which seem to clearly express a relativistic position, for instance, DK 102 reads: "To god all things are fair and just, whereas humans have supposed that some things are unjust, other things just." However, there are numerous fragments concerning the nature of wisdom which give quite the opposite impression. My intention here is to examine these fragments, first and foremost because the insights they contain about the nature of wisdom generally do not receive the attention they deserve, and secondly

---

[1] Even some of the thought-to-be authentic fragments are to be questioned when they do not agree with Heraclitus' overall teaching. One suspects that fragments such as DK 124: "The most beautiful order [in the universe?] [or the or (this?) most beautiful universe] [says Heraclitus] is a heap of sweepings, piled up at random," are incomplete or would have a quite different meaning if placed in their original context.

[2] This is true of Plato; cf. ***Cratylus***, 402a. Aristotle says in the ***Metaphysics***, 1012a25: "The saying of Heraclitus that all things are and are not seems to make all things true . . . and so no statement is true." And Thomas Aquinas says: "Certain people thought that all bodies were mobile, and thinking them to be in continuous flux thought that no certitude of the truth of things could be possessed by us. . . . as Heraclitus said 'it is not possible to touch twice the water of a flowing river' as the Philosopher relates in ***Metaphysics***, Bk. IV." (q. 84, art. 1 of the prima pars of the ***Summa Theologiae***).

because they reveal a Heraclitus who is a humble pursuer of wisdom, rather than a skeptic and relativist. Indeed the two goals are not unrelated inasmuch as the Heraclitus' reputation as a relativist prevents a certain number of people from examining and profiting from his doctrine. I do not intend to address all the arguments advanced by those who maintain that he is a relativist, but rather, I aim at the more modest goal of showing how an alternate reading is plausible.[3]

The fragments which will be examined all respond in some way to the question: what is wisdom according to Heraclitus? Following the method which he himself seems to have adopted, I will start out by determining what wisdom is not for him, and then to proceed to a clearer and clearer notion of what it is. Two passages are rather explicit on the point of what wisdom is not, albeit the second is doubtfully attributed to the sage of Ephesus.

> The knowledge of many things does not give understanding, else it would have given it to Hesiod and Pythagoras, as well as to Xenophon and Hecatea. (DK 40)[4]
>
> Pythagoras,[5] the son of Mnesarchos, cultivated research more than any other person, and having chosen these writings, he

---

[3] There are a very large number of commentaries on Heraclitus, the more interesting of which propose groupings of the fragments by theme. The theme which I am addressing, namely, that of wisdom in Heraclitus, is specifically addressed by a certain number of other commentators who cite remarks about Heraclitus made by other ancient authors and who make reflections worthy of examination (for example, Kostas Axelos, ***Héraclite et la Philosophie*** [Paris: Les Editions de Minuit, 1962]). However, to review all the literature in detail would exceed article-length proportions. Also, as Charles H. Kahn notes: "every age and philosphical perspective, from Cratylus to the Neoplatonists and the fathers of the Church, projected its own meaning and its own preoccupations onto the text of Heraclitus. This is a familiar enough phenomenon in the history of ideas. . . . But Heraclitus is an acute case." ***The Art and Thought of Heraclitus***, (Cambridge: Cambridge University Press, 1979), p. 84. This is a second reason why I make no claim to presenting an interpretation which is definitive, but only one which is plausible and which presents matter provocative of serious reflection about the nature of wisdom.

[4] The Greek text is that found in ***Heraclitus, Fragments and Translation with a Commentary*** by T.M. Robinson (Toronto: University of Toronto Press, 1987). Translations are my own unless otherwise noted.

> made of them his own wisdom—which was only much learning (πολυμαθίην)and bad art. (DK 129*)

There can be no doubt from these passages that wisdom is not to be found in a pure accumulation of the knowledge of many things, or polymathy, as he called it. What it is about beyond or instead of the accumulation of knowledge is not indicated in the first-cited fragment, but a clue is given in the second where he says that such polymathy is bad art; for bad art means art without order. Wisdom, then, must involve order: it cannot be a collection of undigested and disordered knowledge.

Another description Heraclitus gives of wisdom is more positive, although it too involves negation:

> None of those whom I have heard has realized that that which is wise is separated from all things. (DK 108)

The statement is ambiguous: Does it mean the same thing as what Anaxagoras[6] means when he calls the mind unmixed? Or is it an earlier expression of Plato's notion that wisdom requires separation from the world of change? Or does it mean yet something else? Fortunately there is another fragment which speaks of wisdom as some one thing apart from other things:

> The one wisdom is unique: it accepts and does not accept the name of 'Zeus'. (DK 32)

---

[5] It is surprising Heraclitus is so critical of Pythagoras, given that Pythagoras is known for coining the word 'philosopher'. Heraclitus even goes so far as to call Pythagoras "chief captain of swindlers." (DK 81a).

[*] Indicates a fragment which is possibly spurious.

[6] Cf. Aristotle, ***De Anima***, 405a17.

This particular manner of characterizing wisdom seems strange at first sight. What relation could seriously be proposed between the mythological god Zeus and wisdom? Before dismissing this statement as simply an enigmatic remark, let us first ask ourselves a few well-placed questions: Why is wisdom one and unique? And who was Zeus, or rather, who was he supposed to be?

The answer to the latter question is that Zeus was the "Father of gods and men," that is, the one who was the guiding and directing principle of the universe.[7] Thus if we take Zeus as the One who governs all things to their ends, then he is indeed the unique wisdom, and the one separate from all things which Heraclitus spoke about in the fragment cited earlier. If, on the other hand, we take Zeus to be the mythological father of the Greek gods, sitting in his very human form, thunderbolt in hand, on top of Mount Olympus, then we must acknowledge that he is not what we mean by wisdom.[8] We understand, then, this rather indirect definition of wisdom to be saying that wisdom is God himself, the unique being to whom no other is like, the ultimate being who is the cause of the order of all things. This interpretation finds its confirmation in the well-known complete definition of wisdom:

> Wisdom is one thing: to understand the mind which moves all things through all. (DK 41)

Now this mind is the mind of God, and thus wisdom consists in knowing the divine mind.[9]

---

[7] Fragment #30 shows that Heraclitus did not regard "Zeus" as the creator of the universe: "The cosmos, the same for all, no god or man made, but it always was, is, and will be. . . ."

[8] Fragment #15 reveals Heraclitus' scorn for those who believe in the gods of mythology: "If it were not in Dionysius' honor that they make a procession and sing a hymn to shameful parts, their deed would be a most shameful one. But Hades and Dionysius, for whom they rave and celebrate the festival of the Lenaea, are the same." (trans. Robinson).

[9] It is to this same mind that does and does not accept the name Zeus that Heraclitus is referring to when he says: "Thunderbolt steers all things." (#64).

It should be noted that a number of authors argue that Heraclitus is a pantheist.[10] The fragment cited above at very least makes it plain that there is a certain amibivalence in Heraclitus on this point. While it would be interesting to explore further his thought on the matter, it would take us too far from our main purpose which is to examine those fragments which speak of wisdom.

Heraclitus seems to identify wisdom with ***logos*** in other passages:

> Not after listening to me, but after listening to reason (***logos***), it is wise to agree that all things are one. (DK 50)

'Logos', with its multiple meanings, is one of the most difficult words to interpret in ancient Greek philosophy. What does Heraclitus mean by it?

> Of the ***logos***, which holds forever, people forever prove uncomprehending, both before they have heard it and when once they have heard it. For, although all things happen in accordance with this account, they are like people without experience when they experience words and deeds such as I set forth, distinguishing each thing according to its nature, and defining how each properly is. The rest of mankind, however, fail to be aware of what they do after they wake up just as they forget what they do while asleep. (DK 1)

The ***logos*** which people are oblivious to is to be found in the natures of things, properly distinguished from one another and defined. People experience things, but generally do not seem to go beyond this to something which holds forever,

---

[10] Frederick Copleston, S.J. maintains that Heraclitus is a pantheist, although he does add qualifications to this position; cf. ***A History of Philosophy***, vol. I, part 1. (Garden City, New York: Image Books, 1962), p. 60. G.T.W. Patrick, in ***The Fragments of the Work of Heraclitus of Ephesus on Nature***, (Baltimore: N. Murray, 1889), p. 61, maintains: "[Heraclitus'] system was pure pantheism."

the ***logos***, the "ratio," the whatness of things. They do not attain universal knowledge from what they experience; they fail to recognize the common nature shared by different things, and thus each particular thing is new to them.[11] Thus in another place Heraclitus insists:

> That is why one must follow that which is common. Though the ***logos*** is common, the many live, however, as though they have private understanding (ἰδίαν φρόυησιν). (DK 2)

Most people spend most of their time engaged in various practical matters. They solve life's problems in a pragmatic way, rarely seeking out any theoretical justification for what they do, and sometimes not needing any theoretical justification, as is the case when it is the matter of making some product, where success or failure is rather obvious. Yet Heraclitus disagrees with the view that each person's particular experience brings him his own unique know-how: if it works well for him, it is good for him, and if something else works well for others, that is good for them. Heraclitus sees this way of living as not being that of the wise person. His view of such people is that they are asleep:

> [Those who are asleep I think Heraclitus calls] labourers and co-producers of what happens in the universe. (DK 75*, trans. Robinson)
>
> They are separated from that with which they live in most continuous familiarity. (DK 72*)

The latter are physically present, but are mentally absent acting as they do without an intellectual grasp of what they are doing.

---

[11] There is a similar theme in #34, and in #17: "Many people do not understand the sorts of things they encounter. Nor do they recognize them even after they have had experience of them, though they think they do."

Heraclitus characterizes the one who is not asleep as being able to see beyond the multiplicity of things to something which is common:

> Those who speak with insight must base themselves firmly on that which is common to all, as a city does upon law—and much more firmly. For all human laws are nourished by one law, the divine. For it holds sway to the extent that it wishes and suffices for all, and is still left over. (DK 114, trans. Robinson)

The people who are asleep prefer their own private world, their own views of things, rather than seeking to put their opinions on an objective basis of what all commonly experience. Another fragment in the same line reads:

> For those who are awake there is a single, common universe, whereas in sleep each person turns away into his own private universe. (DK 89)

The word 'idiot' comes from the Greek word for proper. For the Greeks, a person who had his own ideas was not deemed original, but rather an idiot. Living in his own world, a world alienated from that known to all through common experience, the idiot cannot be reasoned with, for he rejects the common basis for discussion. And thus when Heraclitus says:

> The world is the same for all. . . . (DK 30)

he is not enunciating a mere truism, but is insisting upon a fundamental truth whose recognition is prerequisite to becoming wise.

Statements such as these show that it is far from plain that Heraclitus is the relativist which he is often made out to be. It is further worthy of note that the reason many attribute a relativistic doctrine to him, namely, because of his

statements concerning the continual flux of the objects of sense perception, is not as telling as it might first seem. For Heraclitus does not only say "it is not possible to step twice into the same river" (DK 91a), but also "we step and do not step into the same rivers" (DK 49a), which latter could be reasonably taken to express the fact of experience: objects of sense both change continually as to certain things, and are stable as to others. Indeed, this is what he himself says in DK 12: "As they step into the same rivers, different and [still] different waters flow upon them."

Fragments such as DK 61: "Seawater is very pure and very foul water: for fish drinkable and life-sustaining, for people undrinkable and lethal" are also taken to be expressions of relativism, but here too another interpretation could reasonable be given. Heraclitus need not be taken as saying that things are and are not in the same respect. In fact in the latter fragment he explicitly points out the difference of respect: for fish water is drinkable, and for people undrinkable.[12] Heraclitus had a propensity for formulating paradoxes, and while certain authors take this as amounting to a denial of the principle of contradiction (or in other words as an affirmation of the coincidence of opposites), there is nothing in Heraclitus which prevents us from taking him to intend them to be legitimate and provocative dialectical problems which can be solved by making the proper distinctions. This is the case of the river fragments discussed above; it would also seem to be the case of DK 60: "The road up [and] down [is] one and the same," inasmuch the implicit paradox stimulated

[12] Indeed no one would call Thomas Aquinas a relativist, and yet he makes the same point regarding prudence, and one even wonders whether his use of fish as an example is not drawn from Heraclitus: "If we would put that that science which is about useful things of the political sort, would be the wisdom which is the head of all, it would follow that there would be many wisdoms. For there cannot be some one reason (ratio) concerning those things which are good for all animals; but it is necessary that concerning each animal there be another consideration considering what is the good for each. . . . For as said above, as is health, so too is the good other for men and for fish. In ***In Decem Libros Ethicorum Aristotelis ad Nicomachum Expositio***, (Turin: Marietti, 1933), #1188; cf. also #1187: "And things of this sort cannot be the same for all; as it is manifest that health and the good is not the same for men and for fish."

Aristotle to make the distinctions which he makes in the ***Physics***.[13] Thus when Heraclitus says that "what opposes unites, [and that the finest attunement stems from things bearing in opposite directions, and that all things come about by strife]," (DK 8*) this can reasonably be taken to mean that the truth is only arrived at by examining opposing points of view.[14] At any rate another fragment speaks in rather plain terms about wisdom having an objective basis in reality:

> Sound thinking is a very great virtue, and wisdom is saying what is true and acting in accordance with the nature of things by assenting to it. (DK 112)

The reason why wisdom consists in conforming to nature was mentioned earlier. The order in nature is the product of the unique ***logos***, God, whose thought the wise man seeks to know. The fool is the one who fails to recognize the need to subject himself to nature, prizing instead novelty, originality, imagination.

Note that in this fragment, the Greek word here translated as 'assent' can also mean 'acclaim', 'praise'. This suggests that characteristic of the wise person is not only a certain state of mind (one conformed to nature), but a certain moral attitude, namely that of reverence.[15]

Let us summarize the meanings of '***logos***' in Heraclitus: (1) Divine Mind, source of the order of things; (2) the order of things as caused and perfectly known by the Divine Mind. As for the relation of the human mind to

---

[13] Aristotle, ***Physics***, 202a20 and b13.

[14] Philip Ellis Wheelwright disagrees: "To be sure, the logicizing intellect will undertake to analyze each of these paradoxes into its elements, explaining in just what pair of respects, or in what pair of circumstances, or from what opposite points of view, something is at once such and not-such. But Heraclitus regards the paradox itself, and not its logical transformation, as more truly representing the real state of affairs." ***Heraclitus***, (Princeton: Princeton University Press, 1959), p. 92; cf. also pp. 98, 103-105.

[15] This theme is developed by Duane Berquist of Assumption College, MA in his unpublished commentary on DK 112.

logos: the order in things, product of the Divine Mind, is the measure of the human mind. The wise man sees the order in the world which presents itself commonly to all; whereas the fool see simply a multiplicity of things or sees things from "his point of view," rather than from the point of view of common experience. The wise person has an objective basis for his statement: the world as all experience it; whereas the fool eschews this basis, preferring instead to follow his personal feelings and intuitions.

We now have a better understanding not only of what Heraclitus understands wisdom not to be, but also of what he understands it to be. At this point, however, we must return to our point of departure, the matter of polymathy, the antithesis of wisdom, for there are problems there which we skipped over. First, Heraclitus seems to contradict himself on the subject, for while claiming that much learning does not cause wisdom, he maintains that:

> Lovers of wisdom ought to be very much inquirers into many things. (DK 35, trans. Robinson)

A second and related problem lies in understanding exactly what polymathy is: Is it knowledge of many things? (As is stated in DK 40: "The knowledge of many things does not give understanding.") Or is knowledge of many opinions which people have held about things? (As is implied in DK 129: "Pythagoras, the son of Mnesarchos, cultivated research more than any other person, and having chosen these writings, he made of them his own wisdom. . . .")

As to the first problem, one way of reconciling the fragments which on the one hand blame Pythagoras for his extensive research into the opinions of others, while on the other insist that there is a need for much inquiry is to say that Heraclitus means to tell us that we should inquire into things, rather than into opinions of others. However, this interpretation does not square with his affirmation that knowledge of many **things** does not give understanding. A better approach to solving this problem lies in pointing out that to deny that knowledge of many things gives understanding is not quite the same thing as to

deny any dependence of understanding upon considering many things. Another passage indicates that this is the correct solution:

> Those seeking gold dig up a great deal of earth and find little. (DK 22)

The seeker of wisdom must know how to go through many things, but must also know how to separate what is worthless from what is worthwhile. Heraclitus, then, does not contradict himself, for he is saying that while much inquiry is necessary to attain wisdom, it of itself does not yield wisdom, but in addition discernment is necessary. Indeed, the importance of discernment is a favorite theme in Heraclitus:

> Poor witnesses for people are eyes and ears if they possess uncomprehending souls. (DK 107, trans. Robinson)

Now let us consider the second question concerning polymathy: Does Heraclitus mean to criticize any research into the sayings of others as being bad art? First of all we must point out that that Heraclitus does not necessarily mean by polymathy collecting opinions of others. The word Heraclitus uses in DK 129* ("Pythagoras, the son of Mnesarchos, cultivated research more than any other person, and having chosen these writings, he made of them his own wisdom—which was only much learning [polymathy] and bad art") which has been translated as 'research' is / στορ| α. The modern senses of 'research' are two: first, it can mean investigating a question by going through what others have said on it; secondly, it can mean investigating by experiment. The word seems to have a similar ambiguity in Greek, meaning a learning by inquiry (Heraclitus uses the same word when he says that we must be inquirers into many things), and a historical narration of what one has learnt from others.[16]

---

[16] Both Aristotle's ***Parts of Animals*** and ***De Anima*** are "***historia"***; both are neither simply narration of opinions of others nor Aristotle's own investigations into things.

The ἱστορία of Pythagoras may simply correspond to the investigations he made into different branches of learning (mathematics, astronomy, etc.), in which case Heraclitus is criticizing him for selecting writings out of which he made for himself a hodge-podge liberal arts program. However, even if Heraclitus in fragment 129 is not in fact criticizing Pythagoras for compiling the opinions of others,[17] in view of getting a better understanding of his notion of wisdom it is worth considering whether he does reject considering the opinions of others in favor of considering the things themselves by oneself. Some fragments give the impression that this is in fact the case:

> [Heraclitus, as though he has made some mighty and august utterance, says:] I investigated myself. (DK 101, trans. Robinson)
> Eyes are more accurate witnesses than are ears. (DK 101a)
> Not after listening to me, but after listening to reason, it is wise to agree that all things are one. (DK 50)

However, in other fragments Heraclitus does not inveigh against teachers, but insists rather that one show discernment as to who one listens to:

> What discernment or intelligence do they possess? They place their trust in popular bards, and take the throng for their

[17] Robinson thinks that this is the case: " . . . we can reasonably guess that what made the enterprise 'disreputable' in Heraclitus' eyes was the fact that it was devoid of a basic understanding of how to investigate the real. Instead of listening to the 'common' ***logos*** and looking carefully at the world, Pythagoras chose to compile a 'private' wisdom or philosophy from the views of others. To make matters worse, his compilation involved selection, and selection without reference to a viable selection principle (such as might have been provided by listening to the ***logos***). The result is, necessarily, claims Heraclitus, just an unstructured mound of learning, devoid of insight, for all the training to which Pythagoras subjected himself." Robinson, ***op. cit.***, p. 164 (comments on #129).

> teacher, not realizing that the majority are bad, and only few are good. (DK 104, trans. Robinson)

Indeed he seems to outline two extremes:

> A stupid person tends to become all worked up over every statement [he hears] (DK 87)
> The dogs bark at everyone they do not recognize. (DK 97)

In other words some people are too gullible, and listen to just anyone, whereas others are too critical, and listen to no one, except those already familiar to them. The mean for Heraclitus lies in listening to others, but not just to anyone, but only to the rare person of wisdom:

> It is law also to obey the counsel of a single one. (DK 33, trans. Robinson)
> One man is ten thousand, provided he be very good. (DK 49)[18]

It is better to listen to one person of outstanding insight, than considering the sayings of ten thousand others. But to recognize the one who is good presupposes discernment. Thus while Heraclitus holds that ultimately it is by assenting to the natures of things and not to the speech of a person, even a wise person, which renders one wise, still he recognizes that the words of the wise provide assistance on one's journey to wisdom.

While Heraclitus does recognize the person of superior wisdom as someone who can help one along the road to wisdom, he does not see this as being true of one's equals, much less one's inferiors. In other words, he recognizes the value of teaching, but not of dialogue. And on the face of it, this seems reasonable, for the master is to be listened to as one who knows, whereas

[18] Cf. also #122 on the superiority of one individual.

one's equals are in the same state of ignorance as oneself. Heraclitus has seen the essential of dialectic which is to formulate problems by arguing for both sides of a contradiction as is clear from his practice of formulating paradoxes. However, he has not seen dialectic as means of formulating and pursuing the solution to such dilemmas by discussing with one's equals. And it is certainly unlikely that he would ever have even entertained the notion that the statements of very inferior thinkers are to be regarded as contributions to the advance of learning, and indeed as contributions necessary for its advance.[19]

Returning then to the question of what constitutes polymathy, we are left with the ambiguity as to whether he means a knowledge of many things or of the opinions held by many different people about things. However, one point is clear, be it a knowledge of a variety of disciplines, or knowledge of a variety of thinkers' opinions, the key defect of polymathy lies in the failure to select the writings by listening to the ***logos***, i.e., by measuring what it said in the writings by the order found in things. Wisdom, thus, can neither be found in a random smattering of a variety of disciplines, nor in a random familiarity with what people have said about various aspects of reality.

Two previously cited fragments are worth a closer look, condemning as they do extreme positions which one may adopt regarding supposed truths proposed to one. The first is: "[It is] a stupid person [who] loves to get excited about everything that is said." (DK 87) The other is: "for dogs also bark at whomever they don't know." (DK 97)

The stupid or gullible people of DK 87 are those who lack the critical faculty and who accept without discernment each new idea proposed to them simply because it is new. At the other extreme are those who are habit-bound and who, like dogs, react by immediate opposition to whatever new thing is proposed to them, without taking the time to give it a fair hearing.

There is no doubt that acquiring wisdom will require digging through much material in order to discover the flecks of gold hidden among the dross of inconsequential discourse that is met with in the intellectual life. If the learner

---

[19] As does Aristotle; cf. ***Metaphysics***, 993b12-b18.

accepts everything without discernment, the gold will be buried and confused with the dross; if the learner rejects whatever is proposed, the gold will simply be thrown out with the dross. The seeker of wisdom must not fall into either excess. Open to whatever is true, the true learner will greet what is new as a possible addition to already acquired truth, but will at the same time insist it be carefully scrutinized before being accepted. Neither a slave to the familiar on the one hand, nor to the new on the other, the wise learner uses judgement to sift and then to order and to assimilate whatever has been found to be good.

These warnings about avoiding the extremes in what is proposed to us have far-reaching applications. They are applicable to those who hold to traditional truths for no other reason than that they are traditional, and who for this reason reject all that is new. They too are barking without just cause. What is old and long-received may be, for as much, false, just as it may happen that what is new is not true, but only "bad art." So too, the runners after fashionable ideas are to be placed with the gullible. Thus we see that these extremes may be met in many forms and variations: someone, for instance, may reject old ideas because these are not familiar to his mind, formed as it is by more recent habits of thought. Such a one is also like the dog, although what he is barking at is not the absolutely new, but only the new for him. And someone else may gobble down whatever is presented as old, simply because it is old.

Heraclitus would remind us that it is reason and the natures of things that are to be our measure, and neither the familiar nor the unfamiliar, the new or the old, the traditional or the fashionable. Reason alone, bringing discourse to account through a confrontation with experience, should be our guide.

The road to wisdom, then, is no easy one, as fragment 22 reminds us. Our seeking of the gold will necessarily require much digging and sifting. This difficulty is an inherent consequence of the divine nature of wisdom. Wisdom is not really a fully human possession. We humans must strive with all our force to come to know ever so little of that mind which moves all things through all things, much as a monkey must to acquire some little rudiments of human language (Heraclitus in fact says we are to God what a monkey is to us (DK

83*). Heraclitus also alludes to the difficulty of wisdom in the following fragments:

> The hidden harmony is better than the apparent harmony. (DK 54)
> Nature loves to hide. (DK 123)
> One would never discover the limits of soul, should one traverse every road—so deep a measure does it possess. (DK 45, trans. Robinson)

Indeed Heraclitus goes so far as to say that:

> Human nature does not have right understanding; divine nature does. (DK 78, trans. Robinson)

Wisdom is something divine and not fully within the power of human beings to attain. Aristotle echos this saying that "it is rightly thought that wisdom is not a human possession."[20] And a definition of philosophy given by Plato points to

---

[20] Cf. Aquinas' commentary on Aristotle's statement in the ***Metaphysics***, 982b30 (also 983a10) "In many ways human nature is slavish,": "Human nature is said to be servile, insofar as it is subject to many necessities. From which it happens that sometimes one sets aside what is worth seeking for its own sake on account of those things which are the necessities of life; as is said in the third book of the ***Topics***, that it is better to philosophize than to make money, granted that to make money is sometimes more to be chosen, e.g., when one is in need of necessities. From which it is manifest that that wisdom alone is sought for its own sake which does not belong to man as a possession. For that which man has as a possession he is able to have at will and to use freely. This science, which is alone sought for its own sake, man cannot use freely, since frequently he is impeded from doing so on account of the necessities of life. Nor even is it subject to the will of man, since a man is not able to arrive at it perfectly. Nevertheless, the little which is had of it outweighs everything else which is known through the other sciences." ***In Duodecim Libros Metaphysicorum Aristotelis***, (Rome: Marietti, 1950), #60. Cf. also #64: "God alone possesses [wisdom], or if not God alone, he chiefly possesses it. For he alone has it according to perfect comprehension. He has it chiefly, insofar as in his mode even it is possessed by men, granted it is not had as a possession, but as something borrowed from him."

the same thing: philosophy is "being like to God **as far as this is possible for man**."[21] The great difference between divine understanding and human understanding is referred to in a number of other fragments as well:

> A man hears himself called silly by a divinity as a child does by a man. (DK 79, trans. Robinson)

A second fragment is even stronger:

> In the matter of wisdom, beauty, and every other thing, in contrast with God the wisest of mankind will appear an ape. (DK 83*, trans. Robinson)

Heraclitus insists a great deal upon the limitation of human powers, even to the point that he says:

> The one who appears to be the wisest knows, preserves, only what seems. (DK 28a, trans. Robinson)

However, other fragments indicate that he does not entirely despair of attaining some degree of wisdom:

> One must expect the unexpected or [one] will not discover it; for it is difficult to discover and intractable. (DK 18)

Wisdom is difficult, but not impossible to discover. The same also seems to be implied in DK 93:

> The lord whose oracle is in Delphi neither indicates clearly nor conceals, but gives a sign. (DK 93)

---

[21] Plato, ***Theaetetus***, 176b1.

For Heraclitus the hidden harmony is **known** to be better than the apparent harmony. Thus, knowledge of the hidden harmony can be arrived at.

Heraclitus, then, insists on the fact that the answers to philosophical questions are rarely easy, without going so far as to deny that there are answers, or at least partial answers. This immediately suggest certain moral attitudes: the philosopher humbly accepts to have to dig a lot, to shun apparent short cuts, and to avoid idle speculation (as he says in DK 47: "Let us not conjecture at random about the greatest of things"). Love of wisdom makes the toil bearable, for the philosopher esteems the little knowledge he digs up as having great worth.

Conclusion

Wisdom according to Heraclitus is not the accumulation of opinions or facts. In the full sense it is nothing other than the mind of God which is the source of the order of the universe. In a derivative sense it is human knowledge which attains to this order. The endeavor of coming to know the mind of God is one of great difficulty, indeed is virtually beyond human power. To the extent that humans can attain wisdom, this can only be accomplished by starting from the world as commonly experienced, and by delving into many things. Having delved into many things, the seeker of wisdom proceeds to formulate paradoxes: On the one hand, evidence points to things being this way, while on the other, it seems to indicate that they are the exact opposite.

The attainment of wisdom demands in addition the ability to get beyond the pure multiplicity of things to an understanding of their common nature. The inquiry into many things of itself does not yield wisdom: discernment and insight must also be present. This insight is necessarily attached to the proper starting point of reflection, the world as all experience it: "Wisdom is saying what is true and acting in accordance with the nature of things by assenting to it." (DK 112)

The difficulty of wisdom commands reverence and humility on the part of the one seeking it. Heraclitus' statement: "There is greater need to extinguish ***hybris*** than there is a blazing fire" (DK 43, trans. Robinson) is especially applicable to intellectual pursuits. Another consequence of the difficulty of wisdom, is that the person in the quest thereof requires assistance, and this is to be drawn from the rare individual who has had some success in this difficult undertaking; here too discernment must be exercised lest one mistakenly choose a fool for a master.

These thoughts on wisdom reveal Heraclitus' true colors as being those of a humble seeker of truth: For while modest in his expectations of attaining wisdom, he nonetheless maintains that if the efforts one makes are judicious and sustained, they are more than amply rewarded.

St. John's University
Jamaica, New York

Truth, Adequacy and Being in Spinoza's *Ethics*

Lance Byron Richey

Despite the enormous amount of ink that has been spilled in the last two generations concerning Spinoza's distinction between the truth and the adequacy of an idea, the scholarship on the subject is still, to say the least, inconclusive. Although the distinction between the truth and the adequacy of an idea is central to Spinoza's epistemology, the little general agreement that is found among scholars tends to be insufficient for most readers to gain a "clear and distinct" understanding of what distinguishes true ideas from adequate ones. The purpose of this paper is to make more clear Spinoza's distinction. By an examination of the central texts from Spinoza which deal with the truth and the adequacy of ideas, as illuminated by some of the more important scholarship which has been done in this area during this century, I hope to cast some light on this important topic.

My goals in this paper are relatively simple and limited. In Section I, I provide a moderately detailed summary of Spinoza's views on true ideas and their connection to a correspondence theory of truth which runs throughout the ***Ethics***. In Section II, a similar review of Spinoza's theory of adequate ideas, and its connection with a coherence theory of truth which is also found in the ***Ethics***, is attempted. In Section III, I briefly explore the relationship between the truth and the adequacy of an idea, that is, Spinoza's epistemology and his ontology. I argue that, for Spinoza, the coherence and correspondence theories of truth which run parallel throughout the ***Ethics*** are not competing truth theories, but are instead complementary. More importantly, I argue that these concepts, indeed Spinoza's epistemology itself, are really the vestibule of his ontology, and that true and adequate ideas are finally subsumed into Substance, or Being, as mere characteristics of it. Finally, a brief summary and review of this article will conclude the discussion.

Truth, Adequacy and Being in Spinoza's Ethics

I
Spinoza on Truth

Very early on in the ***Ethics***, Spinoza offers a theory of truth, writing that, "A true idea must agree with that of which it is the idea" (E1Ax6)[1] This concise formulation seems traditional and clear enough. Unfortunately for students of Spinoza, it is neither. Two problems arise from it. The first is Spinoza's somewhat free use of the word 'truth' throughout his writings, often allowing it to cover both true and adequate ideas. The second is Spinoza's untraditional interpretation of agreement in light of his rejection of substantial forms. For now, I will attempt only to ascertain what Spinoza meant by a "true idea" in his mature thought, and allow earlier positions to go by the way.

In the ***Ethics***, Spinoza chooses to present this description of 'true' axiomatically, not definitively. Some interpreters have taken this as evidence that, for him, axioms are functional descriptions, while definitions are exhaustive regarding the essence of what they describe.[2] By presenting it axiomatically, Spinoza makes his meaning of 'true' rely not on its self-evidence, but on its place in his system. He offers no justification for this usage of 'true', nor does he intend for this axiom to be taken as a definition of what it means for an idea to be "true." Garrett rejects Wolfson's thesis that Spinoza uses definitions, axioms and propositions interchangeably within his system. Instead, he argues that Spinoza puts forward this description of a true idea axiomatically instead of definitively because, while it is not incorrect, "external agreement or correspondence does not fully capture the essence of truth."[3] What Spinoza claims does capture the essence of truth will be taken up later. In any

---

[1] "Idea vera debet cum suo ideato convenire." All latin quotes are from Benedict Spinoza, ***Opera Omnia***, (C. Gebhart ed.; Heidelberg: Carl Winters Verlag, 1925).

[2] Don Garrett, "Truth, Method and Correspondence in Spinoza and Leibniz," ***Studia Spinozana*** 6 (1990), 13-44, p. 18.

[3] Ibid., p. 18.

case, Spinoza does use correspondence with its object as a necessary characteristic of true ideas throughout the ***Ethics***. He later writes that

> All ideas are true in so far as they are in God. Demonstration: All ideas, which are in God, agree completely with the objects of which they are ideas, and so they are all true. (E2P32)[4]

Furthermore, he says that one has a true idea when one has "an idea which corresponds to that of which it is the idea." (E2P43Schol)[5]

It is not only in the ***Ethics*** that Spinoza puts forward correspondence with its object as a necessary characteristic of a true idea. In the ***Tractatus de Intellectus Emendatione***, he says that a true idea is distinguished from a false one by, though not solely or chiefly by, "an extrinsic denomination," that is, a correspondence with the external object of which the idea is an idea (TdIE II,26,15-20).[6] And as late as 1675, Spinoza writes in a letter to Tschirnaus that "the word true refers to the agreement of an idea with its ideatum." (Ep.60)[7] It seems clear from his writings that Spinoza intended that 'true' refer, at least in most contexts, to an extrinsic quality of an idea, namely, the agreement of an idea with its object. At times, he used the word more freely, and sloppily, but if pushed to the wall and asked what his meaning was in using the word 'true' in the ***Ethics***, I believe Spinoza would make it correspondence. Were this not the case, Spinoza would not have had need of the concept and elaborate terminology for 'adequacy' which occupies much of the ***Ethics***.

There is one important point for students of Spinoza's epistemology: Spinoza seems to be running up against the limitations of his language in this matter, since truth has more than one specific philosophic sense. It can be taken

---

[4] "Omnes ideae, quatenus ad Deum referuntur, verae sunt. Demonstratio: Omnes enim ideae, quae in Deo sunt, sum suis (Objectis et) ideatis omnino conveniunt, adeoque verae sunt."

[5] "idea vera, quatenus tantum dicitur cum suo ideato convenire."

[6] "denominationem extrinsicam."

[7] "inter ideam veram et adequatum nullam aliam differentia—jam agnosco, quam quod nomen veri respiciat tantummodo convenientiam ideae cum suo ideato."

strictly as correspondence, as the medievals took it. But truth in its vulgar usage carries a sense of reality or correctness which need not be limited to any one theory of truth (see any dictionary), as is clear from the many differing philosophical theories of truth (i.e., coherence, correspondence, performative, etc.). This is the cause of Spinoza's occasionally free use of the word. While Spinoza would certainly want to claim that his definition of adequacy is true in the sense of not being false, I am not so sure he would want the truth of that proposition to rest solely on a correspondence model. Therefore, the reader is advised to keep in mind that 'true' taken by itself will be used to refer solely to some form of correspondence, while 'true in the broader sense' or similar phrases will be used to denote the popular meaning of 'true' as 'not false'.[8] While this rather vague usage may seem inappropriate, it allows us to avoid tedious digression and is justified, I believe. As long as the reader keeps these distinctions in mind, no undue ambiguity should arise.

It has been shown that Spinoza has a correspondence theory of truth circling the waters of his philosophy. What of it? We have already seen that Spinoza, while holding correspondence to be a necessary characteristic of truth, probably does not hold it to be the defining characteristic. Nevertheless, nearly every commentator has seen the correspondence theory as a major component of Spinoza's thought. Although most have found some difficulties in consistently applying it[9], I believe these difficulties point more to its insufficiency, which Spinoza was quite aware of, than to its falsity. Perhaps the most important exception to this general consensus is found in Joachim:

> It would seem, therefore, that it is strictly impossible for Spinoza to talk of an 'agreement' between idea and ideatum. For, from one point of view, they are so completely one, that

---

[8] For an extended discussion of this topic, see Thomas C. Mark, "Truth and Adequacy of Spinozistic Ideas," ***Southwestern Journal of Philosophy*** 8 (1977), 57-70.

[9] Edwin Curley, ***The Collected Works of Spinoza***, (Princeton: Princeton University Press, 1985), p. 31 (footnote).

> no relation between them is possible—their unity is in no sense a relational unity. And, from another point of view, they are so absolutely two, that they cannot have any community of being whatsoever.[10]

Wolfson's medieval casting of Spinoza some thirty-odd years after Joachim, while outdated in some ways, is still correct in seeing in Spinoza a correspondence theory of truth as well as a coherence theory, and is correct in his subordination of the former to the latter in Spinoza's thought.[11]

G. H. R. Parkinson, in his ***Spinoza's Theory of Knowledge***, follows Wolfson in pointing out the medieval origins of Spinoza's criterion for truth, suggesting that Spinoza probably culled it from the (presumably Aristotelian) logic manuals of the day which he read and owned.[12] Parkinson's valuable study of Spinoza's correspondence theory also provides a logical bridge to the next section of my article. He argues that Spinoza's rejection of Aristotelian forms, which is the second problem mentioned above for the students of Spinoza's epistemology, also entails a reworking of scholastic correspondence theory. He concludes that the Spinozistic correspondence of idea and ideatum is actually a relation of identity, since "this Corollary (to E2P7) says that whatever is in the attribute of extension is present, in the same order and with the same connexions, in the attribute of thought."[13]

This revision also reveals the progress made in Spinoza scholarship during the last few decades. It is an important advance here that, while Joachim saw this relation of identity as precluding any correspondence, Parkinson views

---

[10] Harold Joachim, ***A Study in the Ethics of Spinoza***, (Oxford: Clarendon Press, 1901), p. 148.

[11] H.A. Wolfson, ***The Philosophy of Spinoza***, 2 vols. (Cambridge, Massachusetts: Harvard University Press, 1934), pp. 98-101.

[12] G.H.R. Parkinson, ***Spinoza's Theory of Knowledge***, (Oxford: Clarendon Press, 1954), p. 112.

[13] Ibid., p. 113.

it as the basis for such a theory of truth, and builds a sizable and intellectually fruitful analysis upon it.

And it is here that we begin to come nearer to Spinoza's main beliefs about truth. We now have stepped away from a focus on correspondence, that is, on a one-to-one mapping between the attributes of thought and extension, and towards the intra-attribute connections of ideas. Spinoza does not deny that this extrinsic correspondence is a necessary characteristic for the truth of an idea (in the broad sense), but I believe he would deny it sufficiency for making an idea true (in the broad sense). This now leads us to consider whether truth abides, not in any trans-attribute relationship, but instead in the intra-attribute relations of any given idea. This is what Spinoza refers to as an idea's "adequacy."

II
SPINOZA ON ADEQUACY

We are now starting to close in on Spinoza's main tenets concerning the essential characteristics of truth. Garrett notes that Spinoza gives a definition of an adequate idea, instead of presenting it axiomatically, as he did with true ideas.[14] Spinoza writes,

> By an adequate idea I mean an idea which, insofar as it is considered in itself without relation to its object, has all the properties—that is, intrinsic characteristics—of a true idea. Explication: I say 'intrinsic' so as to exclude the extrinsic characteristic—to wit, the agreement of the idea that of which it is an idea. (E2def4)[15]

---

[14] Garrett, p. 18.

[15] "Per ideam adaequatam intelligo ideam, quae, quatenus in se sine relatione ad objectum consideratur, omnes verae idea proprietates, sive demonimationes intrinsecas habet.

Here, one could ask, if the necessary characteristic of a true idea is its extrinsic agreement with its ideatum, and Spinoza is excluding this from the properties true ideas share with adequate ideas, what else is left to be shared? The obvious solution lies in seeing that extrinsic agreement with its object is not the only characteristic of a true idea. A true idea also has its relationships with other ideas as well as with its objects. The standard view among scholars concerning the nature of these relations is that ideas have logical relations, and objects have causal ones, a view which I see no need to challenge here. And it is here, in the intra-attribute relationships of ideas, that Spinoza believes adequacy is to be found.

Spinoza focuses more on adequacy than on truth (in the narrow sense) in the ***Ethics***, which hints at the primacy over truth that he gave to adequacy. In the propositions devoted to the knowledge the human mind has of its body and the affects of it, he gives almost exclusive emphasis to the adequacy of the idea in the mind, and very little to its truth. Nevertheless, after his definition of adequacy, which Spinoza obviously considers "adequate" for the reader, he does not trouble to detail it further, instead going on to apply it to specific instances of ideas. The next place to look for clarification, then, is the letter to Tschirnaus quoted partially above. The full passage reads:

> I recognize no other difference between a true and an adequate idea than that the word true refers to the agreement of an idea with its ideatum, while the word adequate refers to the nature of the idea itself.[16]

Explicatio. Dico intrinsecas, ut illam secludam, quae extrinseca est, nempe convenientiam ideae cum suo ideato."

16 "Inter ideam veram et adaequatam nullam aliam differentiam agnosco, quam quod nomen veri respiciat tantummodo convenientiam ideae cum suo ideato; nomen adaequati autem naturam idea in se ipsa."

These are the two most explicit treatments of adequacy I can find in Spinoza's writings, and they are certainly the main ones upon which most of the important scholarship done recently has focused.

Now we must explore this idea of adequacy which Spinoza has succinctly (and perhaps too succinctly) laid out for us. Doing so will naturally lead us into a discussion of the coherence theory of truth which he apparently intended by his definition. We should notice that Spinoza is not making extrinsic correspondence a necessary but insufficient ground for an idea to be adequate. There is a radical distinction between the two terms, for

> falsity consists in the privation of knowledge which inadequate ideas, that is, fragmentary and confused ideas, involve. Proof: There is nothing positive in ideas which constitute the form of falsity. But falsity cannot consist in absolute privation (for minds, not bodies, are said to err and be deceived), nor again in absolute ignorance, for to be ignorant and to err are different. Therefore it consists in that privation of knowledge which inadequate knowledge, that is, inadequate and confused ideas, involves. (E2P35)[17]

This passage brings out two important points. First, falsity is not limited to one particular aspect (i.e., the adequacy) of an idea. One can say both of true and of adequate ideas that they are "not false," and therefore that falsity is not to be seen as only opposed to the truth, but not to adequacy.[18] This point becomes more prominent in the next section. Secondly, the adequacy of an idea is not reliant upon any external relationship to an object, but solely upon the

---

[17] "Falsitas consistit in cognitionis privatione, quam ideae inaedequatae, sive mutilatae, et confusae involvunt. Demonstratio. Nihil in ideis positivum datur, quod falsitas in absoluta privatione consistere nequit (Mentes enim, no corpora errare, nec falli dicuntur), neque etiam in absoluta ignorantia; diversa enim sunt, ignorare, et errare; quare in cognitionis privatione, quam rerum inadaequata cognitio, sive ideae inadaequata, et confusae involvunt, consistit."

[18] Mark, pp. 24-25.

relationship of the idea to other ideas. This gives us a model of the classic coherence theory of truth which, not surprisingly, finds its greatest acceptance among rationalist and idealist thinkers (with Hegel being **perhaps** the greatest exponent of it).

There is a clear tension in both Spinoza and his commentators over the level of priority to be given this coherence theory of truth over the correspondence theory in his system. Of the recent writers, perhaps Garrett feels most uncomfortable with radically devaluing the importance of correspondence in favor of coherence, that is, intrinsic denomination of truth. But even he recognizes the ultimate superiority of coherence and believes that Spinoza "assumes . . . that there are internal characteristics (which grant truth to an idea) possessed by all and only those ideas that correspond to their objects."[19] And this internal characteristic is nothing else than the completeness of all necessary connections of the idea with other ideas, that is, the adequacy of the idea.

What makes an idea inadequate, that is, false, is not its lack of correspondence with that of which it is the object, at least not primarily. Rather, it is the idea's lack of proper connections with other ideas which are logically presupposed for the completeness of this idea. (E2P35) As Parkinson writes,

> A coherence theory of truth must not only say that any incomplete idea is false, but also that any false idea is incomplete. And this is what Spinoza seems to maintain when he discusses the nature of falsity or error.[20]

This incompleteness is not a trans-attribute incompleteness, that is, a non-correspondence, but is instead an intra-attribute incompleteness. However, correspondence sneaks in again through the back door. Spinoza writes that "the order and connection of ideas is the same as the order and connection of things."

---

[19] Garrett, p. 32.

[20] Parkinson, p. 120.

(E2P7)[21] If this is so, then any adequate idea which contains all the ideas necessary for its completeness will necessarily correspond to its object due to the idea's adequacy. For the object exists as it is due to its complete causal history, and the idea, if adequate, will contain the logical connections which parallel the causal connections of the object.

We now begin to see the interdependence of correspondence and coherence theories of truth in Spinoza. The sublime beauty of his system is not the dogmatic exclusion of one or the other as a criterion for truth, but rather their interweaving to form a new fabric of truth which transcends both. Section III of this paper will be a brief look at the way in which Spinoza integrates the two.

III
KNOWING AND BEING

We have seen how Spinoza subordinates correspondence to coherence[22], and how he makes the two radically conceptually distinct, neither presupposing the other. At the same time, they are more than simply parallel and coextensive. Their relationship is eventually more intimate than simple co-extensivity and parallelism. The truth and the adequacy of an idea eventually dovetail, and although neither conceptually presupposes the other for its self-sufficiency, their parallelism is necessary, not contingent, and can be demonstrated as such.

Correspondence is subordinate to coherence as a standard of truth (in the broader sense), since Spinoza makes it clear that knowledge of correspondence is not required for certainty of the truth of an idea. He writes,

> if a true idea is distinguished from a false one only inasmuch as it is said to correspond with that of which it is an idea, then

---

[21] "Ordo, et connexio idearum idem est, ac ordo, et connexio rerum."

[22] Parkinson, p. 114.

> a true idea has no more reality or perfection than a false one (since they are distinguished only by an extrinsic characteristic). (E2P43Schol)[23]

Clearly, for Spinoza the intrinsic characteristics of an idea are what characterize its truth for us. But this use of adequacy as a signpost for truth in the broader sense should not be confused with a necessary and sufficient condition for truth. When we have an adequate idea, we know that we have an adequate, that is, a true idea, not because of its adequacy, but simply because we have it. Spinoza writes,

> As to the last question, how can a man know that he has an idea which corresponds to that of which it is an idea, I have just shown, with abundant clarity, that this arises from the fact that he does have an idea that corresponds to that of which it is an idea; that is, truth is its own standard. (E2P43Schol)[24]

Truth, it would seem, does not spring then from correspondence or from coherence, but rather from the self-evidence of an idea. But even more importantly, our knowledge of the truth of an idea, the certainty we feel of the truth of an idea, does not spring from the correspondence we see between our idea and its object, or the idea's coherence, but rather from the self evident truthfulness of the idea we have. As Mark puts it so nicely, "adequacy is not the criterion of truth, truth is the criterion of truth."[25]

---

[23] "si idea vera, quatenus tantum dicitur cum suo ideato convenire, a falsa distinguitur, nihil ergo realitatis, aut perfectionis idea vera habet prae falsa (quandoquidem per solam denominationem extrinsecam interna denominatio)."

[24] "Quod denique ultimum attinet, nempe, undenam homo scire potest se habere ideam, quae cum suo ideato conveniat, id modo satis superque ostendi ex hoc solo oriri, quod ideam habet, quae cum suo ideato convenit, sive quod veritas sui sit norma."

[25] Mark, p. 26. I am not completely convinced, as some are, that Mark is at variance with Curley and others on this point, although I would not assert their agreement with certainty.

This point should be examined more closely, as it lies at the very heart of my topic. We have already seen clearly that adequacy is superior to correspondence in determining an idea's truth in the broad sense. But now we have been told that not even adequacy is sufficient, as it is truth, and not adequacy, that is the criterion of truth. There are difficulties at this point which must be resolved, and I believe Mark's article contains the seeds of the best solution of this question.

Where is Spinoza headed with all his apparatus of truth, adequacy, and so forth? I want to argue that he is trying to move beyond mere epistemology and into an ontology of truth. Spinoza wants to overcome Descartes by moving beyond the question, "What am I sure of?", and into the question, "What is real and true?".[26] And when this question is asked, one is always asking, not about knowledge, but about being. For Spinoza, truth and adequacy primarily perform an indexical role as necessary characteristics of true ideas of which we have knowledge, but that knowledge must always be of particular entities as expressed under the attribute of thought at one time, and extension at another. Mark puts it thus:

> To be sure, what one recognizes in a true idea is an instantiation of being or substance, which is to say that one recognizes the self-completeness that we have linked with adequacy. Nevertheless, the recognition that our idea is true comes about not because adequacy is an indicator of truth, but because reality is self-complete and self-explanatory.[27]

As Mark says, not even adequacy is the essence of truth (that is, the characteristic which is a necessary and sufficient condition for believing an idea

---

[26] This question might also be phrased as, "I know that I have a clear and distinct idea, but what I really want to know is whether or not it is true." The obvious slap in Descartes' face would surely have been made more subtle by Spinoza, though.

[27] Mark, p. 26.

to be true), but instead truth is its own sign, needing no epistemic justification for its truth other than the fact of its truth. Thus, we have moved beyond the sphere of epistemology, that is, the study of knowledge, and into the study of being, of what is.

At last, we have seen the true importance and unity of adequacy and truth, not primarily in Spinoza's epistemology, but in his ontology. Truth and adequacy run parallel and coextensively throughout Spinoza's epistemology, because there one finds a distinction between knower and known, between idea and ideatum. But at the ontological level, the level of substance for Spinoza, such distinctions do not occur, as was demonstrated by him in Part I of the ***Ethics***.

As Mark says, reality is self-complete and self-explanatory. Therefore, all true knowledge must be of reality, and if Spinoza was attempting anything in his system, it was to give a comprehensive schema of reality which could accommodate both thought and extension, as regards both their existence and their explanatory histories. This is where correspondence and coherence theories of truth complement each other in Spinoza's thought. The correspondence theory maps individuals of all attributes onto thought, including the attributes of thought itself, and thus provides for a coextensivity of thought and being, which has always been the goal of western philosophy.[28] The coherence theory of truth provides for a complete inner consistency and integration of each attribute. Usually coherence is thought of as relating to the attribute of thought, but only because the physical world makes much more manifest its complete inter-relatedness and self-containedness. Nevertheless, for every event in the physical world, along with all the levels of relationship that event holds with other physical events, there is a corresponding idea, which exactly maps all the physical inter-relationships onto mental relationships. Thus, Spinoza has accounted for the completeness of each attribute, as well as for their perfect parallelism, their perfect mapping of one onto the other.

---

[28] The coextensivity of the attribute of thought with all other attributes, a necessary feature of this system, is not itself without problems, and the reader should keep this in mind.

Truth, in the vulgar sense, in the broader sense which we have named without naming so often in this paper, is for Spinoza the totality of correspondence and coherence, as it contains all the intra-attribute and trans-attribute relationships there are. In English, we have a much shorter way of expressing this: reality. Indeed,

> If someone were to ask, "How can one be certain that the ideas one has are not merely true but also adequate?", Spinoza would answer: if one is asking for a criterion to determine the adequacy of ideas, which is separate from the idea in question, then there can be no other criteria to establish it except these ideas themselves. They are adequate and true for the simple reason that, being ideas of thought objects correlated with the actual nature of things, they could not (logically) be false or inadequate...By 'true idea' here, Spinoza must mean an idea which is at once adequate and true.[29]

Note well here that truth has been made a necessary but insufficient condition for a true idea in the broader sense of the word. It is now clear that for Spinoza, truth in the common sense lies not in the status of our knowledge as true or adequate, but in the status of being ***qua*** being. Truth is nothing else than being, or substance, or God. Thus he writes, "All ideas are true insofar at they are related to God." (E2P32)[30] Truth in the strongest sense comes, not from any relationship across or within attributes, but in the ground of those attributes, the concept of substance or God. Truth finds its being in God, for Spinoza. The truth and adequacy of an idea come from its being true (that is, real), and not the reverse, as some have argued.

---

[29] S. Paul Kashap, "Spinoza's Use of 'Idea'," ***Southwestern Journal of Philosophy*** 8 (1977), 57-70, p. 69.

[30] "Omnes ideae, quatenus ad Deum referuntur, verae sunt."

As a closing support for this view, I will offer the following syllogism, which the reader can take or leave; my argument does not hinge upon its validity, though it may be useful for illustration. For Spinoza, false ideas contain nothing positive which causes their falsity. (E2P33) Nothing can be or be conceived outside of God. (E2P15) Therefore, the cause of all falsity in non-being, or non-existence. All truth is therefore in God, or Substance. (E2P32) And Spinoza says that "thinking substance and extended substance are one and the same substance, comprehended now under this attribute, now under that." (E2P7Schol)[31] Therefore, truth lies primarily in being, or substance, and only secondarily in the attribute of thought or in any correspondence between two attributes. In understanding this, hopefully the reader more fully understands the distinction and relation between a true idea and an adequate idea, as well as their subsumption in truth ***per se***.

Conclusion

What conclusions should be drawn from this discussion? Spinoza has laid out a criterion for truth which foreshadows much of the talk about "being" that has gone on since Hegel. A true idea, for Spinoza, is an idea which corresponds to the causal history of its object, although reference to this causal history is not required for recognition of the truth of this adequate idea. But for Spinoza, there is a higher level of truth than simple correspondence, a level I will call "truth ***per se***." Truth ***per se*** is the substance underlying every attribute, and containing in it the entity which finds parallel expression throughout all attributes, and complete expression throughout each. Of course, although strictly speaking there is only one entity, God, particular "moments" of God (i.e., individual objects or ideas) can still be true ***per se*** granted they contain adequate notions of the preceding moments, which in the attribute of extension are causes, and in the attribute of thought are ideas. This results in the belief

[31] ". . . consequenter quod substantia cogitans, et substantia extensa una, eademque est substantia, quae jam sub hoc, jam sub illo attributo comprehenditur."

among so many commentators that there is really only one true, adequate idea: the idea of God. While plausible, this belief need not be true for the thesis to stand that truth ***per se*** is the subsumption of truth and adequacy into substance. As such, the primacy of ontology over epistemology in Spinoza's philosophy is made clear.[32]

Marquette University
Milwaukee, WI

[32] I wish to express special thanks to Lee Rice for his generous help at every stage of this paper.

Turing Machines and Semantic Symbol Processing
Why Real Computers Don't Mind Chinese Emperors

Richard Yee

Debate over the computer metaphor of mind appears endless. At issue is the prospect of a computer ever having what could properly be considered a mind. Many proponents of the view that computers could have minds are persuaded in large part by the power of information processing. The essence of mind seems to lie in the processing of various forms of information including goals, perceptions, plans, and actions. Moreover, the **theory of computation** provides powerful support for the view that any sufficiently complete scientific theory of the mind, e.g., a theory relating neurological processes to psychological ones, would be **computable**. This essentially reflects the view that the **Church-Turing thesis** covers mind-brain processes. If human minds were understood in sufficient detail, then it would certainly be possible, in principle at least, to build computers that have real minds.

In contrast, many critics view computers as mere mindless automatons. In the rote execution of a program, how could a computer ever come in contact with any intrinsic meaning in its actions? How could a computer's symbols ever represent anything **to the computer**? Because the ability to process symbols semantically is a key mental trait, the **formal symbol processing** that a computer performs could never be sufficient to endow it with a mind.

The resulting debate between proponents and critics often revolves around the possibility of a computer's having any of a number of key mental qualities including consciousness, understanding, semantics, intentionality, qualia, creativity, and insight. Most such phenomena currently have no completely satisfying or agreed-upon characterizations (cf. Sloman, 1985). However this lack does not inhibit the formulation of arguments and refutations that reflect personal intuitions rooted in diverse backgrounds and biases. It is therefore inevitable that divergent and strongly held views should arise over many central questions, and equally inevitable that recurring attempts should be

made to convince the other side of the compelling force of one's own intuitions. Few are ever swayed, of course, and the debate rages on.

Disagreement over poorly defined mental phenomena is one thing, but too often what divides proponents and critics are their differing intuitions about computers, programs, and computation. For example, consider a computer running a program, which produces some computation of interest. In analyzing this phenomenon, do the key philosophical questions center on the program or on the computer? Which entity has primary responsibility for the computation? If it should happen that proponents focus on programs while critics focus on computers, then the ensuing debate might be vigorous indeed, being fueled by arguments that are largely at cross-purposes. Such a situation, however, should not be tolerated for long because, unlike mental phenomena, computers, programs, and computation should have reasonably precise definitions.

Unfortunately, in most cases the pivotal concept: **computer** is **not** precise. The common meaning of the term denotes programmable machines such as personal computers, workstations, and mainframes. Such machines are physical instances of **universal Turing machines** (UTM's). Often, however, the term is used to indicate any **Turing machine** (TM), not just universal ones. Although in many contexts the distinction between UTM's and non-universal TM's is not stressed, in the debate over the "computer metaphor" of mind, the distinction is crucial.

Countless critiques of the idea that mind is computable stem from the view that "computers" are only formal symbol processors. While such a view might hold for UTM's, it does not hold for all TM's. Shifting the debate to a more rigorous basis—i.e., focusing upon Turing machines proper rather than upon "computers" or "formal systems"—has devastating consequences for two of the most widely known critiques of the computer metaphor: Searle's **Chinese room** argument (Searle, 1980) and the family of arguments based upon Gödel's **Incompleteness theorems** (e.g., Nagel & Newman, 1958; Lucas, 1961; Rucker, 1982; Penrose 1989; Tymoczko, 1990). The Chinese room argument only attacks UTM's, which constitute an exceptional **subclass** of all TM's. The

"Gödelian" arguments attack **formal systems**, which correspond to **static** TM's, again, a peculiar restriction of the full TM model.

One thus finds that the failure to address Turing machines directly has bred volumes of unproductive debate. Both critics and proponents have engaged in disputes over differing "machine intuitions." Critics have formulated attacks upon TM-surrogates: UTM's and formal systems, while proponents have often responded by defending TM's in essence, but they have done so only implicitly—leaving the critics with their surrogate machine intuitions intact. A more profitable course in examining issues of mind and computation would be for both sides to address explicitly only full-fledged Turing machines, the **real** computers.

I
Minds and Turing Machines

To analyze arguments in the debate, it is first necessary to define clearly the questions at issue. Unfortunately, even this step entails some confusion. Probably the best-known position in support of the computer metaphor of mind is the one dubbed by Searle as **strong AI** (Searle, 1980). Searle describes this position as holding that a suitably programmed computer must have a mind—in the same sense in which humans have minds—if the computer exhibits the right inputs and outputs (Searle, 1980, 1987, 1990, 1992). This characterization is problematic, but the fault is not necessarily Searle's, at least not entirely (see Searle, 1992, Chapter 9, p. 200, and Footnote 2). Often the AI community itself has been too imprecise in articulating its own claims.

One problem with this common view of strong AI is that it conflates two independent propositions. The core assertion of strong AI is that mind is computable, i.e., a mind could result from the actions of a physically instantiated Turing machine. A second proposition holds that the **Turing test** (Turing, 1950) is an adequate means for awarding the label **has-a-mind** to a computing machine. In other words, this second proposition holds that probing a system

exclusively through its "linguistic" input-output behavior can provide sufficient information to justify the attribution of true mental processes to the system (e.g., Dennett, 1990).

Clearly, these two propositions—corresponding to Turing's most celebrated contributions—are independent. A position on one does not entail a position on the other. In particular, it would be possible to argue against the adequacy of the Turing test while maintaining the view that mind is computable by Turing machines. Although many supporters of strong AI also support the Turing test, it would be somewhat of a peripheral issue if TM's could never have minds in the first place. Hence, it is important keep these two propositions separate.

This article is concerned solely with arguments impacting the first proposition, namely, that some type of Turing machine could truly possess a mind. This is will be stated as follows:

Mind is a computation producible by a Turing machine. (P1)

Proposition P1 expresses the essence of the computer metaphor of mind. If P1 is true, then there is a non-empty class of Turing machines that, when physically instantiated, would have (or could develop) minds in the same sense in which humans have (or develop) minds. This is exactly analogous to saying that **addition** could result from a real TM. Hence, P1's claim is that, like addition, mental phenomena are computations that some subclass of TM's could perform.

P1 does not entail any position, pro or con, regarding the Turing test. Even more importantly, it does not refer to computers, programs, formal systems, or universal Turing machines. P1 does not refer to **computers** because the term connotes two related but significantly different concepts: Turing machines and universal Turing machines. P1 does not refer to **programs** because programs are merely descriptions of TM's. It does not refer to **formal systems** because, unlike TM's, formal systems do not provide for receiving inputs from external sources. Finally, P1 does not single out **UTM's** because,

contrary to their name, “universal” machines constitute only a subclass of all TM's—those that are universally programmable.[1]

Given P1, attention may turn to possible arguments against it. The following sections examine two such arguments: Searle's Chinese room and arguments from Gödel's Incompleteness theorems. When viewed in light of P1 and the theory of computation, the flaws of each argument become apparent. Although the theory of computation exposes the flaws, it does not necessarily address the original intuitions of critics who would likely persist in their contention that computation is inevitably devoid of semantic understanding. Therefore, Section III presents a basic account of why TM's can process symbols **non-formally**. Section IV concludes with a view toward further investigation of semantic symbol processing in Turing machines.

II
Attacks on Turing machines?

Since there are two prominent arguments that purportedly refute strong AI, it is natural to examine how each pertains to P1. A straightforward analysis will show that the Chinese room (CR) argument (Searle, 1980) has no direct bearing on P1 because the CR fails to address the entire class of TM's. The second argument is actually a family of arguments centered around Gödel's Incompleteness theorems (e.g., Lucas, 1961; Penrose, 1989). Although somewhat more complicated to analyze, it will be seen that Gödelian arguments are similarly flawed in that they attack only static formal systems, which are an inadequate substitute for dynamic TM's. Both arguments thus fail due to erroneous identifications between TM's and “equivalent constructs” that turn out to be **non-equivalent** for the purposes of answering philosophical questions about minds and computation.

---

[1] Moreover, **strict** UTM's are uninteresting from the point of view of mental processing for a number of reasons. In particular, they are among those TM's that cannot re-program themselves, i.e., they cannot learn.

2.1: The Chinese Room Versus Programmed Computers: Searle claims that the Chinese room (CR) argument refutes strong AI. Does it refute P1? Briefly, the CR argument is as follows. Inside a room is a human **U** (Searle), a program **P**, and Chinese input symbols $\mathbf{c_{in}}$ which are composed into strings. Figure 1(a) depicts this situation. **U** can understand no Chinese and is only able to process the inputs $\mathbf{c_{in}}$ according to program **P**. We may assume that the

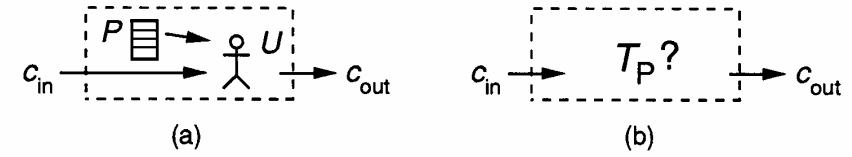


Figure 1: The Chinese room is a UTM. **U** does not understand Chinese, but what about $\mathbf{T_P}$?

CR can convincingly pass a Turing test, and then ask whether it follows that the Chinese symbols are being understood inside the room. At no time can **U** understand the meanings of the symbols $\mathbf{c_{in}}$. Furthermore, argues Searle, there is nothing else about the room which could understand them. The only other entity, program **P**, could be eliminated by having **U** "memorize" it. Doing so would still not enable **U** to understand the symbols $\mathbf{c_{in}}$. Thus, the meanings of the symbols are never understood inside the room.

One conclusion drawn from this argument is that the Turing test is inadequate as a test of understanding: the appearance of understanding need not imply its existence. As noted, however, this issue is not of current concern. A second conclusion is that no "programmed computer" could understand its input symbols because computers always do exactly what **U** does in the room: follow rules for syntactically manipulating symbols without connecting them to any semantic content. Because the semantic processing of symbols is central to mental functioning, this establishes that computers could not have minds.

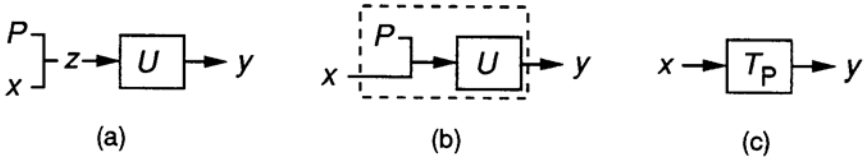


Figure 2: Two types of Turing machines. Part (a) shows a UTM **U** with input **z** composed of a program **P** and a nominal input **x**. Part (c) is an instance of the TM described by **P**. Part (b) is a view of **U** as the machine $\mathbf{T_P}$

TM's, UTM's and Programs: To understand this argument's relationship to P1, it might be useful to review some aspects of so-called "universal" computations. Figure 2 illustrates the relationships between UTM's, programs, and TM's. Fig. 2(a) shows that the total input to every UTM **U** consists of two parts: a program **P** and a nominal input **x**. **P** is a description of some Turing machine $\mathbf{T_P}$ (shown in Fig. 2(c)). **U** interprets program **P** as rules for processing the nominal input **x**. The output of this process, **y**, is the same as the output of $\mathbf{T_P}$ running directly on **x**. Figure 2(b) illustrates the typical operation of a programmed computer in which the program portion of **U'**s input is held constant while the nominal inputs are varied. In this manner **U** is able to simulate the overall input-output behavior of $\mathbf{T_P}$.

It is important to keep in mind the exact sense in which **U** is "universal." **U**, like every other TM, computes one specific function. In general, then, **U** and $\mathbf{T_P}$ are two different TM's that compute two different functions. **U** maps input $\mathbf{z} = \langle \mathbf{P}, \mathbf{x} \rangle$ to the output **y**, while $\mathbf{T_P}$ maps **x** to **y**. Only by restricting attention to **U**'s nominal input **x** (as in Fig. 2(b)) does one observe the reproduction of $\mathbf{T_P}$'s input-output behavior. Internally, **U** operates in a very specific fashion which centers on properly interpreting program **P**, a portion of input **z**. Nominal input **x** is processed only via the instantiation of $\mathbf{T_P}$, not by **U** directly. Hence, to generate $\mathbf{T_P}$'s output on input **x**, it is sufficient for **U** to process **x** purely formally. A UTM, therefore, is simply a TM that treats a portion of its input as a body of rules for formally manipulating the remaining portion of its input.

Consider, for example, the differences between a simple calculator that can perform only addition and a general purpose computer running a program that describes the calculator. There is no software-hardware distinction for the calculator. The calculator itself embodies an algorithm for performing addition. One could never remove the calculator's "program" without losing the whole machine. The computer, on the other hand, embodies a universal algorithm, which is nothing more than an algorithm for interpreting and executing rules. It is clearly not an algorithm for addition. To the computer, a calculator program is just an input, which could be completely replaced without affecting the computer itself. To find out how it would "actually feel" to add numbers, one should consult the calculator because the computer cannot add—it can only instantiate other TM's.

When a computer runs a program, there are at least two identifiable computations, the universal one and that of the TM described by the program, both of which are simultaneously implemented on a single hardware system. It is sometimes said that the program's TM is a "virtual machine," but the truth is that both computations have precisely the same ontological status. Each corresponds to a particular mathematical, or logical, description of the physical system. Thus, neither computation could be considered more or less "real" than the other.[2]

To investigate computational processes properly, therefore, it is advisable to attend only to the processes themselves, ignoring UTM-computers altogether. Unless strictly-universal computations happen to be the objects of interest, UTM's need never enter the picture. They are nothing more than middlemen whose own computations divert attention from the computations that they instantiate. UTM's are mathematical curiosities and engineering

---

[2] This is the formal justification for one-half of the **Systems Reply** to the Chinese room argument. That is, the theory of computation establishes that there are two computations in the CR. Hence, the **possibility exists** that the second computation, the one described by program **P**, is actually understanding the Chinese input symbols. To conclude the second half, that the Chinese is **definitely** being understood, one would either need to believe in the Turing test, or need to know whether the program's TM does what Chinese **brains** do (N.B., not minds).

conveniences, but when it comes to questions about minds and computation they are philosophical quicksand.

The Chinese Room Revisited: The Chinese room argument asserts that executing a program is not sufficient to produce a mind (e.g., Searle, 1992, p. 200). Suppose this conclusion were correct. Would P1 have been scathed? The programmed entity **U** in the CR is exactly a "computer," i.e., a UTM. Thus, at best, the CR only establishes that UTM's can never understand the meanings of their (nominal) input symbols. On the other hand, P1 asserts only that **some** TM's may have such mental qualities. Thus, even if one granted the argument's claims with respect to UTM-computers, there would remain plenty of non-universal TM's for which P1 might still hold.

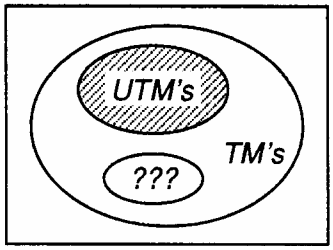


Figure 3: UTM's are a proper subset of all TM's. Even if mind were not a universal computation, it does not follow that it is not in some other Turing-computable class.

Figure 3 illustrates this point, which can be formalized as follows. Consider the following assertion:

Property P is true of every computer. (2)

This statement is ambiguous, and should be replaced by one of the following:

P is true of every Turing machine. (3a)
P is true of every universal Turing machine. (3b)

Statement (3a) is much stronger than (3b). Because UTM's are a proper subset of the set of all TM's, (3a) logically implies (3b), but not conversely. In general, it would be fallacious to form a conclusion like (3a) from arguments that only establish its (3b) equivalent. Although the CR argument only addresses UTM's running programs, it seems to have been widely interpreted as a demonstration that **no TM** could understand the meaning of its input symbols. Clearly, however, such a deduction would be erroneous. P1 is thus immune from any straightforward application of the Chinese room argument.

What does this analysis say about questions of **syntax-versus-semantics**? The preceding comments **do not** dispute Searle's most basic claim—shared by Harnad (Harnad, 1990)—which is that formal manipulation of symbols is significantly different from the semantic understanding of symbols required for minds. Analysis of the CR argument simply shows that it does not address the symbol processing abilities of all (indeed, of most) TM's because the CR examines the internal processing **only** of programmable UTM's. On this point, the universality of UTM's is immaterial because it does not extend below the input-output level, and the argument assumes that performance there looks like the true understanding of Chinese. To establish that all TM's are as numb to certain input symbols as **U**, one would need an argument that specifically addresses the internal processing of all TM's. However, Section III, below, presents arguments to the contrary, showing that TM's are capable of processing symbols in a non-formal fashion.

The Mind-Brain Analogy: Searle points out that the strong AI position is often characterized by drawing an analogy between minds and brains on the one hand, and programs and computers, or software and hardware, on the other, i.e., **mind:brain::program:computer**, or **mind:brain::software:hardware** (Searle, 1980, 1987, 1992). If this analogy is

indeed pervasive, it would not be surprising. Computers and programs are two of the most conspicuous icons of our time. The Chinese room argument is intended to refute this analogy, but such a refutation is unnecessary because the analogy itself is not apt.

Translated into more precise terms, the analogy reads that mind is to brain as the description of a TM is to a UTM, i.e., **mind:brain:: TM-description:UTM**. This borders on the nonsensical. Even if one could make sense of it, there is little point in trying. Turing machines, not simply UTM's, are the objects of interest. The proper analogy to consider, therefore, is the one derived from P1:

mind:brain::computation:Turing-machine.

Could a Turing machine have a mind in the same way that brains have minds?

2.2: Gödelian Truth Versus Formal Systems: The view that Turing machines are merely formal processors, blind to the meanings of their symbols, also underlies critiques based on Gödel's Incompleteness theorems (e.g., Nagel & Newman, 1958; Lucas, 1961; Rucker, 1982; Penrose, 1989; Tymoczko, 1990). All such arguments are based on Gödel's proof that **formal systems** have limitations, which in a certain manner, translate into limitations on Turing machines. Gödel proved that any formal system **F**, sufficient for Peano arithmetic, has a well-formed statement **G(F)** that is clearly seen to be true but is not derivable using only **F**'s axioms and transformation rules. This fact is used as the basis for arguing that certain human powers of thought cannot be reproduced by any formal system or any equivalent TM.

The essence of this argument is a “proof” that runs generally as follows. Suppose that human thought **H** were producible by some formal system **F**, i.e., **H = F**. Since **F** must be sufficient for Peano arithmetic, there must be a statement **G(F)** which is true but which **F** cannot prove via its axioms and derivation rules. Nevertheless, through “mathematical insight” humans **H**

can see that **G(F)** is true. Thus, **H** can prove something unprovable by **F**, showing that **H≠F**: human thought is not producible by any formal system.

The key to this argument is the idea that, thanks to Gödel, humans could always **outmaneuver** any formal system **F**. This is done by forming **F**'s Gödel statement **G(F)** and recognizing its truth. Note that knowing of the existence of some Gödel statement **G(F)** is not the same as knowing that a particular statement **w** is in fact **F**'s Gödel statement, i.e., that **w = G(F).** Gödelian arguments assert that humans are actually capable of establishing the truth of particular statements **G(F)** for systems **F** that are alleged to be equivalent to human thought.

Many point out that such an assumption seems unfounded: if human thought were in fact producible by a formal system, it would undoubtedly be enormously complex, and there is no principled reason to believe that humans could produce its Gödel statement (e.g., Hofstadter, 1979; Rucker, 1982). Others point to related problems regarding the **consistency** of formal systems, which is necessary for applying Gödel's results (e.g., Putnam, 1960; Bowie, 1982; Chalmers, 1990; Davis, 1990; Mortensen, 1990). Although such refutations might have merit, they have not won over many Gödelian critics (e.g., Lucas, 1961; Penrose, 1990; Tymoczko, 1990). Perhaps these refutations do not directly confront the central Gödelian intuition which is something akin to the sense that "I can **easily see** a truth that **provably cannot be derived** by any formal system." A more direct response to this intuition comes in two parts. The first points out that humans are not really so clever. The second points out that TM's could also outmaneuver formal systems, if they were given a fair chance.

Naked Formal Systems: What enables humans to produce Gödel statements for formal systems? Gödel has shown us the key: it is simply necessary to construct a particular self-referential statement about a given system. Surely, humans could reproduce Gödel's "trick" for any given formal system, especially if it were not extremely complex. Let us test this assertion on a very simple

formal system **A**: Peano arithmetic. Of course, it would be too easy to derive **G(A)** if **A** were specified using the usual alphabet of mathematical symbols such as '×' for multiplication or '∃' for 'there exists'. Instead, let us encode the symbols for **A** using arbitrarily chosen Chinese characters: for example, we could replace the symbol '∃' with the Chinese symbol for 'horse'. Call this Chinese encoding of Peano arithmetic $\mathbf{A_C}$.

Now imagine being given $\mathbf{A_C}$ with knowledge of neither the encoding nor even the fact that $\mathbf{A_C}$ represents Peano arithmetic. How could one formulate $\mathbf{G(A_C)}$ and be convinced of its truth? Worse still, suppose one were given an encoded form of a system, **F**, more complicated than $\mathbf{A_C}$? It is absurd to think that anyone could be certain of both forming and recognizing the truth of the Gödel statements for formal systems under such circumstances. Although some might feel that the intended interpretations of the symbols should be provided with the systems, such information is **external** to any formal system. It is information about the relationship between the given formal system and another system—ultimately the "system" of human experience.

The point of this example is simply that humans could only outmaneuver a formal system via Gödel's technique if they were provided with meanings for the symbols of the formal system. Only in such a case could anyone construct meaningful statements (let alone **true** ones) without relying on the system's axioms and transformation rules. This fact is so obvious, it is easily missed. It is easy to assume that **being given a formal system F** further entails being given an interpretation, **Interp(F)**, that assigns meanings to **F**'s symbols.

Semantic Clothing: Given a semantic interpretation for the symbols of a system **F**, it does indeed seem possible to derive a Gödel statement **G(F)**, at least in principle. But if humans require **Interp(F)** in order to form the Gödel statement, it would be unfair to withhold this information from **F**, the system humans claim to be outmaneuvering. **F** should also be given **Interp(F)**. But now this highlights a crucial difference between formal systems and TM's.

Formal systems **do not accept inputs** from external sources, whereas TM's clearly do.

To illustrate the crucial role of inputs, consider the following fallacious argument. I claim that my image can never be captured in any mere photograph. Suppose that, to prove me wrong, you were to take a picture of me. To convince me that you have indeed captured my true image you must give me the photograph. I will then point out that this photograph fails to reflect the obvious fact that I am holding a photograph. Clearly, I can outmaneuver **any** alleged photograph of me because no photograph can portray its own image. I therefore conclude that although photographs might reflect certain likenesses, they can never fully capture my true image. I can always identify a fatal discrepancy in any photograph **that is given to me** (cf. Lucas, 1961; Hofstadter, 1979).

The fact that a person could derive a Gödel statement for system **F**, given an interpretation of its symbols, **Interp(F)**, does not prove that **F** was not an accurate account of the person at the time before **Interp(F)** was provided. In other words, if person **H** were characterized by formal system **F** at time **t**, i.e. if **H = F**, then providing **H** with **Interp(F)** at time **t+1**, would thereby change **H** into $\mathbf{F^+ = F + Interp(F)}$. $\mathbf{F^+}$ has new axioms that may entail **G(F)**, so it would **not** contradict Gödel's theorems for $\mathbf{F^+}$ to derive **G(F)**. Furthermore, the derivation of **G(F)** would in no way imply that $\mathbf{H \neq F}$ at time **t**. There would only be a contradiction if **H** could prove **G(F) without being given Interp(F).**

Gödelian arguments now face the following dilemma. Either (a) humans must be capable of establishing the truth of Gödel statements for formal systems whose symbols are utterly meaningless (to the persons in question), or (b) it must be proved that no formal system $\mathbf{F^+}$ can derive **G(F)**, where $\mathbf{F^+ = F + Interp(F)}$ is a formal system **F** augmented with a suitable interpretation of its own symbols. The first case is impossible because the "self-evident truth" of any Gödel statement requires knowing the meanings of the symbols in question. Establishing the second case, which is not covered by Gödel's theorems, would require a new proof. However, there is good reason to doubt that such a proof exists because Gödel has **shown us how** to construct **G(F)** from a suitable

interpretation of system **F**'s symbols. It is far from obvious that Gödel's process of using **Interp(F)** could not be automated—at least to the same extent that humans are capable of performing it.

As Penrose points out (Penrose, 1989, p. 111-112), there is nothing particularly sacred about the truth of the Gödel statement **G(F)**. **F**'s axioms are also "obviously true" but not derivable within **F**. Hence, there is no more magic required in seeing the truth of **G(F)** than there is in seeing the truth of **F**'s axioms. To be validated, both the axioms and **G(F)** must be interpreted rather than formally derived. In both cases the interpretation process begins by associating meanings with **F**'s symbols. These meanings are essentially correspondences between the symbols and entities in some other system, e.g., a person's knowledge and beliefs about the world. The truths of the axioms and Gödel statement are then ascertained by determining "truth" in the other system.

Learning Machines: The fallacy of Gödelian arguments lies in equating dynamic Turing machines with "static snapshots," i.e., formal systems. Turing machines can interact with external sources of information and, through such interactions, change themselves. Formal systems, on the other hand, are inputless engines for generating a body of theorems. In short, Turing machines can **learn**, whereas formal systems cannot (cf. Arbib, 1987, Section 8.5).

Although most computer programs (Turing machines) that people encounter retain their functionality from one use to the next, making them equivalent to fixed formal systems, not all TM's need be so "faithful." Given inputs, TM's may literally re-program themselves, becoming different machines with new behaviors and abilities. In particular, it is entirely possible that a TM, given an interpretation of its own program, could produce the Gödel statement for that program. There would be no violation of Gödel's theorems because the input interpretation would have first **changed the machine**—just as it would for humans.

III
Formal Symbol Processing Versus Turing Machines

Are the flaws in the Chinese room and Gödelian arguments mere mathematical technicalities? Fundamentally, both arguments are motivated by the same intuition: automatic computation is an inherently meaningless activity, whereas the very essence of mind lies in subjective awareness and the certainty that thought processes are intrinsically meaningful. The CR and Gödelian arguments are simply vehicles that attempt to establish concretely this key difference between computational and mental symbol processing. An essential question, therefore, is whether every Turing machine necessarily processes information in a fashion that is inherently meaningless **to the machine**.

To some it may seem obvious that all TM's are simply formal symbol processors. If this is so, then it should be easy to prove. The first requirement is to define the difference between **formal** and **non-formal** types of symbol processing. The phrases **symbol processing** or **symbol manipulation** are sometimes used as if the processing of symbols were **necessarily** formal, but this is obviously not true. For example, in the CR if **U** were literate in Chinese, then the processing of the input symbol for 'horse', say, would not necessarily be formal. The reason is that, in interpreting the symbol, it would become possible to access **U**'s personal information about horses, information not contained in the symbol itself. If **U** believed that horses are bigger than bread-boxes, for example, this information could play a role in determining **U**'s output.

For the processing of input symbol α_i to be strictly formal, outputs may depend only on objective information intrinsic to α_i itself. This is usually considered information describing α_i's shape (literally its **form**) because that information allows the symbol to be distinguished from other potential symbols. The processing of α_i would be non-formal if (a) the processor were to associate the symbol with subjective information, not intrinsic to α_i's shape, and (b) such information were allowed to influence the course of subsequent processing and, ultimately, the outputs.

Turing Machines and their Transition Functions: To determine whether all Turing machines are restricted solely to formal symbol processing, one must consider the details of how a TM processes its input symbols. A Turing machine **T** is a mathematical object denoted

$$T = (S, \Gamma, \delta), \text{ where}$$

$S = \{ \mathbf{s_0}, \ldots, \mathbf{s_N} \}$ is a set of internal **control states**, which includes the **start state** $\mathbf{s_0}$ but not the **halt state H**

$\Gamma = \{ \boldsymbol{\alpha_1}, \ldots, \boldsymbol{\alpha_M}, \boldsymbol{\gamma_1}, \ldots, \boldsymbol{\gamma_P}\}$ is a set of **tape symbols** containing the subset

$\Sigma = \boldsymbol{\alpha_1}, \ldots, \boldsymbol{\alpha_M}$, the set of **input** tape symbols;

$\delta : S \times \Gamma \rightarrow (S \cup \{H\}) \times \Gamma \times \{$ **Left, Right**$\}$ is the **transition function**.

T has an infinite tape each of whose cells may contain a single symbol from $\boldsymbol{\Gamma}$, and it has a read-write tape head which accesses exactly one cell at a time and which can be moved one cell to the left or right on each application of the δ transition function. The δ function is the heart of a TM. For example, when machine **T** is in control state $\mathbf{s_p}$ and is reading symbol $\boldsymbol{\alpha_i}$, $\boldsymbol{\delta(s_p, \alpha_i) = (s_q, \gamma_j,}$ **Left)** indicates that **T** writes $\boldsymbol{\gamma_j}$ on the tape, moves the tape head one square to the left, and changes to control state $\mathbf{s_q}$. This is how a TM processes symbols.

Is this strictly a formal process? To be so, one must guarantee that the determination of $\boldsymbol{(s_q, \gamma_j,}$ **Left)** from $\boldsymbol{(s_p, \alpha_i)}$ entailed no association of information with $\boldsymbol{\alpha_i}$ that is not intrinsic to the symbol itself. But this is impossible to guarantee. It is entirely possible that in the δ transition, the processing of symbol $\boldsymbol{\alpha_i}$ involved an association with stored information, which in turn influenced the decision to produce the resulting machine configuration $\boldsymbol{(s_q, \gamma_j,}$ **Left).**

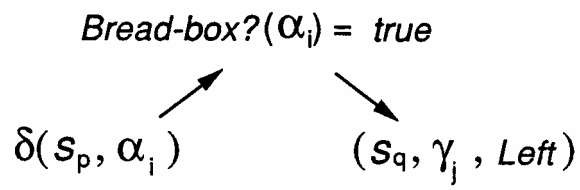


Figure 4: A non-formal δ transition on symbol α_i. The new machine configuration depends on information about α_i that is not an intrinsic property of the symbol.

Figure 4 illustrates how such a non-formal δ transformation might occur. During the transition process the symbol α_i is interpreted yielding, in particular, the fact that α_i is bigger than a bread-box. This information in turn influences the ultimate transition to the new machine configuration. Thus, the transition was not solely dependent on formal properties of the symbol α_i. It also depended on information internal to the machine and on the ability to associate α_i with portions of that information.

Simulated Turing Machines: Weak Equivalence: Why are UTM's formal symbol processors while other types of TM's need not be? A UTM is able to process input symbols on a purely formal basis because it is given the **δ-table** of the TM that it is suppose to simulate. That is, when a UTM, **U**, simulates another TM, **T**, **U** is given a table describing the complete input-output behavior of **T**'s δ function. **U** then performs a **weakly equivalent** (Pylyshyn, 1984) simulation of **T**'s δ function. Weakly equivalent computations need only maintain input-output or black-box equivalence: given the same inputs yield the same outputs. Stronger equivalence entails not only input-output equivalence, but also some degree of equivalence in the manner in which outputs are produced from inputs. Thus, strong equivalence requires some degree of algorithmic equivalence.[3]

[3]Pylyshyn (1984) points out that the notion of strong equivalence has many levels depending upon the level of abstraction at which one describes an algorithm.

For **U** to perform a strongly equivalent simulation of **T**, **U** would not only have to recreate **T**'s δ function's input-output behavior; it would have to do so "in the same way" as **T**. Suppose that the creation of **T**'s δ-table entailed the interpretation of some input symbol. This interpretive process would be completely hidden from **U** because only the result would appear in **T**'s δ-table. As far as **U** is concerned a δ-table is a completely opaque input-output representation of another TM's δ computation.

For a UTM, however, such strong equivalence is never necessary. **U** need only mimic **T** only insofar as necessary to recreate **T**'s global input-output behavior. Because **U** has **T**'s δ-table, it is never necessary for **U** to also recreate the process by which the contents of the table were determined. It is sufficient for **U** to look up the answer—a completely formal process. Thus, a weakly equivalent simulation of **T**'s δ function will always be sufficient, and, therefore, **U** may always treat the input symbols as formal objects only.

On the other hand, when a non-universal machine for **T** is constructed, there is no δ-table. The δ transitions are transitions between physical states of the machine in accordance with the logical states of **T**'s algorithm, to which they correspond. It is impossible to guarantee that, during these transitions, an input signal does not become associated with stored information about its meaning. Such an association is entirely possible. This analysis strongly suggests that P1 is immune, not only to the Chinese room and Gödelian arguments, but to any argument founded on the claim that TM's are incapable of non-formal symbol processing.

Conclusions

The failure of the Chinese room and Gödelian arguments to refute P1 does not, of course, establish its truth. However, most criticisms of the computability of mind are reactions against the process of "merely following an algorithm," i.e., against the rote manipulation of symbols according to a fixed set of instructions. This view of computation is strongly reinforced by UTM-computers and formal systems, both of which faithfully interpret rules and apply them to uninterpreted data. The sole purpose of such activity is to produce a genuine instance of some other TM. Therefore, such computations cannot exhibit key mental traits such as the pursuit of internally determined goals or adaptation through learning from input-output experiences. Understandably, this is a prevalent view of computation, but it is misplaced. It is accurate only with respect to UTM's and other non-adaptive TM's.

Full-fledged Turing machines are not so limited. The true computer metaphor of mind is the **Turing machine metaphor of brain**. The correct proposition to consider is whether human mental processes are within **some** class of Turing machine computations, not simply the universal ones, or ones otherwise prohibited from input-generated learning. General types of TM's are immune from the arguments put forth by critics whose intuitions are based upon surrogate TM's. In particular, it is not true that TM's are incapable of processing symbols in a subjective and dynamically evolving manner.

A Turing machine is governed by its internal programming which may change over time. A TM can be endowed with specific goals. It can receive inputs from an environment, and its outputs can affect its environment. A TM can remember input-output experiences, and it can form generalizations. A TM can associate inputs with current memories and generalizations, enabling it to produce novel outputs. It can develop predictive or causal models of phenomena in its environment, including models of itself. In short, a TM can learn to control its environment in an effort to satisfy its internal goals.

There are many reasons for believing that the control processes of biological machines—including the minds of human brains—are instances of Turing machine computations. These views stem, not from a failure to appreciate the power and intricacy of human thought, but from an appreciation of the power and intricacy of Turing machine computations. Attacking or defending the semantic powers of UTM's or formal systems is a waste of time. The semantic powers of Turing machines are what matter. A more productive exchange of views, therefore, might focus on developing correct intuitions regarding Turing machines, as well as determining what it would take for their computations to be considered truly meaningful.[4]

Department Of Computer Science
University of Massachusetts at Amherst
Amherst, Massachusetts

[4] Special thanks go to Sharad Saxena for many useful and interesting discussions of these issues. I also thank Neil Stillings, Andrew Barto, David Banach, Kevin Staley, John Moore, and Lance Williams for comments on drafts of this work.

References

Arbib, M. A. (1987). ***Brains, Machines, and Mathematics*** (second ed.). New York: Springer-Verlag. Previous edition: 1964 by McGraw-Hill, Inc.

Bowie, G. L. (1982). "Lucas' number is finally up." ***Journal of Philosophical Logic***, 11, 279-285. Slightly revised version reprinted in Jay L. Garfield (Ed.), ***Foundations of Cognitive Science: The essential readings***, New York: Paragon House, 1990.

Chalmers, D. J. (1990). "Computing the thinkable." ***Behavioral and Brain Sciences***, 13(4), 658-9.

Davis, M. (1990). "Is mathematical insight algorithmic?" ***Behavioral and Brain Sciences***, 13(4), 659-60.

Dennett, D. C. (1990). "Can machines think?" In Kurzweil, R. (Ed.), ***The Age of Intelligent Machines***, Cambridge, MA: MIT Press, 48-61.

Harnad, S. (1990). "The symbol grounding problem." ***Physica D***, 42(1-3), 335-346.

Hofstadter, D. R. (1979). ***Gödel, Escher, Bach: An Eternal Golden Braid***. New York: Vintage Books.

Lucas, J. R. (1961). "Minds, machines and Gödel." ***Philosophy***, 36, 120-124. Reprinted in A. R. Anderson (Ed.), ***Minds and Machines***, Englewood Cliffs, NJ: Prentice-Hall, 1964, 43-59.

Mortensen, C. (1990). "The powers of machines and minds." ***Behavioral and Brain Sciences***, 13(4), 678-9.

Nagel, E. & Newman, J. R. (1958). ***Gödel's Proof***. New York: New York University Press.

Penrose, R. (1989). ***The Emperor's New Mind***. New York: Oxford Universe Press.

Penrose, R. (1990). "The nonalgorithmic mind." ***Behavioral and Brain Sciences***, 13(4), 692-703.

Putnam, H. (1960). "Minds and machines." In Hook, S. (Ed.), ***Dimensions of Mind: A Symposium***. New York University Press. Reprinted in A. R.

Anderson (Ed.), ***Minds and Machines***, Englewood Cliffs, NJ: Prentice Hall, 1964, 72-97.

Pylyshyn, Z. W. (1984). ***Computation and Cognition: Toward a Foundation for Cognitive Science***. Cambridge, MA: Bradford Books/MIT Press.

Rucker, R. (1982). ***Infinity and the Mind: The science and philosophy of the infinite***. Boston: Birkhauser.

Searle, J. R. (1980). "Minds, brains and programs." ***Behavioral and Brain Sciences***, 3(3), 417-457.

Searle, J. R. (1987). "Minds and brains without programs." In Blakemore C. & Greenfield, S. (Eds.), ***Mindwaves: Thoughts on Intelligence, Identity and Consciousness***, Chapter 15, 208-233. Oxford, U.K.: Basil Blackwell, Ltd.

Searle, J. R. (1990). "Is the brain's mind a computer program?" ***Scientific American***, 262(1), 26-31.

Searle, J. R. (1992). "***The Rediscovery of the Mind***." Cambridge, MA: Bradford Books/MIT Press.

Sloman, A. (1985). "What enables a machine to understand?" In Josi, A. (Ed.), ***Proceedings of the Ninth International Joint Conference on Artificial Intelligence***, 995-1001, Los Altos, CA: IJCAI, AAAI, Morgan Kaufmann.

Turing, A. M. (1950). "Computing machinery and intelligence." ***Mind***, 59 (236). Reprinted in D. R. Hofstadter and D. C. Dennett (Eds.), ***The Mind's I***, New York: Bantam Books, 1981, 53-66.

Tymoczko, T. (1990). "Why I am not a Turing machine: Gödel's theorems and the philosophy of mind." In Garfield, J. L. (Ed.), ***Foundations of Cognitive Science: The essential readings***, Chapter II-4, 170-185. New York: Paragon House.

Leibniz on Innate Ideas

Shaun A. Champagne

G.W. Leibniz, in his book ***New Essays on Human Understanding***,[1] argues for the existence of innate ideas. In order to argue for this position, Leibniz uses the dialectical method. Within the dialogue, Theophilus represents Leibniz (who is a rationalist), while Philalethes represents the position of John Locke (who is an empiricist). For the sake of clarity, it is important to state the basic beliefs of both men. Philalethes begins by denying the notion of innate ideas and innate principles. He believes that all of our knowledge may be gained without the aid of innate impressions. Theophilus, not only accepts innate ideas (especially Descartes' innate idea of God) which could not **originate** from the senses, but he also says that **all** thoughts and actions come from within us. Although this is his belief, he will "conform to the accepted ways of speaking" and say that we have innate ideas while the senses may be said to be mediate causes of our thoughts.[2] This notion, which will appear several times throughout the dialogue, is (either directly or indirectly) the main focus of Philalethes' questioning. Theophilus' main point is that many people (namely Locke) have "not adequately distinguished the origin of necessary truths, whose source is in the understanding, from that of truths of fact, which are drawn from sense-experience and even from confused perceptions within us."[3] It is evident that most of the dialogue is focused upon the exact nature of innate ideas.

Leibniz first wants to show that an innate principle is a sort of inclination that we have within us. Philalethes tells Theophilus that if he can name a proposition that is innate, then he should name it. Theophilus answers that the propositions of arithmetic and geometry are innate ideas. However, he makes it clear that it is not the actual knowledge that is innate in us, but only the potential knowledge. To help one visualize how we might have potential

---

[1] G.W. Leibniz, ***New Essays on Human Understanding***, trans. by Peter Remnant and Jonathan Bennett (Cambridge: Cambridge University Press, 1981), p. 74-88.

[2] Ibid., p. 74.

[3] Ibid., p. 75.

knowledge, Theophilus compares it to the veins in a slab of marble which exist prior to being uncovered by the sculptor. Theophilus continues by saying that these innate truths, existing within us without our awareness of them, are "tendencies or dispositions" rather than thoughts. Philalethes is perplexed. How can these truths exist in us without our mind having thought of them? Again, Theophilus makes reference to the marble; it is not hard for us to conceive of the veins which exist within the marble prior to their being uncovered. Theophilus goes even further when he puts his last two points together to conclude that if these innate truths do not need any thought and are "tendencies or dispositions," then there is nothing preventing the possibility of truths existing within us that "have never and will never be thought about by us."[4]

Philalethes then questions the claim that there are general maxims that exist innately. For we find no trace of these maxims in "children, idiots and savages." Theophilus answers by stating that innate principles only appear when attention is paid to them.[5] Therefore, these people (children, idiots, and savages) do not show any traces of these truths due to their inability or unwillingness to focus upon the sense-experiences which unlock these maxims. To make the analogy (of the veins, marble and sculptor as compared to the innate ideas, us and sense-experience) clear and complete, we should consider it this way: Just as the sculptor chips away at the marble and reveals the veins that lie within, so does our sense-experience impact us so as to uncover the innate ideas which lie dormant in our understanding until then. If the sculptor does not chip away at the marble, then the veins will remain hidden. Likewise, if sense-experience does not affect us, then innate ideas will remain hidden within us.

Essentially, Theophilus wants to show that there exists within us innate truths which are made evident by sense experience. Philalethes not only questions this, but he also doubts that there are any innate truths on the grounds

---

[4] G.W. Leibniz, ***New Essays on Human Understanding***, trans. by Peter Remnant and Jonathan Bennett (Cambridge: Cambridge University Press, 1981), p. 86-87.

[5] Ibid., p. 87.

that such principles would imply universal agreement on the part of all people; yet, we find no such agreement.

Accordingly, Philalethes constructs a two-part argument with the hopes of disproving innate ideas by showing that Theophilus' assumption, "we witness the existence of common notions in the minds of **all** men," is false. First of all, Philalethes says that universal notions do not necessarily imply innate ideas. Secondly, he says that we do not witness any truths that are agreed upon universally; some people do not even **know** about the two great necessary truths: "Whatever is, is" and "It is impossible for something to both be and not be at the same time." Theophilus avoids these problems by saying: (1) General acceptance of a principle by most people (as opposed to universal acceptance by all people) is only a sign that hints towards its innate origin. Its being innate can only be proven when the certainty of the principle "comes from only what is within us." (2) Principles need not be known in order to be innate, they only need immediate acceptance as soon as they are made known. To make his point, Theophilus uses the example that all people have within them an **inclination** to worship a "higher power" due to the wonders of the universe that we are witnesses to. "A child deaf and dumb from birth has even been seen to worship the full moon."[6] The idea of God would be an innate idea; we are inclined to this notion due to our observation of nature. From what we witness, we immediately worship due to this innate idea of a "higher power."

It is in the above argument that we are able to see a real clash between the empiricism of Philalethes (Locke) and the rationalism of Theophilus (Leibniz). Philalethes thinks that he can disprove innate ideas empirically (and logically): "If ideas are innate, then they would exist in all men. It is obvious that there are no ideas which exist within all men. Therefore, innate ideas do not exist." This argument is a valid one (modus tollens). And in order to disprove this argument, Theophilus will have to show that one of the premises is false. The premise Theophilus attacks is the second premise: "There are no

---

[6] G.W. Leibniz, ***New Essays on Human Understanding***, trans. by Peter Remnant and Jonathan Bennett (Cambridge: Cambridge University Press, 1981), p. 75-76.

ideas which exist within all men." For Philalethes, this premise is the most certain; it is empirically obvious that not all men possess the "innate" principles, "It is impossible to both be and not be at the same time" and "There is a God." Theophilus refutes this by replacing the empiricism of this second premise with his rationalism. Innate ideas, the fact that they are in all men, may be proved "by its being shown that their certainty comes only from what is within us."[7] Therefore, innate ideas, insofar as they are in the understanding (which all of us possess), are in all men.

Two paragraphs ago, I emphasized the word 'know' for Philalethes and the word 'inclination' for Theophilus. These two words are important in defining what it means "to be in the understanding." Philalethes says it is a contradiction to claim that innate ideas are in us without our **knowing** them.[8] In other words, he is questioning what Theophilus means by the phrase, "to be in the understanding." But Theophilus again makes the point that an innate idea does not imply our constant awareness (or knowledge) of that idea; "to be in the understanding" does not signify actual knowledge. Instead, we should recognize innate principles as potential knowledge. By potential knowledge, Theophilus does not mean the capacity of our understanding to know certain truths. But rather, he means our mind's **inclination** (or disposition) to accept the truths contained within our understanding once they have been uncovered by sense experience.[9] Philalethes is strictly focused upon truths of fact, which are most certainly gained through sense experience; whereas Philalethes would simply call this knowledge, Theophilus would label it actual knowledge. Philalethes, then, because he sees the possibility for only one type of knowledge where Theophilus sees two, is perplexed. But as was pointed out earlier, Theophilus wants to distinguish between the origin of truths of fact and that of necessary truths (under which innate ideas are categorized).

---

[7] G.W. Leibniz, ***New Essays on Human Understanding***, trans. by Peter Remnant and Jonathan Bennett (Cambridge: Cambridge University Press, 1981), p. 76.

[8] Ibid.

[9] Ibid., p. 80.

So Philalethes errs (according to Theophilus), insofar as he does not acknowledge a difference of origin when speaking of the necessary truths. Theophilus explains later that the mind is capable of apprehending both truths of fact and necessary truths. And while truths of fact have their origin in sense-experience, necessary truths must come from the understanding itself; no matter how often we experience a truth, we could never infer its necessity inductively.[10] Therefore, this necessity must come from the understanding itself, due to its disposition to readily accept a particular truth. Theophilus claims that only through reason are we able to know that a truth is necessary. These necessary truths are what Theophilus would call innate ideas. And these innate ideas are only potential knowledge (i.e., they are readily and necessarily accepted as true) until they become actualized due to our sense experience.

St. Anselm College
Manchester, NH

[10] G.W. Leibniz, ***New Essays on Human Understanding***, trans. by Peter Remnant and Jonathan Bennett (Cambridge: Cambridge University Press, 1981), p. 81.