Where Reason Meets Poetry:

Toward the Unity of the Rational and the Romantic

A collection of articles by Mark F. Sharlow

Where Reason Meets Poetry: Toward the Unity of the Rational and the Romantic

Introduction

This collection of papers, essays, and blog posts is meant to support for three conclusions.

1. The scientific view of the universe is fully compatible with the poetical or romantic outlook recorded by poets, artists, and certain philosophers. There is no real conflict between the scientific perspective and the romantic outlook, provided that both perspectives are understood correctly.

2. The scientific view of the world is fully compatible with many important beliefs that ordinary people hold about the nature of reality. People normally believe that there are real moral values in the world; that beauty in art or in nature is more than just an illusion; and that persons are important beings (not mere specks of dust in the universe) who have intangible mental or "inner" qualities as well as physical properties. *The scientific view of nature is fully compatible with these commonplace insights.* The fact that some other popular beliefs are scientifically unsound does not change this in the least.

3. It is incorrect to claim that *only* the particles and forces that make up the physical world are ultimately real. Persons and the macroscopic things among which they live, and the everyday properties of those persons and things, are just as significant, and as ultimately real, as the particles and forces revealed by science.

This collection of writings covers several topics, but as a whole it supports these three points. I have touched on these points many times in my writings, but I did not bring these writings together into one place until now.

This collection is only a start; it does not build a complete case for all of the three points. For the rest of the line of argument, consult my books *From Brain to Cosmos* and *The Unfinishable Book*, and my writings on the philosophy of religion. Most of these sources are available on my website and in the online archives where my papers are stored.

Some of the articles in this collection are drafts and preprint versions of philosophical papers, while others are informal writings such as blog posts. For these reasons, readers familiar with my published papers (such as those in *Analysis*, *The Journal of Symbolic Logic*, and physical science journals) might find that some of the papers in this collection seem unfinished and rough. This might change in some future update of the collection (though I can't promise). It can take a while to get feedback from colleagues on the preprint version of an article. Some of the papers here have not yet been circulated enough to reach their finished form.

For most of the articles I used the versions from my website with few or no changes. This explains the puzzling references to my website URLs that exist in some of the papers.

I'd like to hear your comments on this project. As of the time of this writing, my e-mail address is msharlow@usermail.com . In case this address changes, you can probably find my new address on my website or in my profiles in the archives where my papers are found. The URL of my website, at the time of this writing, is http://www.eskimo.com/~msharlow .

- Mark Sharlow

Contents

Note: The page numbers in this table of contents are logical page numbers for the collection as a whole; the title page of the collection is page 1. Some of the articles in the collection have their own visible page numbers which are different from the numbers in the table of contents.

Restoring the Foundations of Human Dignity:

Upholding the reality and significance of persons in an age of cynicism

The importance of the person is an endangered idea in today's philosophical thought. Many traditional philosophical views emphasized the freedom, autonomy, and dignity of persons. Today, philosophical doctrines that marginalize personality seem to have gained the upper hand. Among these doctrines are:

Scientism, which teaches that science is the only legitimate form of knowledge. (If taken seriously, this leads to the view that a person is only a mass of chemicals.)

Behaviorism and **eliminative materialism,** which teach, in different ways, that the human mind is unimportant and perhaps even unreal.

Determinism in its **incompatibilist** form, which teaches that persons do not have free will.

Postmodernism, which sometimes teaches that persons are mere fictions of language, and that personal qualities like reason are social constructs invented by "oppressors" (ethnicities or genders that the postmodernists do not favor).

Most of these lines of thought seem scientific at first glance. Postmodernism is the exception; it does not pretend to be scientific, and it tends to be antiscientific. Despite their differences, all these doctrines deny or undermine the reality and dignity of the person.

Forget what you have heard from the overconfident followers of these beliefs. **Science has not confirmed any of these doctrines - and philosophy has not confirmed them either.**

Indeed, some of the more scientific-sounding of these ideas are scientifically untestable, so there is no chance science will confirm them.

The literature of philosophy contains many arguments against doctrines like these. This literature is too extensive to list here, though I would like to do so. Anyone who searches the literature deeply enough will find that **all** of these doctrines are controversial. None of them has found general acceptance by all serious philosophers. There are arguments for and against all of these ideas. Sometimes scientists who are not philosophers come out in favor of these views - but the philosophical literature already contains arguments that refute their pronouncements.

There is no scientific or philosophical "proof" for any of these antipersonal viewpoints. The truth of each of them remains an open question at best. There still is plenty of room for confidence in the opposite views - and for confidence in the importance and dignity of persons.

On this page, I will summarize some of the main points of my own view of persons. In some places I will provide links or citations to relevant points (or at least related points) in my writings.

Skepticism about the reality of consciousness is untenable. To claim that you only seem to be conscious is, in effect, to claim that things don't really seem a certain way - they only *seem to seem* that way. This latter claim leads to inconsistency. The claim that consciousness has no phenomenal or subjective character is untenable for the same reason. ([1]; see also [2])

Skepticism about the reality of mental states is untenable. So-called "folk psychology" - the commonplace set of beliefs that people generally hold about the human mind - has a solid core that is not in danger from science-driven skepticism. Science can cast doubt on some beliefs about the mind, but it cannot show that humans do not have thoughts, feelings, desires, and the like. [3]

The conscious subject is a single, unified entity. Disunifying phenomena, such as self-division and unconscious influences on the will, cannot compromise the basic unity of the subject, though they can seem to do so. [1]

Science has not refuted free will. Many philosophers today are *compatibilists*; they hold that free will could exist even in the presence of *determinism* (the causal determination or predictability of all physical events). I concur with the compatibilist view. Even if determinism were true, there could be free will. (There are plenty of compatibilist arguments already in the literature.) The possibility that our actions are controlled entirely by unconscious neural events is perhaps a greater threat to free will than is simple physical determinism. But even this circumstance would not rule out free will - because even a so-called "unconscious" brain event may actually lie within the scope of personal consciousness, and therefore be one's own doing. ([4], [1])

Reality does not consist of concrete physical objects alone. It also contains *abstract objects*, such as properties, relations, and sets. These are not concrete objects made of matter or energy. Hence materialism is not a complete view of the universe. (Note that the incompleteness of materialism does not imply supernaturalism. There is nothing "supernatural" about properties, relations and sets.) The idea that abstract objects are fully real is called *ontological realism*. This is a very old idea in philosophy. I argue that ontological realism is not the extravagant doctrine that some say it is. Indeed, ontological realism requires us to believe very little beyond what we already know from everyday experience. [5]

The self is real - and no scientific discovery about the mind can prove otherwise. It is plausible to identify the self with a fully real abstract object of the kind discussed in the point about abstract objects, above. ([6], [7]) Since abstract objects are genuinely real, a self of this kind would be genuinely real too. Even if neuroscience found no evidence of a self, this abstract object could be the self, and its existence would not be falsifiable by science. Some authors seem to think that if the self were "only" an abstract object, then the self would not be real. This argument fails if we accept that abstract objects are fully and genuinely real. (It is unwise to say that anything is "only" an abstract object.)

The qualia, or subjective qualities of conscious experience, are real. Qualia are the subjectively felt features of personal experience - for example, the "feel" of the color red, of a particular pain, or of the musical note middle C. Qualia are abstract objects. As I said earlier, abstract objects are real entities. If we identify qualia with suitable abstract objects, we find that the existence of qualia is not falsifiable by science. The possibility that neuroscience has no need for qualia cannot weigh against the reality of qualia. [8]

Language really can refer to reality; this reference is not merely a social construct or a political fiction. Once one understands how language is related to the way things seem, one finds that language can refer to an objective reality. Hence, postmodern critiques of the referentiality of language must fall apart at some point. ([1]; see also [2] and [9])

The existence of different cultural perspectives does not rule out the reality of objective truth. Although there are many different cultural perspectives, there still is such a thing as objective truth - a truth which, in a sense, encompasses all the perspectives. Hence, postmodern dismissals of objective reality and truth are extravagant and pointless. [1] If one wants to respect all cultures, one should assert that there is objective truth, instead of denying this as so many postmodernists do. (If there were no objective truth, the claim that different cultures deserve respect could not be true.)

Conscious subjects play important roles in physical reality. The physical universe is objectively real, is not a mental construct, and is vast compared to humanity. Nevertheless, the physical universe is deeply intertwined with consciousness. All physical facts have logical ties to the actual and possible experiences of observers. Physical facts are dependent - not causally, but in a certain logical manner - upon facts about experience. Thus, conscious observers are not mere trifles. Consciousness plays a key part in the physical universe. [1]

Science is a valuable source of knowledge, but it is not the only legitimate knowledge. The view that science is the only legitimate form of knowledge is called *scientism*. Scientism, if taken seriously, would imply that philosophical knowledge is impossible. A follower of scientism cannot consistently adopt any philosophical positions - including scientism itself. There are other forms of legitimate knowledge besides science; philosophy is one of these forms. Also, the notion of truth is too rich to be exhausted by any single methodology, including that of science. [1] The statement that scientism is false is not a criticism of science itself, and does not alter the facts that science "works" and that truth is objective. (I should mention in passing that the fashionable postmodern critiques of science are hopelessly off track. Among its other faults, postmodern antiscience demeans people whose lives have been saved by modern scientific medicine.)

These philosophical points, taken as a whole, point to a new view of the person - a view that leaves abundant room for freedom, dignity and autonomy. This new view is based on reason and is fully compatible with science. It is not a finished philosophical system, but is open-ended and exploratory in character. Nevertheless, this view clearly overlaps with two enduring philosophical traditions: **humanism** and **personalism**. (By "humanism" I mean humanism in its original sense, not the scientism-based movement called "secular humanism.") Personalism and humanism both recognize the importance of persons. I suggest that the philosophical ideas presented here could serve as the seeds for a restoration of a truly humanistic and personalistic outlook in the twenty-first century.

References (by the same author)

[1] *From Brain to Cosmos*

[2] "Notes on From Brain to Cosmos: Questions and Answers about Subjective Fact"

[3] "As True as 'You Think': Preserving the Core of Folk Psychology"

[4] "Yes, We Have Conscious Will"

[5] "Getting Realistic about Nominalism"

[6] "I Am an Abstraction, Therefore I Am"

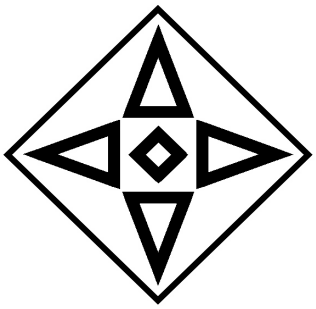
[7] "Platonizing the Abstract Self"

[8] "Qualia and the Problem of Universals"

[9] "How Subjective Fact Ties Language to Reality"

By Mark F. Sharlow. An earlier version of this document appeared on the author's website in 2007.

Poetry's Secret Truth



From the Beauty of Words
to the Knowledge of Spirit

by Mark F. Sharlow

Poetry's Secret Truth

Revised Edition

Table of Contents

1. Prelude

Poetry, it is said, can reveal truth. Yet despite the best efforts of philosophers and poets to describe this truth, very few understand what *kinds* of truth poetry can convey.*

One fact seems clear: only a few of the truths of poetry can be captured equally well in prose. Poetry also conveys truths of a different kind — truths that seem to exist on a level entirely different level from that of ordinary, factual truth.

Some poems try to teach moral or practical lessons that also could be stated in prose.* But this is not the kind of truth that puzzles philosophers and critics. Poetry also can tell another kind of truth — a truth that may be mystifying to scholars, but that is well known to anyone who becomes acquainted with poetry in an intimate way. This kind of truth cannot be spoken of or contemplated on the same terms as ordinary fact.

What is the nature of this strange, yet familiar kind of knowledge that poetry can bring to the human mind?

---

* * *Important Note:* In this book I discuss a number of ideas originated or discussed by other authors, plus some ideas of my own that are close to other authors' ideas. I have cited these items (and others) by placing asterisks near them and listing them by page number in the "Credits and Notes" section. This lessens the distracting effect of the notes.

2. A Different Kind of Fact

The special knowledge that poetry brings has many aspects. One of these aspects — perhaps the most basic — is the knowledge of *subjective facts*.

What is a subjective fact?

A subjective fact is a fact about how things *seem* to some conscious being.* These facts are different from the ordinary, objective facts that we usually recognize as "facts." When you encounter a flowering apple tree in spring, the fact that the tree is a certain number of feet tall is an objective fact. But the way the tree seems to you — the look of its leaves and flowers, the rustling noises it makes in the wind, the feelings and thoughts it brings to mind — all this is a matter of *subjective* fact. Subjective facts are facts about how things seem to a particular observer at a particular moment. They are facts about how the world seems in a certain momentary instance of conscious experience.

Poetry, with its known power to evoke experiences and associations,* is able to express *subjective facts* about a thing or situation. Poetry, at its best, can evoke *a new realm of subjective facts* for its reader. This realm may involve feelings, thoughts, and sensations that are new to the reader, along with others that are more familiar.* In either case, this realm of subjective fact is new, because it differs from the domain of subjective facts that came with the more routine experiences that the reader was having before reading the poem. Some people regard poetry as a record of experiences or feelings. This belief can be the case if the poem creates for the reader subjective facts like those that the poet has experienced, or like those that the poet wanted to evoke. (The same is true of other arts besides poetry. Much of what I will say here also holds true for the other arts.)

The subjective facts that poetry can evoke are inner facts of sensation, feeling, and

perception. These facts can encompass a wide spectrum of contents, ranging from familiar emotions (happiness, longing, concern) and sensations (colors, sounds, scents), to content of a far subtler and more enigmatic sort. These subtler contents include such things as the elusive, almost indescribable sensations that fill one's awareness when one encounters a flowering apple tree in spring.

Anyone who has fully and deeply experienced an apple tree in full flower in a rolling, rustling spring meadow will know what I mean by this.

Anyone who has become fully aware of the mysterious looming of the clouds in the hours before rain, or of the charged, green freshness after the rain, or of the almost audible silence of some warm summer afternoons, will know whereof I speak.

The best poetry evokes not only the coarser sensations and emotions, but these subtle feelings and sensations as well.* The subtle impressions that poetry can communicate could be described as feelings, but they have little in common with the noisy and simplistic emotions of everyday life. Often, these feelings seem more like sensations or perceptions than feelings — but perhaps they can be all of these at once.

Poetry, at its best, is capable of evoking subtle experiences of this kind. It is capable of putting the reader in a frame of mind from which he or she may become more aware of the fullness of subjective experience, and hence of subjective fact.

3. A Secret and a Wider View

It is no secret that in ordinary life, we tend to experience only a few of the impressions that we can receive from the world.* We pass by the apple tree inattentively, thinking of something else. We may note its more obvious features — that it is tall, has leaves on it, and so forth — but we miss the full inflowing of experience.

Poetry makes possible the experience of a *much fuller* range of subjective facts — of subjectively beheld characteristics of things and of the world around us.*

The evocation of a widened range of subjective facts is not the only possible function of poetry. However, it is this function that most concerns me personally. In my view, this is the most important function of poetry. The other functions of poetry depend on this function, in the sense that the other impacts of poetry can occur best (or only) when subjective facts are effectively evoked.

In some instances, a poem actually may communicate subjective facts from the poet to the reader. This is what happens when a poem succeeds in communicating feelings and experiences as the poet intended.* In other instances, a poem simply reminds the reader of subjective facts of which he or she already is subliminally aware, but has not given proper attention.* The successful evocation of subjective facts requires an effort on the reader's part as well as skill on the part of the poet. The poem must be read, not just mechanically traced over with the mind. It is well known that the reading of a poem requires the active use of the imagination.*

The realm of subjective fact that a good poem brings to light includes things of which we normally fail to be aware. Thus, poetry does not only tell the truth; it also tells a secret. The secret is the new realm of experience and feeling that poetry reveals. No one

is hiding this "secret" from us. Usually we hide it from ourselves.

One can ask several philosophical questions about subjective facts. Some of these questions have a bearing on what I am saying about poetry.

The first of these questions is whether subjective facts belong to the real world, as objective facts supposedly do, or whether they are imaginary, unreal, or "all in the mind." The answer to this question is not obvious, but is simple: yes, subjective facts belong to reality. When you experience a certain impression while contemplating an apple tree in spring, it is a fact that this impression occurred. This fact, though subjective, is objectively true — it really is true that things seemed *just that way at that moment*. A subjective fact, though truly subjective, also is as objective as any so-called "objective" fact.

Normally we do not understand the relationship between the subjective and the real.* We regard the realm of the subjective as merely imaginary and not real — whereas actually, facts about how things seem are as true as any other facts in the world. This is the case whether the subjective facts result from sense experiences or from pure imagination. Strangely enough, subjective facts are objective!* Of course, there is no real contradiction in this. It is one of those paradoxes that flouts our assumptions but nevertheless happens to be true.

Of course, this *objectivity of subjective fact* does not imply that everything which seems to be the case really is the case in the physical world. By saying that subjective facts are objective, I am not saying that because someone sees a unicorn in a dream, there really is a one-horned, physically existent animal. Instead, I am saying that there is a fact that that there *seemed* to be a unicorn — and that fact is totally objective. The facts about how things seem are the subjective facts of the world. (Whether there ever were unicorns in the past is a zoological question that I won't take up here!)

4. In the Mind, or in the Tree?

Another, related, philosophical problem is the question of whether the subjective facts can be said to be facts about the external world at all. Think about the flowering apple tree again. Don't the subjective facts evoked by the tree come from processes in the observer's mind or brain, instead of from anything in the apple tree? My answer to this question is twofold. First, the existence of subjective facts does indeed depend upon the state and presence of the observer. Second, the subjective facts are not just features of the observer's mind or brain. They are real features of *the observer plus the object being observed*. The whole system, observer plus object (in other words, you plus apple tree), is the source or seat of the subjective fact. The experience is an experience *of* the object, not only an experience *in* the observer.

Subjective facts are not "only in the mind." They are characteristics of the observer-object couple. They are relative, but only in the same way that certain measurable features of the physical world are relative. According to modern physics, the size and mass of an object depend upon the state of the observer (specifically, the observer's state of motion) as well as upon properties of the object in itself. But this relativity of size and mass does not mean that an object's size and mass are unreal or are "all in the mind." Subjective facts, like facts of size and mass, are simply relative to the state of the observer. However, in the case of subjective facts, it is the observer's state of mind, not state of motion, which matters. In spite of their dependence on the observer's mental state, the subjective facts about the apple tree are every bit as real — or as unreal — as the tree's size or mass.

This relativity of subjective facts also encompasses what happens when people experience very different things during encounters with the same object. A particular

scene may seem happy to one person, sad to another — perhaps due to the observers' past experiences, mental associations, and the like. This only means that the subjective facts depend on the state of the observer as well as upon the state of the object. It does not mean that the subjective facts are unreal.

One can think of the many possible subjective appearances of an object as *possibilities* inherent in the object* — all of them equally real features of the object, or perhaps of the world. Some possibilities may be more crucial, or more important to our understanding, than others. Yet all of the possibilities are there, and all of them are parts of reality. (Philosophers belonging to the school of thought called *phenomenology* have argued that we should take into account the multiple possible ways of perceiving things. What I am proposing is different, with a different range of possibilities, and involves something more — an additional element or factor.)

Some people believe that poetry can help the mind to grasp the true nature of things — that poetic experience can produce a deeper contact with reality, and a more complete view of reality, than can ordinary experience.* When we consider poetry in the light of the objective reality of subjective fact, we begin to see that this must be the case.

Poetry, then, can speak the truth — a truth different from the truths of intellect. Poetry is capable of revealing, or pointing to, the rich stratum of subjective fact that permeates the world in which we live, and that often goes almost entirely unnoticed during our routine, unobservant existence.

5. Art, Science, or Both?

Another philosophical question about poetry concerns the relationship between poetry and science. It has been said that poetry and science make use of different ranges of experience, and that poetry can treat of any subject matter.* Is the truth that poetry can speak related to scientific truth? Is there any common ground?

To address this question, we must face up to an important fact about science: that scientific knowledge, like poetic knowledge, *is based on subjective facts*. Scientific conclusions are supposed to stand or fall according to the evidence provided by experiment and observation. The data of experiment and observation ultimately grow from the soil of someone's experience. A physicist sees a meter pointing to a certain number. A biologist observes a particular bird engaging in a particular mating ritual. A chemist hears a hissing noise when chemicals are mixed, and sees a certain color change. All scientific knowledge, if it is indeed scientific, is judged by means of experience — and hence by means of subjective facts.

This is not to say that science is based only on subjective facts. But science does require subjective facts and cannot exist without them. Science rests upon experience, and hence upon subjective facts.

It is interesting to realize that the subjective facts which science requires form a very narrow subset of the total fullness of subjective facts available to human consciousness.* The sensation of a meter reading — of a needle pointing to the number 2, for example — provides a subjective fact that science might need. The fact that a flowering apple tree in spring looks a certain way — *a certain unique, indescribable way* — involves a subjective fact that science overlooks, but that is equally real and true. (A scientist who is a

psychologist instead of a physicist might even use someone's report about the apple tree as data for a theory about human experience. But even this does not amount to the use of the subjective fact itself as data.)

Poetry, unlike science, can take the way the apple tree seems as "data." Poetry also could take as "data" the way the physicist's meter looks — if some poet cared to take the meter as a subject of a poem. Poetry has no restrictions on the range of subjective facts it can explore. Science deliberately restricts its attention to a narrow set of subjective facts.

6. A Key to the Spirit

The deep relationship between poetry and subjective facts brings us to a crucial observation about poetry. This is that poetry, through its power to reveal subjective facts, *actually reveals the realm of the spirit to the human mind.*

What is the spirit? This is an ancient and persistent question. We do not have to have a complete answer to this question to see where poetry can lead us. Different religions and philosophies have different ideas about the exact nature of the human spirit. Many teachings, both religious and philosophical, hold that the spirit is something besides the body. Some philosophies teach that the spirit is simply an aspect of the activity of the human body, especially of the brain. Some people feel that the spirit is the same as the personality or mind. Others believe that the spirit is something more than, and deeper than, what we usually call the mind or personality. Yet all of these systems of belief have something in common. All of them connect the concept of spirit with the concept of *consciousness*. And all of these belief systems are right about this.

A being with spirit can only be a *conscious* being. It cannot be an unconscious, dead lump devoid of any awareness. Such a lump would be lacking in spirit. (I am not claiming that such a lump really could exist. Some people, including the noted philosopher Leibniz, have even thought that everything contains spirit or consciousness.* But that is a separate question that I will not try to answer here.) Spirit, whatever it may be, involves consciousness. Even the unconscious mind, which some people regard as having spiritual qualities, is not truly "unconscious"; it is mental in character, is connected to the conscious mind, and may well be a kind of consciousness itself.*

Spirit involves consciousness. Consciousness, in turn, always involves *subjective fact*.

Philosophers have long noted that the most distinctive feature of consciousness is its subjective side — the inner experience or "feel" of being conscious. A being that lacks this subjective side would not be truly conscious, even if it appeared to act at times like a conscious being. At best, it would be a mere robot.* Without subjective facts, there is no real consciousness at all.

Because spirit involves consciousness and consciousness involves subjective fact, we must conclude that spirit, whatever it may actually be, involves subjective fact in an essential way. At least some of the subjective facts that conscious beings can encounter are spiritual facts. But are not all subjective facts ultimately spiritual in character? Subjective facts are facts of a special kind. They are not like facts about the visible behavior of the human body, or about the outward physical characteristics of conscious beings or of physical objects. Subjective facts pertain to the subjective "feel" of conscious experience — the aspect of conscious experience that makes experience truly conscious, truly personal, truly "alive," and truly *inner*.

Thus, all subjective facts belong to the spiritual side of conscious experience. This is the case even though only some subjective facts have to do with what we normally call "spirituality." The realm of subjective facts is a spiritual realm.

Depending on one's personal beliefs about the spirit, one might want to regard the realm of subjective facts as only a part of spiritual reality, instead of as the whole. But whatever one's view on this question, subjective facts still are facts of spirit. All conscious experience reveals subjective facts, and hence contains a spiritual element. The more that we notice the world around us with all its qualities and possibilities, the greater this spiritual element in experience will be.

Anything that lets the mind encompass a larger realm of subjective fact is a key to a broader knowledge of the spirit. Poetry, which is the key to new worlds of subjective fact, can perform that function.

7. A Surprising Conclusion

The ideas that I have presented here imply a rather dramatic conclusion: that *poetry can reveal many actual facts about the world, and some of those facts are beyond the scope of science.** What is more, those facts tell us something about the world of spirit.

This conclusion does not make science any less true, objective, or important. In its own sphere of operation, science sets the standard and calls the tune. But there are areas of reality, and not only of reality but of *fact*, into which scientific theory and observation simply cannot enter. What is found in these areas — in the greater parts of the realm of subjective fact — can be expressed and evoked by means of poetry. And these areas lie within the realm that we call the spiritual.

This conclusion will remain true even if science someday manages to explain how the human mind experiences things. Some people — who, for reasons of their own, choose to regard the meter readings more highly than the tree — will argue that the experience of seeing an apple tree can be explained completely in terms of the activity of the brain. I will not comment on this argument here, except to point out that such a complete explanation does not exist and may, for all we know today, be impossible. But even if a physical cause for subjective facts were found, the realm of subjective facts would continue to be real, and the *truth and meaning* of subjective facts would remain unchanged. Poetry indeed conveys truth — and this truth remains true, regardless of whether its roots lie partly in the earth or entirely in some other world.

About the Author

Mark F. Sharlow is both a philosopher and a scientist. He holds a Ph.D. degree in chemistry, and has worked as a chemistry professor and as a scientific computer programmer in the space field. His philosophical work has been published in the philosophy journals *Notre Dame Journal of Formal Logic*, *Analysis*, and *The Journal of Symbolic Logic*. His scientific work (done alone or with coauthors) has appeared in *Annals of Physics*, *Journal of Physical Chemistry*, and other journals. He has written a full-length philosophical book, *From Brain to Cosmos* (Parkland, FL: Universal Publishers / uPUBLISH.com, 2001), which is the source for some of the ideas used in the book you are now reading.

Dr. Sharlow's e-mail address (at time of this publication) is **msharlow@usermail.com**.

Credits and Notes

In this book I have discussed several ideas and questions about the truth and meaning of poetry, and have tried to assess these ideas and questions in terms of my own concept of subjective fact. I am not, by any means, the first to write about these topics concerning poetry, or to argue for the reality of poetic truth. I have learned much from those who came before. I wish to acknowledge particularly the work of Clyde S. Kilby, who, in his book *Poetry and Life* (reprint ed.; Plainview, NY: Books for Libraries Press, 1975), has given an able presentation of many key questions and ideas about poetry. Several of the page-numbered notes below point out specific instances in which I have discussed ideas and problems mentioned in Kilby's book, or in which I have come to conclusions similar to Kilby's views or to other ideas mentioned in Kilby's book. Also, I am grateful to Kilby for his clear overall treatment of the problems of poetic truth and meaning and of the nature of poetry. (See especially Chapters 1-2 and pp. 325-328 of *Poetry and Life*.)

The idea of subjective fact is developed in my earlier (and much longer) book, *From Brain to Cosmos* (Parkland, FL: Universal Publishers / uPUBLISH.com, 2001). In that book, I introduced the notion of subjective fact in a more technical way, and explored the relationship of subjective facts to philosophical knowledge. Some of the philosophical ideas used in the book you are now reading — including the objectivity of subjective fact and the conscious character of the unconscious mind — are developed and explained more fully in *From Brain to Cosmos.* (However, *From Brain to Cosmos* is not a book about poetry.)

Specific Notes:

Page 1 (and elsewhere): The idea that poetry can reveal truth and convey knowledge is a very old idea. Kilby accepts this view (see Kilby, Chs. 1-2, pp. 325-328, and especially pp. 70-77).

Page 1, "Some poems ... in prose": I am thinking particularly of didactic poetry (on which see Kilby, pp. 328-331).

Page 2: The notion of subjective fact is developed in *From Brain to Cosmos*.

Pages 2 and 3: Kilby discusses the power of poetry to evoke experiences, including familiar ones (see Kilby, Ch. 2).

Page 3, "... subtle feelings and sensations ...": Compare Kilby, pp. 326-328.

Page 4, first two paragraphs (especially the sentence "It is no secret ... from the world"): Kilby, pp. 56-57.

Page 4, "This is what happens ... the poet intended.": On the capacity of poetry to communicate feelings and experiences, see Kilby, pp. 65-69.

Page 4, "In other instances ... proper attention.": See Kilby, p. 64.

Page 4, "It is well known ... of the imagination.": See Kilby, p. 24.

Page 5, last two paragraphs: Presumably this is part of the reason why imagination can reveal reality. (See Kilby, pp. 325-327.) Note also that the idea of the objectivity of subjective fact is developed in *From Brain to Cosmos*.

Page 7, "One can think of ... inherent in the object": The possibilities I have in mind here may include, but are not restricted to, the imaginative possibilities noted by Kilby (pp. 57-59).

Page 7, "Some people believe ... than can ordinary experience.": This idea (which has a long lineage) is discussed, and relevant references are cited, in Kilby's book (Chs. 1-2, especially pp. 8-11, and pp. 325-328).

Page 8, "It has been said that poetry and science make use of different ranges of experience, ...": Kilby, pp. 70-75, and references therein.

Page 8, "... and that poetry can treat of any subject matter.": Kilby, p. 3

Page 8, last paragraph, and Page 9, first paragraph: Some ideas like these are explored in Kilby, pp. 71-73, and references therein.

Page 10: On G.W. von Leibniz's idea mentioned here, see his book *Monadology* (which exists in various editions).

Page 10: The idea that the unconscious mind has a kind of consciousness is discussed in *From Brain to Cosmos*.

Page 11, "At best, it would be a mere robot.": Various philosophers have speculated on the possibility of a being that acts just like a human being but has no experiences. They call this imaginary being a "zombie" — though it isn't much like the zombies of Haitian occultism!

Page 12, first and second paragraphs: The notes for Page 1 (first note) and for Pages 8-9 are applicable here as well.

As True as "You Think":
Preserving the Core of Folk Psychology

Mark F. Sharlow

ABSTRACT

In this paper I argue in defense of an important fragment of folk psychology. Specifically, I argue that many propositions about the ontology of mental states and about mental causation are true largely because of certain observable features of human linguistic behavior. I conclude that these propositions are immune to common avenues of eliminativist criticism. I compare and contrast this argument with some previous arguments about the truth of folk psychology.

1. Introduction

Folk psychology is supposed to be an informal theory about the mind—a theory that people normally acquire early and accept unthinkingly. The propositions that people have feelings and thoughts, that people's thoughts and feelings can cause them to act, and so forth—all these commonplace propositions, which people normally use to think about

the mind, constitute what philosophers call "folk psychology." According to a common position in the philosophy of mind, folk psychology is an empirical theory that might be false. Some philosophers—the eliminativists—have taken the extreme position that folk psychology is quite false, and that mental terms like "belief" and "desire" do not refer to anything at all. Eliminativist arguments typically rest on neuroscience; in one way or another, they try to establish that neuroscientific discoveries about the mind show (or probably will show) that folk psychology is false.[1] Other philosophers have mounted serious challenges to eliminativism. According to some of these challenges, no future discovery in neuroscience or cognitive science could give us a strong reason to abandon folk psychology.[2]

In this paper, I will not take up the usual question of the truth of folk psychology. Instead, I will argue for a weaker conclusion: that an important *part* of folk psychology is true. I will do this by means of two arguments. Each argument shows that a crucial class of propositions of folk psychology is independent, not only of scientific discoveries, but of many philosophical considerations as well. Together, these arguments show that folk psychology has a "safe" core—a set of central propositions that no set of scientific discoveries can refute, and that do not depend on the fate of philosophical arguments defending the whole of folk psychology. Thus, a significant part of folk psychology is independent, not only of scientific discoveries, but of the usual philosophical debates as well.

The line of argument presented here is only partly new. It partially overlaps, or at least coheres well with, previous defenses of folk psychology by Graham, Greenwood, Horgan, Margolis, and McDonough. For now I will cite the relevant works in an endnote.[3] In section 3 I will discuss in detail the similarities between these authors' ideas and mine.

2. Mental Language and the Classification of Situations

In the next several paragraphs I will present some general observations about the nature of folk psychology. These observations are neither new nor deep. I will present them with the help of the traditional language of folk psychology, with the caveat (for the sake of argument) that such language might ultimately be eliminable.

The commonsense psychological language that people use every day relies heavily upon the *classification of situations involving human organisms*. Greenwood's line of argument, with its emphasis on "classificatory descriptions of human action",[4] points us firmly toward this fact.[5] The following example, which has a precedent in that line of argument,[6] makes this point. Consider what happens when a child learns the word "think." The child learns to utter that word when certain situations occur that involve his own organism. He learns to utter tokens of sentences like "I'm thinking." In learning to use the word, the child learns to apply the word in connection with certain situations that the child's cognitive apparatus can recognize. The child's brain is able to discriminate these situations from other situations. While learning a language, the child learns to make utterances like "I'm thinking" in response to those situations.

The situations for which the child learns to say "I'm thinking" are more or less those situations that experienced speakers of the same language would call "situations in which the child is thinking." It is an observable fact that a child with typical language capabilities can learn to recognize these situations. *How* this happens—the neural mechanism of the discrimination, its social context, etc.—is beside the point for my argument.[7] I am not ruling out the possibility that the recognition ultimately is verbal in character—that learning the ways to use the word "thinking" is what gives the child the capacity to pick out situations of thinking.[8] I am not even ruling out the possibility (discussed by Greenwood[9]) that recognitions of this sort are theory-laden.

Regardless of the details of the mechanism, the child learns to apply the word "thinking"

to certain situations. If the child has typical neural capacities, he will be able to pick out certain situations from among other situations involving humans. He learns to label those situations as situations of "thinking." Of course, this learning involves the absorption of linguistic norms from the child's social surroundings. Learning to use the word "think" involves learning to discriminate situations that may properly be called situations of "thinking" from situations that may not be so called. In the preceding sentence, "may" indicates the "permission" obtained from the child's linguistic environment. If the child is bouncing a ball and says, "Look, Ma, I'm thinking", the child might be told, "No, that's not called thinking, that's called bouncing a ball." But suppose that the child says, "Look Ma, I've been thinking. Two plus three makes five", which is a fact that the child didn't know before but figured out on his fingers. Then an appropriate adult reaction is "Yes, you have been thinking." A certain physical situation occurred; the features of that physical situation are such that we are warranted in asserting that the child has been involved in a situation that we would call a situation of the child thinking. (Whether this situation is reducible to the child's behavior, or to functional states of the child, or to anything else, is a large and old question which, despite its importance, is completely irrelevant to my present argument.)

People come to regard certain situations involving human organisms (and perhaps other organisms or machines as well) as states of thinking. If some standard forms of materialism are true, then these situations are situations of brains being in states of certain kinds. Externalistic views of mind might equate these situations to situations involving both the organism and its surroundings. Regardless of the truth of these views, the process of learning how to use the words "think" and "thinking" is mainly a matter of learning which situations may correctly be labeled, in one's language, as situations of thinking. According to the rules of a given language (such as English or one of its dialects), certain situations involving a human organism are to be called situations of "thinking." Other situations involving a human organism, like situations of ball-bouncing, are not to be called situations of thinking; they are to be called other things instead. So, to learn to use the word "think," one must learn to discriminate some situations involving the human machinery and/or its surroundings from the rest of the

situations involving those elements.

Needless to say, the remarks I have made about "think" and "thinking" can be extended, mutatis mutandis, to other mentalistic words and phrases, like "feel," "want," and "fear"; and also to more specific mentalistic phrases like "thinking of a pear" and "wanting some money."

What does all this have to do with eliminativism? According to one standard line of eliminativist argument, neuroscience has shown (or might eventually show) that there is nothing in neurobiological reality that is much like a mental state. If this happens, the argument goes, we should not believe in mental states. (I condense and simplify a number of different arguments here,[10] but I believe I have captured their gist.) Suppose someone says "I am thinking of a pear." Someone else (an eliminativist) could say "That isn't true. There's nothing real corresponding to what you, in your ignorance, call 'thinking of a pear.' The phrase 'thinking of a pear'—and, for that matter, the word 'thinking' itself—can't find homes in neurophysiology, so you really should give them up."

My answer to the eliminativist runs as follows. Even if the classification of some states as thinking states has no basis in neurobiology, *it still has a basis in physical reality.* At very least, this classification is part of the linguistic practice of human organisms—and that practice is part of physical reality! Regardless of one's views of the ontology of language, the physical utterance of tokens of words and sentences is a process in the physical world.[11] It is as much a part of physical reality as is any other physical phenomenon. The fact that organisms of a particular species are able to respond to certain situations with certain sounds or markings is a genuine physical fact. This fact forms the basis for a real distinction among situations. One cannot sensibly claim that the word "thinking" corresponds to nothing in physical reality. That word picks out a class of states definable in terms of the physically real behaviors of certain physically real organisms. Standard eliminativist arguments cannot get around this fact. At worst, they might be able to show there is no *neurophysiological* basis for the application of the word

"think." But they cannot do away with the fact that there is a *physical* basis for this application—a basis rooted in the physical features of certain easily observable linguistic practices.

One cannot sensibly deny that physical reality picks out a class of situations of thinking. In like manner, physical reality picks out a class of situations of thinking about a pear. It also picks out a class of situations of feeling happy, a class of situations of wanting money, and so forth. Physical reality manages to pick out these classes of situations—and it does so *regardless* of the facts of neurophysiology or the alleged limitations of folk psychology.

Since all these kinds of situations are firmly rooted in physical reality, it follows that we are correct in speaking as though situations of these kinds really existed. (We can speak as if they really existed because they do really exist.) This implies that ascriptions of mental states to humans, made in the customary fashion, normally are correct. We can easily convert talk about mental situations into talk about mental states: for X a human organism, X is in a state of thinking if and only if there is a situation of X thinking. This is not to say that these mental states have all the powers that folk psychology attributes to them. I will take up that question later.

The fact that *people sometimes think* is true largely, though not solely, because of the way that the word "think" is ordinarily used. This is not the only condition for the truth of that fact, but it is an important one. Given certain empirical facts about the physical nature and behavior of humans, we can deduce, by considering the standard usage of "think," that it is correct to assert that people sometimes think. It would be wrong to conclude that people don't think just because it turned out that there is nothing in neuroscience corresponding to thinking. The set of mental situations of a given kind, such as situations of an organism's thinking of a pear, might not be a neat set of situations involving the activity of brains. Instead, it might be a very disjunctive set of such situations, having little in common except that they are picked out verbally in the way I have described.[12] Alternatively, mental situations could be situations that are *not*

confined to the brain. Externalism goes in this direction, as do the sociocultural accounts of folk psychology proposed by Margolis[13] and McDonough[14]. All these ideas remind us that we should not uncritically picture the domain of mental states as a neat, clean, easily definable set of brain states. As long as we can pick out, by observable physical means, the states that constitute thinking states, then it is perfectly acceptable to use predicates like "is thinking" to describe people. It may well be that such predicates are of no value to brain science, but that's the worst we can say about them. "Thinking" may not be a useful term for physiology, but certainly it's a good term for some other purposes. The word "thinking" does correspond to something in physical reality, though this "something" has more to do with socially conditioned organismic behaviors than directly with neural states.[15] In this sense, the word "thinking" is truthful. We should not feel any imperative to give up this word just because thinking, when analyzed neurobiologically, doesn't fall apart along the lines that neuroscientists might want it to. This, of course, goes not only for "thinking," but also for "feeling," and for "wanting a rose," and for other mental terms. These terms apply to situations involving the human apparatus. Mental terms are applicable to these situations by virtue of the *physical* facts about how the mental terms are used.

3. Some Philosophical Precedents

The above claim about the application of mental terms is close to a number of earlier arguments about folk psychology. It comes quite close to an important pair of anti-eliminativist arguments by Greenwood.[16] Greenwood argues that facts about the causal roles of intentional states should not make us throw out our beliefs about the existence of such states.[17] He points out that there is evidence for the existence of such states, quite apart from our beliefs about their causal roles. This evidence comes from self-knowledge and communication, and does not stand or fall with beliefs about intentional states' causal powers.[18] Greenwood also reminds us that a child can learn to recognize states of thinking, etc. without holding any theoretical beliefs about the causal roles of such states.[19] Thus, according to Greenwood's view, we may safely suppose that intentional

states exist, even if our beliefs about the causal powers of those states are wrong.

The main difference between my argument and Greenwood's is that I am trying to do less. My argument says little about intentional states in general, or about our knowledge that a state is intentional or representational. My argument shows only that we can be sure of the reality of *particular kinds* of *putatively* intentional states—states that philosophers normally classify as "intentional." There is a nonempty class of states normally called "states of thinking," another nonempty class of states normally called "states of feeling," and so forth. These classes of states are firmly grounded in physical reality, regardless of what neuroscientists might discover. In Greenwood's arguments, representation plays an important role; he suggests that "our theoretical classificatory descriptions of human action" could have been wrong if we had lacked "empirical evidence for the intentional direction of human actions."[20] On my account, we could preserve those classifications with even less evidence than that. It is enough that human organisms are able to respond behaviorally to the states in the way that they presently do. This capacity is enough to ensure that the phenomena of thinking, feeling, etc. are grounded in physical reality.

One also can think of the account proposed here as a stripped-down, minimalist version of the view that folk psychology is culturally grounded and hence does not need the support of neuroscience or cognitive science. Margolis[21] and McDonough[22] have proposed accounts of this latter sort for folk psychology. These two accounts are of great interest, and (I believe) are compatible with my approach. However, my view appears to have an added strength: it does not depend on specific understandings of, or detailed arguments about, the relation of folk psychology to culture (as do the views of Margolis and McDonough). Instead, my view depends mainly on certain general observations about how individual humans use mental words. My approach also has an added *weakness* compared to these earlier approaches: my argument does not address the preservation of folk psychology as a whole, but only the preservation of a fragment of folk psychology. The fragment in question consists of attributions of mental states. (Later in this paper I will extend this fragment, but even then it will not encompass all of

folk psychology.) Since language is a cultural practice, my suggestion is a version of the view that folk psychology is culturally grounded. However, my view may be more robust than other ideas of this kind, since it depends less on facts or concepts about culture and more on physical facts.

My claim also comes close to Horgan and Graham's ideas about the "austere" character of folk psychological commitments[23]. According to my view, folk psychology commits us to very little besides the existence of certain obviously real phenomena. My argument does not use the conceptual apparatus of Graham and Horgan, with its classification of theses as "austere" and "opulent." Nevertheless, my proposal overlaps the approach of Horgan and Graham in a key respect. Like them, I have assigned *linguistic competence* the key role in grounding the truth (or at least the warrant) for folk psychological propositions. In Graham and Horgan's account, linguistic facts form the fundamental piece of evidence for the truth of folk psychology. In my account, only the simpler physical aspects of linguistic practice are crucial; these give us confidence only in a part of folk psychology, by grounding that part in the physical. According to my account, the truth of folk psychology rests on the ways in which human linguistic practices are embedded in a physical world.

It would be interesting to make a detailed comparison of my argument with the ideas of Horgan and Graham. One could regard my proposal as a claim about the extreme austerity of folk psychological concepts. On the other hand, one could regard my proposal simply as a suggestion that we can back off from most of the commitments of folk psychology without losing what is most central to folk psychological knowledge.

4. Mental Causation and the Meaning of "Cause"

Folk psychology does not consist of mental state attributions alone. Another important part consists of propositions about causation by mental states. Are these defeasible by neuroscience or cognitive science, as eliminativists often claim? I will argue that some of

these propositions seem less defeasible once one gives them a slightly more charitable reading.

Consider the proposition that my subjective impression of the color red *causes* me to feel excited, and that my desire for a rose *causes* me to seek a rose. Read naively, these propositions seem to say that mental contents (or states) are literally and simply causing other mental contents (or states). Thus, the statement I just made about red might plunge you into the middle of the debate about the causal role of qualia. My statement about desire for a rose might plunge you into the debate about the reality of propositional attitudes.

This paper is not the place to review the known accounts of mental causation. Instead, I will make a suggestion that (I believe) is somewhat orthogonal to the traditional debates about this topic. I suggest that statements about mental causation are more ambiguous than we usually realize. Specifically, I suggest that the meaning of the word "cause" when that word is applied to mental phenomena may not be quite the same as the meaning of the word "cause" when that word is applied to simple physical phenomena.

Sometimes a familiar word turns out, unexpectedly and surprisingly, to have had two incompatible usages all along. In these cases, the best way to understand the incompatibility is to assume that the word has two slightly different senses. To use an old example from physics, people often use the word "heavy" and its derivatives in two incompatible ways in different contexts. Compare the sentence "This ten-pound dumbbell is heavier than this five-pound dumbbell" with the sentence "Gold is heavier than water." The conflict between these two usages becomes evident when the user is faced with certain puzzles, such as whether a gram of gold is heavier than a gram of water. When we learn introductory physics, we learn that the word "heavy" is best understood as having two meanings in these two contexts—closely related meanings perhaps, but different ones. It turns out that "heavy" is equivocal between two meanings. "Heavy" means "having a large weight" when used in weight contexts; it means "having a high density" when used in density contexts. But it takes some reflection, or at least

some new knowledge, to figure this out. This is a case in which people use a word in different contexts, with slightly different senses or slightly different extensions, and don't really think about it. They just do it.

Perhaps this happens with the word "cause" too. People say "my desire for money caused me to do this"; they also say "the impact of the cue ball caused the eight ball to move." Maybe if they learned more, thought carefully, and reflected deeply on physical and mental cause and effect, they would end up saying something like this: "When I said my desire for money *caused* me to act, I didn't mean quite the same thing as when I said the impact of the cue ball *caused* the eight ball to move. I didn't realize it before, but maybe I am using 'cause' in two slightly different ways."

People sometimes use the same word in two different, though related, senses. If the two senses are sufficiently similar or entangled, people may do this without even knowing it. The sameness of words sometimes may deceive people into making false assumptions about the sameness of things. But the fact that the things aren't really the same doesn't give one grounds for throwing out the words. It only acts as a reminder that one must be careful with words. (As if philosophers didn't already know that!) When we use "cause" in the context of talk about mental states, perhaps we are not using it in precisely the same sense as when we use it in regard to physical things. If we are using "cause" in mental and in physical contexts, and we think it has the same sense in both cases, then perhaps we are a bit confused—just as we would be if we had learned a word for the first time and didn't quite know how to apply it in some cases. Equating mental causality to physical causality may be a mistake, but if it is, then it is an understandable mistake. The mistake arises from equating the meaning of the word "cause" in mental contexts with the meaning of the word "cause" in physical contexts.

Someone might try to rebut this by saying "But the two instances of 'cause' *do* mean the same thing! Nobody draws that distinction of meaning when they talk about mental states 'causing' things. They just mean what they normally mean by 'causing.' The second meaning of 'cause' is your invention alone. Therefore, causation by a mental

state is the same phenomenon as causation of one billiard ball movement by another." My reply to this rebuttal is as follows: If people really were using "cause" in slightly different senses in mental and in simple physical contexts, would they inevitably *know* that they were doing so? As I just pointed out, people sometimes use words in slightly different senses without even noticing it.

If "cause" is ambiguous, we cannot say for certain that the simple physical meaning is primary or paradigmatic. If the word really has two legitimate senses (as does "heavy" in the density example), then neither sense is a strained or quotation-marked sense. But even if one sense is privileged, then the simple physical sense is not necessarily the privileged sense. For all we know, the simple physical sense of "cause" might be demonstrably *non*-paradigmatic. Perhaps the concept of mental causation, which lies so close to our own experiences, somehow underlies or permeates all our ideas about causality. Perhaps learning about mental causation helps to set the stage for learning about other kinds of causation—including the billiard-ball kind, which is more alien to the observer. Note that I said "for all we know"; I am not claiming to know whether the last two sentences, with the "perhaps" removed, are true. But in any case, there is no conclusive reason to put one sense of "cause" above the other and to claim that one sense is more standard or correct than the other. (At least there is no reason for this outside the psychology clinic or the physics lab, where special jargons prevail and "cause" may well not have quite its usual richness of meaning.)

Is "cause" really ambiguous in the way I have suggested? As a matter of observable fact, people use the word "cause" to refer to relationships among mental states and also to relationships among obviously physical states of matter. The usages of "cause" in these two contexts are not obviously identical; if they were, there would be far fewer philosophical puzzles about the relationship between what philosophers call "mental causation" and what physical scientists call "causation." Thus, for prephilosophical language, there is little doubt about the double usage of "cause." One cannot get around this by claiming that "cause," as used in physical science, has only one sense. That sense of "cause" amounts to a term of art particular to the physical sciences. It may well not be

the same as the prephilosophical meaning of "cause," or as the sense of "cause" when that word is used in psychological contexts. Psychologists should be interested in whether mental states "cause" each other in the full, uncut, unsimplified sense of the word "cause." They should not be equally concerned about whether a term of art from physics happens to apply to their subjects' thoughts and feelings.

If "cause" is ambiguous in the way I have suggested, then we have no grounds for a blanket denial of claims that mental states (or situations) can play causal roles. Such a denial may even begin to appear a bit extravagant.

Perhaps mental causation is very different from the causation that happens when billiard balls bump. Maybe it can even have a different time ordering. Maybe, as some well-known experiments suggest, an action begins before we are conscious of the decision to act.[24] But this peculiarity of timing should not be too surprising, for time always is measured with clocks that make use of the *other* kind of causation—the simple physical kind. If mental state A causes mental state B, does A have to "cause" B in the physical sense? Perhaps not. Perhaps physical causation is not crucial to mental causation; perhaps a certain commonality of information between mental states, or some other relationship (functionalistic?) between states, is more important. Thus, for all we know, the time ordering of A and B might not be too important for mental causation. (Needless to say, this suggestion does *not* involve reverse causation in the physical sense.)

Once we recognize that "cause" is ambiguous, we can preserve most or all of the part of folk psychology that deals with mental causation. We can do this while leaving open the questions of the reducibility of mental causation in neuroscience, cognitive science, or physics. We can now accept a fragment of folk psychology, even if we do not have answers to many of the ontological questions.

The upshot is that many folk psychological beliefs about mental causation come out true if you give those beliefs a *somewhat charitable reading*, by recognizing that "cause" means something a little different in mental as versus simple physical contexts. All we

need to do is let "cause" have its standard meaning, instead of one of its jargon meanings. We should not assume uncritically that "cause," when used in mental contexts, is a word borrowed from freshman physics with no change in meaning. We can admit that mental causation is a relationship of the kind that one actually finds in psychology, instead of a relationship of the kind that one finds on the billiard table. Once we admit this, many folk psychological beliefs about mental causation simply come out true.

5. Concluding Remarks

In this paper I have suggested that some propositions of folk psychology are true mostly by virtue of the way mental terms are used in natural languages. Propositions about the existence of mental states, such as "I am thinking," often come out true because the human organism can learn to tag certain physical situations in certain systematic ways. Propositions about mental causation often come out true because of the way in which the usage of "cause" accommodates both mental and physical contexts.

We do not know whether these findings will let us preserve folk psychology as a whole. However, they do preserve a key fragment of folk psychology. This fragment, I suggest, is vitally important to our picture of ourselves as persons. It contains the crucial propositions that people have mental states, and that mental states sometimes are causes. This fragment is the vital core of folk psychology—and this core is true largely (though not entirely) because of the way mental language is used within a physical world. Hence future discoveries in cognitive science and in neuroscience will not refute this core, nor will standard lines of philosophical argument erode it. This finding, though not a defense of folk psychology as a whole, is enough to preserve what is most human in us from present and future critiques by eliminative materialists.

REFERENCES

Churchland (1991). Paul M. Churchland, "Folk psychology and the explanation of human behavior". In Greenwood (1991a).

Davidson (1995). Donald Davidson, "Laws and cause". *Dialectica*, **49** (1995).

Dennett (1991). Daniel C. Dennett, *Consciousness Explained*. Boston: Little, Brown and Co., 1991.

Graham and Horgan (1994). George Graham and Terry Horgan, "Southern fundamentalism and the end of philosophy". *Philosophical Issues* **5**, 1994.

Greenwood (1991). John D. Greenwood, "Reasons to believe". In Greenwood (1991a).

Greenwood (1991a). John D. Greenwood (ed.), *The Future of Folk Psychology: Intentionality and Cognitive Science.* Cambridge: Cambridge University Press.

Horgan and Graham (1991). Terence Horgan and George Graham, "In defense of southern fundamentalism". *Philosophical Studies* **62**, 1991.

Horgan and Woodward (1985). Terence Horgan and James Woodward, "Folk psychology is here to stay". In Greenwood (1991a). (Reprinted from *The Philosophical Review* **94**, 1985.)

Margolis (1991). Joseph Margolis, "The autonomy of folk psychology". In Greenwood (1991a).

McDonough (1991). Richard McDonough, "A culturalist account of folk psychology". In Greenwood (1991a).

Ramsey et al (1991). William Ramsey, Stephen Stich, and Joseph Garon, "Connectionism, eliminativism, and the future of folk psychology". In Greenwood (1991a).

NOTES

[1] See, for example, Churchland (1991).

[2] This is how I read Horgan and Graham (1991) and McDonough (1991).

[3] In alphabetical order: Greenwood (1991); Graham and Horgan (1994); Horgan and Graham (1991); Margolis (1991); McDonough (1991).

[4] Greenwood (1991), p. 70.

[5] See Greenwood (1991), especially pp. 73-75.

[6] The precedent is this: Greenwood (1991, pp. 80-83) points out that a child can learn to use words for mental states without understanding the supposed causal roles of intentional states.

[7] Greenwood (1991) touches on some of these issues; see especially pp. 80-83.

[8] But see Greenwood (1991, p. 83) for a likely counterexample involving shame.

[9] Greenwood (1991), p. 82.

[10] For example, Churchland (1991); Ramsey et al. (1991).

[11] This does not exclude the possibility, supported by Margolis (1991; see especially p. 245) and McDonough (1991), that human culture is not reducible to the physical sciences.

[12] This statement about disjunctive states brings to mind Davidson's anomalous monism (see especially Davidson (1995)) and the remarks about "gerrymandered" structures and events in Horgan and Woodward (1985), sections 3 and 4. Comparing these three sets of ideas might prove fruitful. What are the differences and similarities?

[13] Margolis (1991).

[14] McDonough (1991).

[15] Once again I am in the vicinity of Margolis' and McDonough's claims about the cultural nature of folk psychology (Margolis (1991); McDonough (1991)). Those accounts certainly are compatible with, though not entailed by, what I am saying in this paper.

[16] Greenwood (1991).

[17] Greenwood (1991), especially pp. 75-81.

[18] Greenwood (1991), especially pp. 81-87.

[19] Greenwood (1991), pp. 80-83.

[20] Greenwood (1991), p. 74.

[21] Margolis (1991).

[22] McDonough (1991).

[23] Horgan and Graham (1991), Graham and Horgan (1994).

[24] I am thinking, of course, of the Libet experiments and other related findings (see Dennett (1991), chapter 6, for relevant information).

Yes, We Have Conscious Will

Mark F. Sharlow

ABSTRACT

In this paper I examine Daniel M. Wegner's line of argument against the causal efficacy of conscious will, as presented in Wegner's book *The Illusion of Conscious Will* (Cambridge, MA: The MIT Press, 2002). I argue that most of the evidence adduced in the book can be interpreted in ways that do not threaten the efficacy of conscious will. Also, I argue that Wegner's view of conscious will is not an empirical thesis, and that certain views of consciousness and the self are immune to Wegner's line of argument.

Introduction

In this paper I will assess Daniel M. Wegner's line of argument against the causal efficacy of conscious will, as presented in his book *The Illusion of Conscious Will* (hereafter cited as ICW).[1] In sections 1-4 of the paper I will examine the nature of Wegner's thesis about the illusory character of conscious will. While doing this I will explore some concepts and terms used in his argument. In sections 5-10 I will show that much of the evidence Wegner uses can be interpreted in ways that do not support his conclusions. Also, I will suggest that some of Wegner's interpretations of the evidence

beg important philosophical questions. In section 11 I will point out some views of self and of consciousness that appear to be immune to Wegner's argument against conscious will. Section 12 contains some concluding remarks.

In composing this reply to Wegner, I drew on the work of many other authors, including earlier critics of ICW. In some cases I have adopted these critics' arguments intact or nearly so. All of these debts are acknowledged in the text or in the endnotes.

1. Wegner's Determinism vs. Ordinary Causal Determinism

Before I begin, let me try to situate Wegner's thesis within the overall free will vs. determinism debate.

According to Wegner, conscious will is to be regarded as a feeling (ICW pp. 1-28, especially p. 3). As the title of his book suggests, Wegner argues that this feeling is an illusion. By this he means that "*the experience of consciously willing an action is not a direct indication that the conscious thought has caused the action*" (ICW, p. 2; italics in original). In other words, conscious will is only a feeling that makes it seem to us that we are consciously originating our actions. According to this view, the real sources of our actions are unconscious, and often have little to do with our conscious reasons for action (ICW, especially pp. 26-28).

Wegner's critique of conscious will is deterministic, but it goes beyond the physical determinism that philosophers traditionally view as a threat to free will. Many philosophers today are *compatibilists*: they hold that free will could exist even if the universe were governed by causal determinism. (For the record, I am a compatibilist.[2]) The paradigmatic image of the causal determination of our wills is, perhaps, Laplace's famous claim that a being with complete knowledge of the present state of the universe also could know the entire future (Laplace 1902, p. 4)[3]. Compatibilists typically hold that this kind of causal determination would not rule out our actions being free, on some

understanding of what it means to be free. But Wegner goes farther than Laplace. Wegner's critique does not merely imply that our conscious choices ultimately are determined by previous causes. Instead, it implies that we do not really make consciously willed choices, in any objective sense, at all. Many of us hold that determinism of the Laplacian sort is not a real threat to our status as conscious doers. This assessment does not carry over to Wegner's deterministic thesis. If Wegner is right, then we are not merely predictable conscious doers, as in Laplace's scenario. Instead, we are not conscious doers in any authentic sense at all. What is more, Wegner's view implies that we are not even clear-headed enough to recognize that fact.[4] Thus, Wegner's determinism strikes at the heart of the concept of a person in a way that physical determinism alone does not.

Wegner speaks of his argument as a way of combining conscious will and determinism (ICW, pp. 2, 26). However, he argues elsewhere (pp. 318-325) that the problem of free will vs. determinism, at least in its traditional form, is misconceived. Wegner even dismisses the philosophical literature on this problem as "shocking in its inconclusiveness" (ICW, p. 26). I would say that Wegner has not managed to combine *real* conscious will with determinism (see section 2 below), and that the traditional problem of freedom vs. determinism remains as important as ever. In any case, Wegner's deterministic view goes beyond the usual parameters of the freedom-determinism debate by portraying human action, not only as constrained and predictable, but as (one might say) puppetlike. Saying that our conscious will is an *illusion* is different from merely saying that our conscious will is *predictable*. Wegner's determinism is not the only possible version of determinism. One does not need to believe Wegner's version to be a determinist.

2. Is "Illusion" the Right Word?

Next I will make a few remarks about Wegner's claim that conscious will is an illusion. I am not the first to wonder whether the word "illusion" really fits. Other authors,

including Wegner himself, have cast doubt upon the suitability of this word. According to Wegner, certain other words, including "construction," would be as good as "illusion" to describe the nature of conscious will (ICW, p. 2 footnote).[5] Other authors (for example Heyman (2004), Jack and Robbins (2004), Ainslie (2004)) have suggested, in various ways, that the concept of illusion does not fit well with Wegner's evidence about conscious will. Here I will state my own view on this topic—a view that agrees with or overlaps that of several earlier authors.

Reread the quote from Wegner near the beginning of Section 1 in this paper. Think carefully about that statement. If the experience of conscious will does not tell us directly about the causation of our actions, then the experience of conscious will is not what we sometimes think it is. But is it an *illusion*?

Wegner admits that conscious will is more than just a mistake. He points out that the feeling of conscious will often accurately indicates mental cause and effect (ICW, pp. 15, 327), and that this feeling can help people become more effective in their actions (ICW, chapter 9). He even calls this feeling "the mind's compass." (ICW, p. 317) In reading these parts of the book, one gets the impression that conscious will is more like a half-accurate perception than like an illusion. The feeling of conscious will might not be direct awareness of a causal relationship. But does that really matter, if it is good enough *indirect* evidence? (See Ainslie (2004) for ideas relevant to this last point.)

The word "illusion," as used in ICW, bears a heavy rhetorical and ideological load. Wegner has fully acknowledged this fact (ICW, p. 2 note; 2004b, p. 682). Perhaps the heaviest part of this load is a strong suggestion of unreality—a suggestion that goes beyond the mere claim that the feeling of conscious will is a fallible and indirect indicator of the truth. The claim that conscious will is *illusory* is a stronger expression of skepticism than is the alternative claim that conscious will is less powerful and important than we usually think it is.

Wegner's definition of "conscious will" raises other questions. Early in his book (ICW,

p. 3), Wegner claims that conscious will is a kind of feeling. As Hardcastle (2004) has pointed out, this seems incorrect, for the experience of a thing is not the thing experienced. This is an extremely important point. Consider this example (which is not Hardcastle's): One could say "I willed that I would get up from the chair, and I felt that I was willing it." This sentence might not occur in ordinary speech outside of discussions about will and willpower, but still it reflects the standard usage of the word "will." The feeling of conscious will is a *feeling* that one is willing something. Having this feeling is not the same as willing something. Wegner acknowledges this seeming discrepancy (ICW, p. 3), and uses a Humean argument to equate conscious will to the feeling of conscious will (ICW, pp. 3, 13-14).[6] However, an abuse of language still is an abuse of language, even if the name of Hume can be invoked in its favor.

Other authors have raised objections in the same vein. Ainslie (2004) has questioned the identification of conscious will with what Wegner (ICW, p. 317) calls "the mind's compass." Jack and Robbins (2004) have pointed out the difference between will and the experience of will.

The distinction between conscious will as a feeling, and some other kind of will, is not foreign to ICW. Wegner introduces the notion of "empirical will" (ICW, p. 14), and claims that this is real and causally efficacious, and that the feeling of conscious will sometimes (but not always) reflects this empirical will (ICW, pp. 15, 327).

3. Confabulations or Historical Reconstructions?

Wegner suggests that the explanations we give for our actions often are confabulations—that is, fictional stories manufactured in the brain (ICW, pp. 171-184). We have to be very cautious about this claim, for the following reason.

First, a confabulation may not always be just a made-up story. It also can be a reconstruction of past events, based on indirect evidence. Suppose that I drink a glass of

water, and then tell a story about the origin of my action ("I thought it would benefit my health"). Suppose neuroscience shows that my conscious thought played no part in the immediate causation of the action, but appeared after the fact, in the wake of the action. Do these facts alone imply that my story is a mere fiction? They do not! For all we know, the confabulation might be a fairly accurate historical reconstruction of what happened in my brain. Perhaps my brain monitored my current behavior, together with past circumstances that predisposed me to behave that way (like my past health worries and my drinking of water in connection with these worries), and then fabricated a fairly good guess about what led up to my action. This indirect way of knowing why I did things is not infallible, but it may be good enough for many purposes. Dismissing it as mere "confabulation" seems rather silly. To call the story a "confabulation" instead of a "historical reconstruction" is to beg the question of the reliability of the story.

Human observers often reconstruct recent past events in their external surroundings in much this way. Often we do this intuitively and very quickly, without any apparent reasoning. ("I saw broken glass and tire marks in the street, so I knew there had been a car accident." "I heard the cry of a baby from next door, so I knew the neighbors had finally had their baby." "I saw ketchup on the ceiling, so I knew my nephew Boris was visiting again.") Such impromptu explanations, based on memories or other traces of past events, often are remarkably accurate.[7] These explanations are stories based partly on guesses—but we should not demean these stories by calling them "confabulations." These stories are nothing less than historical reconstructions of an informal kind. Why couldn't our brains do the same thing to explain our actions? Perhaps our brains are natural born amateur historians. (In view of their evolutionary history, wouldn't they have to be?)

By making this point, I am not claiming that our pronouncements about the origins of our actions are infallible or "direct." I am only pointing out that we cannot dismiss these pronouncements as mere confabulations. Perhaps these stories are after-the-fact reconstructions based on incomplete information. However, calling them "confabulations" serves no useful purpose—though this word may have a rhetorical

effect by creating a feeling of the uselessness of conscious will. (For most of us, it is easier to dismiss a "confabulation" than to dismiss a "historical reconstruction"!)

An adherent of ICW might reply that these reconstructions are too inaccurate to be trusted even provisionally. This reply is refuted by the fact that the reconstructions often are accurate, and that we often rely on them without bad results. Nevertheless, ICW is full of examples that seem to support the inaccuracy of our feelings of will. Later in this paper I will defuse many of these examples, by pointing out alternative interpretations that suggest the feeling is more accurate than one might think.

4. What Is an Action?

One pervasive and puzzling feature of Wegner's line of argument is the conception of action that it seems to require. This conception is clearest in Wegner's account of *automatisms*—behaviors that appear, from the outside, to be conscious, but that are not consciously willed by the person having the behavior.

An automatism is a series of movements that appears to be a conscious action, but is not accompanied by a feeling of conscious will. Evidently, Wegner regards automatisms as actions unaccompanied by the feeling of conscious will (ICW, pp. 9, 11). This conception of automatisms lends support to Wegner's general thesis about the disconnection between action and the feeling of conscious will (see especially ICW, pp. 143-144).[8] But there is another way to interpret automatisms: we can simply recognize that an unwilled sequence of movements *is not an action at all*. Philosophers have long recognized that an action-like movement of the human body need not be an action.[9]

The following ugly example illustrates this view. Suppose John has just become brain dead, and some form of artificial electrical stimulation of the peripheral nerves causes his arm to make the movements ordinarily called "reaching for a glass of water." Suppose that these movements look very lifelike; the overall motion is not some jerky

approximation, but is the "real thing" from a purely mechanical standpoint. Would this be an action? According to the standard prephilosophical usage of the word "action," this would *not* be an action. It would be a sequence of bodily movements, but calling it an action would be an abuse of language. Now roll back time to when John is alive and well. Suppose that he performs a very similar sequence of movements twice: once when wide awake and reaching for a glass of water, and once when deeply anesthetized and under electronic stimulation of the peripheral nerves. Suppose that these two sequences of bodily movements are mechanically identical for all practical purposes, and also are mechanically identical, for all practical purposes, to the movements we saw after John's death. Are all of these sequences of movements *actions*? No, they are not. Only the movement made while John is awake is an action.

This example, by itself, does not push any claims about the feeling of conscious will. It simply points out that our prephilosophical notion of action does not cover just any sequence of movements that happens to look like an action. An action-like sequence of movements, by itself, does not necessarily count as an action. There are other conditions that must be met for movements to be actions. This is not merely a peculiarity of the prephilosophical notion of action. Philosophers of action also have recognized that bodily movements must meet specific conditions to qualify as actions.[10]

A skeptic might say that all this is irrelevant, and that a sequence of movements that looks like an action just *is* an action, period. But then the skeptic would be redefining the word "action" to such an extent that the word no longer corresponds to standard usage. This is just a fallacy of redefinition. When ordinary people worry about whether their actions are freely willed, the "actions" they are worrying about are not actions in the skeptic's sense, but actions in the standard sense. The skeptic also would be begging a host of philosophical questions about the relationship between actions and physical movements. But that topic will have to wait until the next section of the paper.

Viewed in this light, Wegner's use of "action" to encompass things like automatisms is just an abuse of language. One wonders how much this abuse of language influences the

rhetorical pull of Wegner's argument. If you count automatisms as actions from the outset, then how hard can it be to show that actions aren't much more than automatisms? But aside from this linguistic and rhetorical issue, there is a deeper conceptual issue at stake. This has to do with the ontology of actions, and more specifically, with the individuation of actions.

5. The Individuation of Actions

The line of argument in ICW makes much of the idea that our conscious thoughts sometimes do not cause the actions they purport to explain. Wegner cites examples in which people come up with reasons for their actions *after* they act—reasons that seemingly did not exist in their minds before they acted (ICW, pp. 149-151, 171-186). Let us temporarily grant, for the sake of argument, that many or all of the actions we believe are caused by our conscious thoughts are not actually caused by those thoughts, which only come later. Then consider the following typical scenario.

Suppose that I reach for a glass of water. A second or two after I do this, someone asks me why I reached for the glass. I reply that I was thirsty. However, I did not think about being thirsty when I was reaching for the glass; this behavior "just happened." Afterwards, I think that I reached for the glass because I was thirsty.

This would seem to be a perfect case in support of Wegner's thesis. It looks as though my thought played no role in my action—and that I mistakenly believed it did play a role. But think again! Before we accept this easy interpretation, we should look more carefully at the concept of *action*.

Consider the water glass example of two paragraphs ago. Imagine a parallel scenario in which I perform the same bodily movements, but then have a different conscious thought: I think that I reached for the glass because water is good for my health. In this alternate scenario, I have performed an action. But is this action *the same action* as in the first

scenario? One feels intuitively that it is *not quite* the same action.

The problem of the individuation of actions is a recognized philosophical problem.[11] I will not try to review the literature on this topic, nor will I adopt any particular account of the individuation of actions.[12] Instead, I will show that the very existence of this problem raises serious doubts about some of Wegner's claims concerning human action.

Consider these three actions:

(1) my drinking a glass of water when I am thirsty

(2) my drinking a glass of water when I am not thirsty, but have long believed in the health benefits of drinking lots of water

(3) my drinking a glass of water when I am not thirsty, but am about to go hiking in the desert

These three actions may involve sequences of bodily movements that are, for all intents and purposes, the same. However, these three sequences of movements "fit in" with my past, present, and future history in different ways. The first action is a satiation of thirst. It coheres with the biological fact that I am now slightly dehydrated. The second action is an act of hygiene. It coheres with my previous thoughts, worries, and doctor visits in a way that the first action does not. The third action is an act of preparation. It coheres with my projected future behavior: because I am about to go into the desert, the act of drinking the water is not a mere movement, but is a safety measure. Biologically, it is a strongly survival-positive act.

Are these three actions really just exact copies of each other? If I did (2) instead of (1), would I be doing the same action that I otherwise would have done? What if I did (3) instead of (2)?

I am not proposing answers to these last two questions. I am merely pointing out that the

answers are not immediately obvious.

Interestingly, the third action might be the action that it is, not only because of my beliefs about my future, but because of the way my future really will be. If I were merely under a delusion that I was about to go hiking in the desert, would drinking the water be the same action described in (3)? Philosophers have long considered that actions or events might be individuated by their effects[13]—and, of course, these effects are in the *future* of the action or event. Thus, we should not rule out offhand the possibility that future circumstances make present actions the actions that they are. (Needless to say, there is nothing truly mysterious about this, and there is no hint of reverse causation.)

By giving these examples, I am not trying to show conclusively that actions can be individuated by the circumstances mentioned in the examples. Also, I am not going to defend any particular account of individuation here. My point is that it is *not* blatantly obvious that actions are *not* individuated by such circumstances. We are not entitled to dismiss this possibility out of hand. The question "Which past, present, or future circumstances make an action what it is?" is a question that cannot be answered off the top of one's head. Philosophical reasoning is required. Philosophers have devoted serious effort to this nontrivial question.

Actions might be individuated by circumstances besides the bodily movements involved in the actions. The situations that precede an action might play roles in the individuation of the action. The situations that *follow* an action also might play roles in the individuation of the action. This last point is especially cogent for the effects of an action.

What does all this have to do with Wegner's arguments?

Let us go back to the first water glass example near the beginning of this section. When I pick up the glass, I do not yet have a conscious thought of my reason for reaching for the glass. Later I have such a thought. The thought comes later than the action, so seemingly

it can play no role in the making of the action. But does this really follow? We know that the thought played no role in *causing* the action. However, we have not ruled out the possibility that the thought plays a part in the *individuation* of the action. After all, an action's effects can help to individuate the action—or at least we cannot dismiss offhand the possibility that they do. Perhaps the thought about being thirsty helps to individuate the action in this way. Perhaps if the thought had not occurred, the movement would not have been the action that it is. Then it would be true to say that the conscious thought, though causally irrelevant to the *movement*, is necessary for the occurrence of the *action*. Without the conscious thought, the same movement would have occurred—but the movement would not have been *that* action. The movement would have occurred; the action, as it actually did occur, would not have occurred.

The upshot is that our actions may depend upon our thoughts, even if the thoughts do not cause the actions. Our conscious thoughts can play roles in our actions, not only by causing physical movements, but by helping to individuate the actions—by making an individual action what it is. In effect, conscious thoughts can transform bodily movements into actions. The relationship between an action and the thought explaining it might not be a causal relationship, with the thought causing the action. Instead, it might be a logical and conceptual relationship grounded in individuation.

Note that I when I said "transform bodily movements into actions," I was not speaking of a fictitious or illusory transformation. An adherent of ICW might be able to live with that phrase if I added a disclaimer like this: " 'Transform bodily movements into actions' really means 'make the brain interpret bodily movements as actions.'" But I will not add such a disclaimer, for that is not what I meant. I was speaking of the *real* individuation of *real* actions. In the scenario I described, conscious thought does not only make the physical movement seem like an action—*it really makes the physical movement into an action.* The thought's occurrence insures that the movement belongs to a different ontological category, and hence is a different kind of item, which the movement would not be if the thought did not happen.

Needless to say, one cannot read the word "makes" naively here. The thought might not exert any causal influence on the physical movement, or on the action. But still, the conscious thought insures that the movement is an action. The thought "makes" the movement into an action in a logical and ontological sense of "makes." It "makes" the movement into an action in roughly the same way that being human, adult, male and never-married at the same time makes one a bachelor.[14]

Again, I should stress that I am not defending any particular account of the individuation of actions. My suggestions about individuation by conscious thought might turn out to be correct, or might need revision. But the mere existence of open questions about individuation of actions casts serious doubt upon Wegner's argument. If certain views of individuation turn out to be right, we might be correct in believing that our actions cannot occur without the conscious thoughts that seem to explain them. We might be correct even if the thoughts come after the actions. Perhaps if you did not have the thought, the bodily movement you made would not be the action that it is. *That* action would not have existed. In its place would be some other action—or perhaps only an automatism—involving the same sequence of bodily motions as *that* action.

This argument about the individuation of action gives us two separate ways to undermine Wegner's thesis.

First, this argument suggests that the feeling of conscious will could be a good indicator of real doing, even if that feeling fails to trace the causal origins of actions. Suppose that the after-the-fact conscious thought, which seems to explain an action, really plays a role in the individuation of that action. If the presence of such a thought is what makes a mere movement into an action, then we are quite right in feeling that the thought "adds something" to the action, and even makes the action what it is. In this case, the feeling of conscious will is trustworthy. If you feel certain that your conscious reason for acting really explains your action, then your conscious thought is in fact responsible for that action's existence. This is the case even if your conscious thoughts do not *cause* the actions they describe.

The second undermining argument is similar, except that the feeling of conscious will takes the place of a conscious thought. It could be the case that the feeling of conscious will itself helps to individuate the actions that it accompanies. If this is the case, then the feeling of conscious will might be a very good indicator of the presence of real action.

One can reach similar conclusions without considering the individuation of actions, by noting that the feeling of conscious will can be *logically necessary* for the action to occur as it does, without actually being the cause of the action. See Krueger (2004) for discussion of a possibility of this general sort. This is a third way to undermine Wegner's argument.

In this paper, I will not try to show conclusively that any of the above three undermining arguments is right. I am only pointing out that if any of them were right, Wegner's argument for the illusoriness of conscious will would be in trouble. The existence of this open question about individuation leaves an opening for accounts of individuation that undermine Wegner's argument. This further implies that Wegner's argument depends implicitly upon ignoring the possibility that certain accounts of individuation are true. However, the truth of these accounts is a philosophical problem, not a scientific one. Thus, Wegner's treatment of action actually depends on a strong nonempirical philosophical commitment. *A fortiori*, Wegner's argument for the illusoriness of conscious will is not entirely a scientific argument.

The existence of open questions about individuation of actions also casts doubt upon the the concept of unwilled action. Once we have admitted that the circumstances surrounding a sequence of movements can individuate an action, we have opened the door to the possibility that a movement physically resembling an action might not be an action at all. As I stated earlier, Wegner's treatment of automatisms as actions involves an abuse of language. The study of individuation of actions shows that the difficulty is not merely linguistic. We cannot freely assume that so-called unwilled actions really are actions—for there may be action-like sequence of movements that are not actions. If

such a sequence takes place under appropriate circumstances (like automatism, hypnosis, or artificial brain stimulation, all of which are discussed in ICW[15]), then it might not be an action at all. If there were no genuine unwilled actions, Wegner's view of the separation of action and conscious will would be considerably less plausible.

6. Individuation of Actions and Hypnotic Suggestion

Wegner uses cases of hypnotic suggestion as evidence for the separation between action and the feeling of conscious will. In these cases, the subject of hypnosis performs actions suggested by the hypnotist—yet the subject feels that the actions are his own, and even comes up with reasons why he did them. In one case cited by Wegner, the subject was told to shelve a book that was lying on a table, then later claimed that she did it because the book on the table looked "untidy" (ICW, p. 149). My earlier argument about the individuation of actions suggests a different way to interpret these cases. (As I will point out later, this interpretation has something of a precedent in Ainslie (2004).)

First, note that hypnotic suggestion never is the sole cause of the hypnotic subject's action. Past states of the subject's neural apparatus also causally influence the action. The suggestion can cause nothing without the help of this apparatus, which is laden with capacities and dispositions. The hypnotic suggestion is only one of many causal influences on the final action. The action still originates within the subject.[16] (Those of us who think about these hypnotic suggestion cases may tend to underrate the role of other influences besides the suggestion. These other causes are at least as important, and presumably are more important than the single brief input of a suggestion.)

Next, note that non-hypnotic circumstances could lead the subject to perform the same bodily movements that occurred after hypnosis. Without hypnosis, the subject could have moved the book for many reasons—including the stated reason involving untidiness. There are many sets of possible circumstances that could have led the *un*hypnotized subject to perform the same bodily movements for the same reason that the

hypnotized subject gave. For example, if there was a disorderly pile of toys next to the book, the subject might have been more strongly inclined to move the book for tidiness' sake. If the subject had just been to a library, then the subject might have been thinking of neatly shelved books before looking at the table, and might have reacted more vigorously to an out-of-place book—and moved it for tidiness' sake. And so forth. Even in cases involving bizarre actions (like the one discussed in ICW, p. 150), one can make up a story about *reasons*—a story which, if true, would make the subject's bodily movements well-motivated. One can think up enough possible combinations of thoughts, emotions, mischievous impulses, and so forth to show that a wide variety of posthypnotic behaviors could occur under the right non-hypnotic circumstances.

In the hypnotic cases in question, the subject makes certain movements and then claims to have a reason for those movements. Under suitable nonhypnotic circumstances, the subject would have made those same movements for that same reason—but in that case, we would classify the reason as a plausible reason. This implies that the subject has the capability of doing those very movements for that very reason. Before being hypnotized, the subject already had capabilities for doing many different sequences of movements for various reasons. I am speaking here of "good" reasons—that is, reasons that would seem sensible and that would seem to us to justify the bodily motions. Many different "motion-reason pairs" of this sort lie within the capability of the subject.

In ICW's favored interpretation of hypnotic suggestion cases like this (ICW, pp. 149-151), the subject makes the movements for no reason (but with a cause), and then invents a bogus reason. However, there is an alternative interpretation: one could suppose that the process of suggestion does not just cause a movement, but causes the subject to *do an action for a reason*. When the subject does the action, the stated reason really is the reason for the action—but the subject has been caused *to do the action for that reason*. The process of suggestion does not only cause a bodily movement; it also brings a reason to light. It brings a possible "motion-reason pair" into actuality. One could say that the suggestion activates one of the subject's preexisting capabilities for performing a motion for a reason, and causes the subject to exercise that capacity.

An adherent of ICW might object to all this as follows: the subject could not have been doing the action for the stated reason, because the subject was not thinking of the reason before the action occurred. My earlier discussion of the individuation of actions should put this objection to rest. A bodily movement may become an action because of the contents of a thought that happens later in time. Perhaps this is what happens in the hypnotic suggestion cases.

In cases like these, one can view hypnotic suggestion as a process in which the hypnotist causes the subject to perform a series of movements, and also causes the subject to discover one of the possible good reasons for that movement. Perhaps the hypnotist causes this discovery indirectly (by causing the movements first) and even inadvertently—but nevertheless, the subject does manage to find a reason as a result of the suggestion. This reason is sufficient to justify the subject's movements. Because of the way actions are individuated, the reason is a genuine reason for those movements. One cannot say that the reason was entirely made up on the spur of the moment, because the reason had real precedents in the subject's preexisting capabilities to perform actions for reasons.

According to this interpretation, hypnotic suggestion does not represent a failure of conscious will as much as a *disturbance of mental focus.* The hypnotist did not simply control the subject like a marionette. Instead, the hypnotist caused the subject to focus on an already existing possibility for action—a possibility that the subject otherwise might not have noticed. The hypnotist is not a puppetmaster as much as a *magician*—one who misdirects the attention of a relatively passive subject. A magician usually directs one's attention to an external object or event, causing one to overlook the mechanism of the trick. The hypnotist directs one's attention (or perhaps deeper levels of neural processing) to a possible action, causing one to overlook other possible actions and their reasons.

This interpretation of hypnotic suggestion does not imply that the hypnotic subject is

morally responsible for the suggested action. One can argue that the hypnotic subject is in a state of diminished moral responsibility, even though the action is a genuine action and is done for a reason. Perhaps one could argue that the subject is not morally responsible for the action because the subject was directed away from other possible courses of action and did not have a fair chance to choose among them. One could say that hypnotic suggestion sharply reduces a person's ability to choose, but does not eliminate the person's ability to act. (The subject still can choose the details of how to carry out the suggested action.)

This interpretation upends Wegner's use of hypnotic cases as examples of an inaccurate feeling of conscious will. If a feeling of conscious will occurs in these cases,[17] then according to this interpretation, that feeling is accurate. The suggested action is a genuine action. Perhaps it is best to describe these cases as cases of *conscious* will with severely limited *freedom* of will.

Ainslie (2004), commenting on Wegner's work (2004a), pointed out that magnetic brain stimulation can "predispose directly to one alternative" among possible behaviors of the subject (Ainslie, 2004, p. 660). This suggestion about brain stimulation seems quite close to what I just said about hypnosis, though of course hypnosis and brain stimulation are quite different in their mechanics.

7. Split Brain Cases

Wegner points to split brain cases as examples of confabulation (ICW, pp. 181-184). In the most interesting of these cases, the right brain receives a stimulus and initiates behavior; then the left brain (which in most people controls speech) originates an utterance about the reason for the behavior. Sometimes this reason seems to have nothing to do with the original stimulus. In one example (ICW, pp. 182-184), which I summarize fairly closely here, the subject viewed pictures in an experimental setup that insured that each hemisphere received different pictures. Then the subject observed other

pictures in the normal way, and selected pictures pertinent to the pictures in the first set. The right hemisphere was shown a snow scene. Later, the hand controlled by the right hemisphere pointed out a picture of a shovel. The left hemisphere was shown a chicken claw; then the hand controlled by the left hemisphere pointed out a chicken. So far, so good. The trouble is that the subject later claimed to have selected the shovel because of its pertinence to the *chicken claw*, not to the snow scene. The subject is quoted as saying "…you need a shovel to clean out the chicken shed" (ICW, p. 184). This reason, standing alone, sounds sensible enough. The problem is that the hand which pointed out the shovel was not controlled by the left hemisphere—so the left hemisphere's stated reason for pointing out the shovel seemingly could not have been the true reason for that choice.

After reading this example, it is easy to feel that the explanation originating from the left hemisphere must be bogus, on the grounds that the left hemisphere did not cause the action. Indeed, Wegner takes examples like these to be examples of confabulation (ICW, p. 181). It is not hard to see how this supports Wegner's view that conscious will is illusory.

There is another interpretation that does not lend support to Wegner's view. To find this interpretation, we must recognize that the two sides of a split brain are not as separate as we usually think. The right and left hemispheres of a split brain patient do interact; they are causally connected in various ways (see Marks (1980), pp. 17-19, 26-28). The severance of the corpus callosum does not stop all interaction, or all causal connections, between the hemispheres; it only closes off the main channel. There are ongoing interactions between the hemispheres, which occur even when the corpus callosum is severed (see Marks (1980), pp. 17-19, 26-28). Some sort of physical interaction is inevitable as long as the two hemispheres sit side by side in the same living body, bathed in the same fluids, interacting with the same organs. Therefore, not all the neural events causally influencing the action (the choice of the shovel) were in the right hemisphere. The left hemisphere interacted with the right hemisphere during the period when the action was developing. Hence there was a *single physical process*, involving both

hemispheres, that led up to the act of pointing to the shovel. The fact that the left hemisphere was involved only in a marginal way does not change this fact. After all, a loosely connected physical process still is a physical process.

The right hemisphere's unstated reason for the choice (you need a shovel to deal with the snow) was a good fit to the observed behavior. That reason alone would be enough to justify the behavior. But the left hemisphere's stated reason, alone, *also* could fully justify the behavior. Thus, either of these reasons can justify the action originated by the *single* overall brain process. Instead of saying that the right hemisphere had the real reason and the left hemisphere had a fake reason, why not just say that *both* hemispheres had sensible reasons for the action—an action caused by a single physical process involving *both* hemispheres?

On this view, the action occurs, not for one reason, but for two. There is nothing mysterious about an action having multiple reasons. All of us sometimes act for multiple reasons. ("I'm going to eat the yogurt because I'm hungry, and also because it's good for my health.") The split brain case differs from these standard cases in two respects: the subject can talk only about one of the reasons, and one hemisphere does almost all the work of initiating behavior.

Applying our earlier remarks on the individuation of actions, we find that the left hemisphere plays a role in individuating the action of choosing the shovel. Even if the left hemisphere did not make up the reason until after the right hemisphere acted, the left hemisphere's reason still could play a role in making the action into *that* action and not some other. The marginality of the left hemisphere's role in causing the movement does not change this. There is no rule against a single action being caused by events that influence each other only weakly.

According to this view, the feeling of conscious will is not an illusion in this split brain case and others like it. The left hemisphere did participate in the causal origination of the action (albeit marginally), and the left hemisphere helped to make the action what it

finally was.

8. Illusion of Control, or Simple Mistake?

Wegner presents cases of the "illusion of control,"[18] in which we feel we are consciously willing something that we do not in fact control (ICW, pp. 9-11). In one of Wegner's examples (ICW, pp. 9-10), a person feels he is controlling items on the screen of a computer game when actually the joystick is not affecting the screen at all. Wegner claims this is an instance in which the feeling of conscious will exists without real doing (ICW, p. 9).

O'Connor (2005, p. 224) has shown clearly why Wegner's interpretation of these cases is wrong. According to O'Connor's alternative interpretation, cases like the joystick case do not involve false feelings of doing; instead, they involve mistakes about how far the effects of one's actions extend. (Jiggling the joystick is what you are doing. Your feeling of conscious will is correct. However, your belief that things on the screen are affected by your action is mistaken.) O'Connor also argues (2005, p. 224) that one of Wegner's examples involves only belief change, not a real illusion of control. These alternative interpretations weaken Wegner's case by eliminating so-called illusions of control as plausible instances of false conscious will.

Another alternative, a slight variation on O'Connor's idea, would be to say that you can be mistaken about which physical events are *parts* of your action. (Your action at the computer includes your jiggling the joystick, but you only think it includes the movements of items on the screen.) This interpretation is plausible because the spatiotemporal extent of an action can sometimes be hard to determine.[19] This kind of error may be what happens in cases of the movement of phantom limbs, as described in ICW (pp. 40-44). (See Ainslie (2004) for an analysis of phantom limb movements consistent with this view.)

This mistake about parts of actions also may account for what happens when subjects seemingly take other people's movements to be their own doing (as in ICW, pp. 41). Perhaps we could describe these cases as follows: a person performs a real *mental* action, and then mistakenly thinks that a non-mental bodily movement (someone else's) was part of the action. (Wegner recognizes that mental actions are legitimate actions (ICW, p. 44).) Alternatively, we might invoke O'Connor's interpretation in its original form, and say that the person mistakenly believes the bodily movement was an effect of the mental action. Either of these interpretations undermines the claim that there was an illusion of doing. Instead, there were only mistakes about the details.

9. A Note on the *I Spy* Experiment

Another case akin to the "illusion of control" cases is the "*I Spy*" experiment described in ICW (pp. 74-78). (The original reference is Wegner and Wheatley (1999).) This experiment involved a situation in which two persons (a real subject and a confederate of the experimenter) acted together to move one object (a sort of computer mouse). The subject had to report, on a 0-to-100 scale, how much influence his or her own actions had on the resulting events. This experiment showed (among other things) that the subjects could not always tell when they, and not the confederate, were stopping the movement of the mouse. This seemingly showed that a person's feeling of conscious action can be inaccurate. However, there is a simpler interpretation. By consenting to the experiment, the subject already has, in effect, agreed to be a coauthor of a set of physical movements—at least to the extent of helping to move the mouse. Thus, in a sense, the subject is a co-originator of the movements, even when the immediate cause of some of the movements is the activity of another person. On this interpretation, the subject is not mistaken about whether he or she is acting, but is only misestimating the extent of his or her contribution to the action, while performing an action jointly with another person. This is another variant of the misestimation of the extent of one's actions—a phenomenon that I discussed in the previous section.

10. Action Projection, or Two Other Simple Mistakes?

Wegner points to cases of "action projection" as examples of failures to recognize that we are doing something (ICW, chap. 6). In some of the cases he cites (such as the famous "Clever Hans" case), a person influences another organism through unconscious bodily movements, and attributes the resulting action to the other organism. (In the Clever Hans case, the other organism was a horse.) According to Wegner, "the sense of authorship" is disrupted in these cases (ICW, p. 187). However, we can easily find an alternative interpretation of these cases. Earlier I raised the possibility that an action that is not consciously willed is not really an action at all. If that is the case, then the person doing the influencing is *not* performing an action, and has no authorship to lose. Instead, that person is simply undergoing *movements*, not doing an action—and the movements cause the other organism to perform actions (or perhaps just movements). There is no real misattribution of actions.

The preceding argument does not apply to all cases of "action projection," but only to the ones associated with what Wegner calls "*The Inaction Fiction*" (ICW, p. 218; italics in original)—namely, the erroneous belief that one is not doing anything to influence the other subject. In other case of "action projection," the person knows he is doing something to, or with, the other subject (ICW, pp. 218-220). These other cases involve mistakes about the causes, effects, or extent of one's actions, or mistakes about other subjects' actions—but not mistakes about whether one is "doing." The mistakes in these cases are much like the errors involved in the "illusions of control" that I discussed earlier; one misjudges the effects or extent of one's actions. Interpreted this way, these cases of so-called "action projection" do not pose a threat to the belief that one really is doing something.

11. A Larger and Divided Self

The ultimate challenge to Wegner's thesis comes from the possibility that the conscious self may be larger than it seems. By this I mean the following: (1) Much of what we consider unconscious processing might actually be conscious in some sense. (2) Many supposed instances of divided consciousness might not amount to real divisions of the conscious subject. I will explore these two possibilities in turn.

Block has suggested that contents in "the Freudian unconscious" might actually be conscious, provided we understand this consciousness as "phenomenal consciousness" and not as "access consciousness" (Block 1996, p. 457). According to Block's suggestion, such a content might be "experienced" (Block 1996, p. 457), even if the subject cannot know about this experience in the customary way. Elsewhere I have concurred with Block's view; I have suggested that much of what psychology traditionally calls "the unconscious" actually is conscious in the sense that it is associated with a way things seem (Sharlow 2001, pp. 230-234).[20] Perhaps some of the mental or neural processes that we regard as unconscious really are conscious. Perhaps these processes even give rise to full-blown phenomenal experiences, but the subject cannot know about these experiences. Presumably this would be a special case of a known phenomenon: failure of metacognition.[21] (It also might be a special case of what Wegner calls "deep activation" (ICW, pp. 163-164).)

We also must face the possibility that so-called divisions in consciousness are not as divisive as they seem. Split brain cases provide the most dramatic examples of supposed disunity of consciousness—yet one can argue that this disunity is only intermittent and does not affect the unity of the mind itself (Marks, 1980). One even can argue that a split brain patient has a single consciousness at all times—a consciousness that has all of the conscious phenomenology associated with either hemisphere, but which (in a certain sense) does not have all of that phenomenology *together*. I explored this possibility in Sharlow (2001, chaps. 11-12). Using a notion of "subject" that equates a subject to a

single persisting consciousness (pp. 89-111, 215), I suggested that things can seem one way to a subject, and also seem another way to the subject, without it ever seeming to the subject that things are both ways at once. In other words, it seems to the subject's single consciousness that P, and it seems to the same consciousness that Q, but it never seems to that consciousness that P & Q. (This last sentence is very close to Marks' characterization of a *non-unified* consciousness (1980, pp. 13, 17, 39). On my account it is compatible with the unity of consciousness.) I suggested that division of this kind might occur in split brain cases (Sharlow 2001, pp. 266-267), and also in less dramatic cases of disunity, such as repression and compartmentation of belief (pp. 235-242). Further, I argued that none of these apparent disunities can pose any real threat to the unity of the conscious subject (pp. 242-244).[22] In those writings, I did not explore the physical basis of the phenomena of self-division or of inaccessible consciousness. I simply tried to codify their structure using ideas from modal logic. One can describe these cases in more cognitive terms as failures of metacognition.

These conceptions of divided and inaccessible consciousness are of interest in connection with Wegner's argument. Indeed, these ideas completely undermine Wegner's strategy for tracing our actions to unconscious causes. They open up the possibility that the so-called unconscious causes actually are conscious after all. Perhaps the neural events leading up to action, such as the precursor events found in the Libet experiments (Libet 1985), actually are *conscious* events. This last idea has precedents in the work of Holton (2004) and Velmans (2003, 2004), which I will examine and compare below. (There also are other precedents, which I will mention in the notes, but the suggestions by Holton and Velmans seem closest to what I have in mind here.) According to this idea, the precursor events are conscious, but we do not know that we contain them. If this idea were true, it would destroy the view that so-called consciously willed actions really are nonconscious at their cores. The neural processes immediately preceding our actions could be genuinely conscious, and perhaps even accompanied by the phenomenal feel typical of conscious doing. This could be the case even if we do not know of any conscious thought or feeling until later.

Holton (2004) has made a similar suggestion in a review of ICW. More specifically, he suggested that the precursor events might be genuinely mental events of which the subject is not aware until later (2004, pp. 220-221). Holton (2004, pp. 219-221) showed that this possibility weakens Wegner's thesis about conscious will. Holton made this suggestion in the context of higher-order thought models of consciousness, but he pointed out that these models are not necessary for his idea. In my estimation, Holton's argument is an important objection to ICW. Another precedent comes from Velmans, who has argued that the events immediately leading up to conscious actions can be genuine parts of the self (2003, 2004), and in some instances are conscious in certain senses of the word "conscious" (2004). (The 2004 article by Velmans was a response to Wegner (2004a).) I will mention several other precedents in an endnote.[23] What I am suggesting is, perhaps, not quite the same as Velmans' idea. I am suggesting that the precursor events are phenomenally conscious throughout their course, while Velmans suggested that they can be conscious in the sense of being accessible to consciousness at some time. But Velmans has found an important objection to Wegner's argument. Velmans suggests that the self encompasses some unconscious events as well as conscious events, and that the unconscious beginnings of actions in us can be truly our own doing (2003, 2004). I would agree, and I would go further. If Block's suggestion about the unconscious is correct, then perhaps the precursors of action are not only truly ours, but also truly parts of our conscious lives. Perhaps the so-called "confabulated" reasons for some of our actions are simply conscious reasons, indirectly known. And (for all we know) perhaps the feeling of conscious will plays a part in the phenomenal feel of the precursor events—making that feeling one of the wellsprings of action after all. This possibility does not strictly follow from the ideas of Block, Holton, or Velmans, but it is a possibility nonetheless.

This view of the "unconscious" precursor events also has consequences for our attitudes toward human creativity. Wegner points to creative inspiration as an example of action without conscious will (ICW, p. 81-84). According to the view I am presenting here, one's creative inspirations may well be products of conscious processes that are genuinely one's own, but that lie outside what one normally regards as the self. If this is

true, then one's creative productions are truly one's own, even when they arrive by way of a "Eureka!" experience that seems involuntary. They simply come from a place outside of one's *everyday* consciousness.

12. Concluding Remarks

In this paper I have tried to undermine Wegner's argument in two ways. First, I have shown that it is possible to understand most of Wegner's evidence in ways that do not support his view that conscious will is an illusion. Second, I have pointed out that certain views of the self make Wegner's evidence nearly irrelevant to the question of the efficacy of conscious will. In most of my arguments I have not proposed positive accounts of anything. I have simply pointed out that certain possible accounts of things would render Wegner's argument unpersuasive. But that is enough to defuse Wegner's argument.

One lesson we can learn from this study is that the argument about conscious will in ICW is not entirely empirical. That argument depends crucially upon philosophical assumptions—or, more precisely, upon leaving out certain philosophical issues. If we pay closer attention to these issues, we find that Wegner's argument has crucial weak spots.

Neither science nor logic forces us to accept Wegner's pessimistic view of conscious will as presented in ICW. That view is neither an empirical hypothesis nor the conclusion of a persuasive philosophical argument. Instead, it is a curious philosophical position haunted by many unanswered questions. In particular, Wegner's book does not give us convincing grounds to believe that science has debunked the efficacy of conscious will.

References

Ainslie, George (2004). The self is virtual, the will is not illusory. *Behavioral and Brain Sciences*, **27**, 2004, pp. 659-660. (Open Peer Commentary article on Wegner (2004a).)

Block, Ned (1996). How can we find the neural correlate of consciousness? *Trends in Neurosciences*, **19**, 1996, pp. 456-459.

Davis, Lawrence H. (1970). Individuation of actions. *The Journal of Philosophy*, **67**, 1970, pp. 520-530.

Hardcastle, Valerie Gray (2004). The elusive illusion of sensation. *Behavioral and Brain Sciences*, **27**, 2004, pp. 662-663. (Open Peer Commentary article on Wegner (2004a).)

Heyman, Gene M. (2004). The sense of conscious will. *Behavioral and Brain Sciences*, **27**, 2004, pp. 663-664. (Open Peer Commentary article on Wegner (2004a).)

Holton, Richard (2004). *The Illusion of Conscious Will*, by Daniel Wegner. *Mind* **113**, 2004, pp. 218-221. [Book review of ICW.]

ICW. Same as: Wegner, Daniel M. (2002).

Jack, Anthony I., and Philip Robbins (2004). The illusory triumph of machine over mind: Wegner's eliminativism and the real promise of psychology. *Behavioral and Brain Sciences*, **27**, 2004, pp. 665-666. (Open Peer Commentary article on Wegner (2004a).)

Kihlstrom, John F. (2004). "An unwarrantable impertinence". [Quotation marks in original title.] *Behavioral and Brain Sciences*, **27**, 2004, pp. 666-667. (Open Peer Commentary article on Wegner (2004a).)

Krueger, Joachim I. (2004). Experimental psychology cannot solve the problem of conscious will (yet we must try). *Behavioral and Brain Sciences*, **27**, 2004, pp. 668-669. (Open Peer Commentary article on Wegner (2004a).)

Laplace, Pierre Simon (1902). *A Philosophical Essay on Probabilities*. Trans. by Frederick Wilson Truscott and Frederick Lincoln Emory. New York: John Wiley & Sons, 1902.

Libet, Benjamin (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *The Behavioral and Brain Sciences*, **8**, 1985, pp. 529-539.

MacKay, Donald M. (1985). Do we "control" our brains? *The Behavioral and Brain Sciences*, **8**, 1985, p. 546. (Open Peer Commentary article on Libet (1985).)

Mackie, David (1997). The individuation of actions. *The Philosophical Quarterly*, **47**, 1997, pp. 38-54.

Marks, Charles E. (1980). *Commissurotomy, Consciousness and Unity of Mind.* Montgomery, Vermont: Bradford Books, 1980.

O' Connor, Timothy (2005). Freedom with a human face. *Midwest Studies in Philosophy*, **29**, 2005, pp. 207-227.

Richards, Norvin (1976). *E pluribus unum:* a defense of Davidson's individuation of action. *Philosophical Studies* **29**, 1976, pp. 191-198.

Schooler, Jonathan W. (2002). Re-representing consciousness: dissociations between experience and meta-consciousness. *Trends in Cognitive Sciences*, **6**, 2002, pp. 339-344.

Sharlow, Mark F. (2001). *From Brain to Cosmos*. Parkland, FL: Universal Publishers, 2001.

Van Gulick, Robert (1985). Conscious wants and self-awareness. *The Behavioral and Brain Sciences*, **8**, 1985, pp. 555-556. (Open Peer Commentary article on Libet (1985).)

Velmans, Max (2003). Preconscious free will. *Journal of Consciousness Studies*, **10**, 2003, pp. 42-61.

Velmans, Max (2004). Why conscious free will both is and isn't an illusion. *Behavioral and Brain Sciences*, **27**, 2004, p. 677. (Open Peer Commentary article on Wegner (2004a).)

Wegner, Daniel M. (2002). *The Illusion of Conscious Will*. Cambridge, MA: The MIT Press. [Cited in this paper as ICW.]

Wegner, Daniel M. (2004a). Précis of *The illusion of conscious will*. *Behavioral and Brain Sciences*, **27**, 2004, pp. 649-659.

Wegner, Daniel M. (2004b). Frequently asked questions about conscious will. *Behavioral and Brain Sciences*, **27**, 2004, pp. 679-692. (Response to Open Peer Commentary article on Wegner (2004a).)

Wegner, Daniel M., and Wheatley, Thalia (1999). Apparent mental causation: sources of the experience of will. *American Psychologist* **54**, 1999, pp. 480-492.

Wood, Charles C. (1985). Pardon, your dualism is showing. *The Behavioral and Brain Sciences*, **8**, 1985, pp. 557-558. (Open Peer Commentary article on Libet (1985).)

Notes

[1] In using the abbreviation ICW, I follow Wegner (2004b), who uses the same abbreviation (albeit in italics) to refer to his own book. Many of the points made in that book also are made in an article, Wegner (2004a), which is a condensation of ICW. In these cases of duplication, I cited the book itself.

[2] This does not imply that I am a determinist. I do not think we know enough about physics to make a final decision on determinism. I am a compatibilist because I do not think determinism, *if true*, would rule out free will.

[3] Wegner quotes Laplace on this in ICW (p. 1, footnote).

[4] I think this is quite clear from ICW, especially pp. 26-28 and chapter 5.

[5] See also Wegner (2004b), p. 682, where Wegner takes seriously the issue of the appropriateness of the word "illusion" and discusses the nature of the illusion he had in mind.

[6] Wegner later defended the identification of conscious will as a feeling, in Wegner (2004b), pp. 681-682. I do not think this defense adds anything important to what is in ICW.

[7] Velmans (2003, p. 44) has pointed out the "reasonably accurate" character of most conscious perceptions. Velmans' remark, which is right, was made in the context of a discussion of will.

[8] Note also that in ICW, a "voluntary action" is characterized as "something a person can do when asked" (p. 32). This characterization leaves open the possibility of non-consciously-willed voluntary action, if one reads "something a person can do" to mean a sequence of movements a person can undergo.

[9] See, for example, Davis (1970), pp. 520 and 524, for mention of the idea that a movement must meet certain criteria to count as an action.

[10] Davis (1970, pp. 520, 524) mentions the idea that a movement must meet certain criteria to count as an action.

[11] For some interesting papers on this topic, see Davis (1970), Richards (1976), and Mackie (1997).

[12] Wegner cites some of the relevant literature in ICW (pp. 19, 159). Indeed, he has used the idea of "multiple identifications or descriptions" of action (p. 159)—a topic closely related to individuation—in his work on "*action identification theory*" (p. 159; italics in original). This work is described briefly in ICW (pp. 160-161). However, ICW does not trace the *full* impact of issues of individuation on Wegner's view of conscious will. If Wegner had done this in ICW, he would have had to soften his dismissive view of conscious will. In the present paper I will try to show why.

[13] For some discussion of this idea, see Richards (1976), p. 193.

[14] Of course, this analogy cannot be pushed too far, because the statement about bachelors depends more obviously upon the meanings of words than does the statement about actions.

[15] These three classes of phenomena are discussed in various places in ICW, most notably chapter 4 for automatisms, chapter 8 for hypnosis, and pp. 45-49 for stimulation of the brain.

[16] Velmans (2003, p. 60) notes that the "unconscious and preconscious mind/brain" are within the self. See also Velmans (2004).

[17] The feeling does not always occur; see ICW, pp. 286-287, and Kihlstrom (2004).

[18] Wegner credits Ellen Langer for this terminology (ICW, p. 9).

[19] To see what I mean by this, see Mackie (1997), pp. 46 and 50. Interestingly, Wegner comes very close to confronting this difficulty in its general form. He mentions (ICW, p. 18) that some actions "seem to be nested within" others. Later he points out, correctly,

that "even holding perfectly still can be a variety of acts" (ICW, p. 157).

[20] In a related vein, Velmans (2003, pp. 42-44) has suggested that some so-called unconscious processes may be conscious in certain senses. (These senses do not appear to coincide with what either Block or I had in mind.)

[21] On metacognition generally, see Schooler (2002).

[22] Presumably this same view of disunified consciousness can be applied to multiple personality disorder (discussed in ICW, pp. 255-263). Perhaps it also can be applied to other examples of what Wegner calls "Virtual Agency" (ICW, p. 221) in which an imagined or believed-in agent, like a spirit, seems to take possession of a person (ICW, chap. 7).

[23] Jack and Robbins (2004), commenting on Wegner (2004a), suggested that intentions can be conscious without being metaconscious, and pointed out that this fact hurts Wegner's argument. Ainslie (2004), also commenting on Wegner (2004a), suggested that some of the phenomena Wegner describes involve "a split of consciousness" (p. 660). MacKay (1985) suggested that the precursor events in the Libet experiments are deeply rooted in the processes that underlie consciousness. Wood (1985), in a commentary on Libet's work, pointed up the fact that a conscious system may have unconscious parts (p. 558). Van Gulick (1985), also commenting on Libet's work, differentiated two senses of "conscious mental state" (p. 555), and suggested that the precursor states in Libet's work might be conscious in the sense that they are objects of awareness, with the awareness coming after a time delay.

Still No Disproof of Free Will

Has science debunked free will? A recent article *Nature Neuroscience* [1] tells of some research that suggests the answer is "yes." An article in *The Wall Street Journal Online* [2] explores this research - and its implications for free will - in less technical terms.

According to the research, our brains can show specific kinds of activity about 10 seconds before we make conscious decisions. The findings suggest that when you make a conscious decision, your brain already has "decided" as much as 10 seconds earlier. So what is the role of your conscious decision? Does your act of deciding do anything? It seems as if your feeling of conscious decision is just a side effect of brain activity that already has happened. As one of the researchers pointed out (in [2]), this makes things look bad for free will.

It seems as if science might have debunked free will.

But wait a minute! Things just aren't that simple.

There is a way of understanding these findings that does NOT rule out free will. Maybe your brain starts a decision a while before you consciously decide. However, you can believe this and still believe in free will. All you have to do is admit that your actual consciousness includes more than your so-called conscious mind.

Psychologists (especially psychoanalysts) have long claimed that people have unconscious minds as well as their ordinary conscious minds. Philosopher Ned Block [3] has suggested that contents of the so-called unconscious might actually be conscious in a sense. This raises the possibility that your so-called unconscious mind might not truly be empty of consciousness, but might have a consciousness of its own. This would be a consciousness that you normally can't think or talk about - but that is a real part of you anyhow. (I've explored this idea further in my book, *From Brain to Cosmos* [4].)

Now what if you made a decision, but the decision happened in your unconscious mind? Since your unconscious mind is part of you, the decision truly would be your own - just as if you had made it with your ordinary conscious mind. For all we know, it could even be a free choice. (Some of the people who commented on the Wall Street Journal article made these two points about the unconscious. [5]) But what is really interesting is that your so-called unconscious choice might really be a *conscious* choice. This would happen if the so-called "unconscious mind" has some consciousness. Even if this were the case, you might not be able to think or say that you had decided, or act on the decision.

This might be what is happening in the study in *Nature Neuroscience*. The brain events that happen 10 seconds before the "conscious" decision might really be, or contain, the person's own free decision, involving conscious processing of a sort. However, it is a decision that he or she cannot yet think or talk about, or act upon.

In other words, free will and conscious choice might exist even if the neuroscientists' findings are right. The findings might show that we don't understand ourselves as well as we think. Specifically, they might show that the unconscious parts of ourselves are much more important than we usually suppose them to be. But the findings cannot debunk free will.

Just think about that!

(The argument I used in this post is not new. It's based on the one in my paper, "Yes, We Have Conscious Will." [6] That paper is a response to another line of argument against free will - not the same as the one discussed here, but in the same vein. If you're interested in the details of my argument, in further references on these topics, and in some other rebuttals to arguments against free will, read that paper.)

References

[1] Chun Siong Soon, Marcel Brass, Hans-Jochen Heinze and John-Dylan Haynes, "Unconscious determinants of free decisions in the human brain," *Nature Neuroscience*, 11, 543-545, April 2008.

[2] Robert Lee Hotz, "Get out of your own way," *The Wall Street Journal Online*, June 27, 2008, p. A9.

[3] Ned Block, "How can we find the neural correlate of consciousness?", *Trends in Neurosciences* (Reference Edition) 19, 456-459.

[4] Mark F. Sharlow, *From Brain to Cosmos*. Parkland, FL: Universal Publishers, 2001.

[5] WSJ.com Forums, linked from reference [2].

[6] Mark F. Sharlow, "Yes, We Have Conscious Will," 2007. Available at http://philsci-archive.pitt.edu/archive/00003778 .

Slightly modified 10/9/2010 (one link updated).

A Note on the Next Article

The next article deals with an old philosophical problem: the question of the reality of abstract objects. This article might not seem relevant to the three points I set out in the Introduction. If you read this article and then read the next few pieces, you should find that this article is relevant indeed. Questions about abstract objects can have a decisive bearing on our view of the nature of the self.

Getting Realistic about Nominalism

Mark F. Sharlow

URL: http://www.eskimo.com/~msharlow

ABSTRACT

In this paper I examine critically the relationship between the realist and nominalist views of abstract objects. I begin by pointing out some differences between the usage of existential statements in metaphysics and the usage of such statements in disciplines outside of philosophy. Then I propose an account of existence that captures the characteristic intuitions underlying the latter kind of usage. This account implies that abstract object existence claims are not as ontologically extravagant as they seem, and that such claims are immune to certain standard nominalistic criticisms.

I. What Do People Really Mean by "Exist"?

There appears to be a marked difference between the way in which philosophers use the word "exist" and the way in which many other people use that word. This difference often shows itself when beginners in philosophy encounter philosophical positions that deny the existence of seemingly familiar things. Take, for example, nominalism — a view according to which multiply exemplifiable entities, such as properties and relations, really do not exist. (This definition may not do justice to all versions of nominalism, but it is close enough for our present purpose.) A strict nominalist has to deny, for example, that there are such things as colors. He can admit that there are colored objects; he even can admit that we usefully speak as though there were colors. But he must deny that there actually are colors, conceived of as multiply exemplifiable entities.

A newcomer to philosophy might hear about the nominalist view of colors, and say in amazement, "How can anyone claim that there are no such things as colors? Look around the room — there they are!" To lessen this incredulity, a nominalist might explain that he is not denying that we experience a colorful world, or that we can usefully talk as if there are colors. He may claim that he is not really denying the truth of what the beginner is trying to say about the world when the beginner says "Look around the room — there they are!" The nominalist is only denying that there are entities called colors, in addition to the colored objects that we find around us.

A similar incredulity occurs when some philosophers deny that mathematical objects, like numbers, are real. The reaction of a newcomer to philosophy might be "Do you *really* think that there aren't any numbers? Then what do you count with, anyhow?"

In my opinion, these beginners' reactions point up a difference between two slightly different readings of statements asserting existence. To see what this difference is, first take note of the fact that the phrase "there is" and the colloquialism "there is such a thing as" often are used interchangeably in non-philosophical English. The phrase "there is such a thing as" is not always used to assert the existence of an entity — or at least when it is

used, one cannot quite be sure than an entity is being posited. As examples of such usage, consider the following sentences: “There‘s such a thing as meanness in the world.” "There is such a thing as cold." "There is such a thing as hope." For each of these sentences, a speaker who asserts the sentence is asserting a fact about certain phenomena in the world — but it is not clear that the speaker means to postulate entities named "meanness," "cold," or "hope." In view of this, it is interesting that colloquial usage does not draw any distinction, or at least any clear distinction, between the meaning of “there is such a thing as” and the meaning of "there is." This near-equivalence of "there is" and "there is such a thing as" may help us, in a roundabout way, to understand why the beginner in philosophy finds the nominalistic claim "There are no colors" to be amazing. To the beginner, denying the truth of "There is a color" is tantamount to denying an obvious fact about phenomena in the world — namely, the fact that we can find the phenomena of color out there in the world. No color-perceiving lay person, and certainly no color-perceiving artist, would want to do this.

In metaphysics, however, it is standard to use "there is" in a slightly different way — to indicate the existence of an entity, or entities. This is the othe part of the reason why the philosophical beginner finds it so hard to swallow the nominalist’s claim that there are no colors. The nominalist simply does not mean that our world is colorless. He simply means that there are no *entities* which one can call colors, *in addition to* the colored entities in the physical world. Occasionally, philosophers seem to use "there is" in an even more conservative sense, to indicate a kind of existence that is somehow irreducible. (Consider the following sentence, which some philosophers have believed: "Elementary particles really exist; strictly speaking, tables and chairs do not.") But even philosophers who do not go this far tend to use “there is” to indicate the existence of an entity. A nominalist philosopher can plausibly deny that there are colors, because when she says that there are no colors, she doesn't mean that we don't find coloration, or colored things, in the world around us. Instead, she means that there are no real entities called colors that exist in addition to colored objects. Judging by the way that lay people often talk about colors, it is not at all clear that "There are no colors" means exactly the same thing to most

lay people that it means to most philosophers.

If we examine more closely these two usages of "There are colors," we begin to discern more clearly what the lay person and the philosopher may really mean by this sentence. In prephilosophical usage, existence statements like "There are colors" are used to express facts about the real world. Sometimes, these facts seem obviously true. According to common non-philosophical usage, "There are colors" safely can be asserted if one can find color in the world — or, more precisely, if the real world is, at least in part, colored. I think this is the way that most non-philosophers use "There are colors." They are prompted to assert this sentence because they see color, or because they believe that something is colored. They assert "There are colors" without taking any thought upon the problem of whether colors are separate entities, or are in any sense ultimately real. To say that there are no colors, or that colors do not exist, is for most people tantamount to saying that we live in a colorless world. This is why the philosopher's denial of the truth of "There are colors" may meet with the non-philosopher's incredulous response: "You mean there are no such things as colors? But *look* at them!" When the philosopher says that there are no colors, what he means is that there are no entities that can be regarded as colors, above and beyond the concrete physical objects in the world. This is not what most non-philosophers mean by "There are no colors" — at least if we can guess something about what they mean from how they speak.

It appears that existential statements are used a little differently in metaphysical discourse than in ordinary discourse. To improve our understanding of what this difference really is, we will consider nominalism again.

II. When Nominalists Go Bad

Consider the following example.

An artist has a painting hung in a gallery. Some time later, an art critic writes that the colors used in the painting are similar to those used by 19th-century house painters. Then the artist, who also happens to be a nominalist, rebuts this charge as follows: "It is not

true that I used colors of that kind, for *there are no such things as colors*."

Clearly, this rebuttal is silly. It would be silly even if both artist and critic were nominalists — and even if nominalism were in fact true. But are we sure that we understand *why* the rebuttal is silly? It is clear that the critic is trying to say something, and that what the artist is saying in rebuttal does nothing to prove the critic wrong. How can we best understand the fact that the artist's rebuttal misses the point?

The explanation that many nominalists probably would offer runs along these lines: The critic's statement can be true even if there are no such things as colors. What makes the critic's statement true is not a fact about items called "colors," but circumstances of other sorts. The precise nature of these circumstances depends upon which version of nominalism is assumed correct [1]. If Resemblance Nominalism is true, then these circumstances could involve resemblances between the painting and other physical objects. If Predicate Nominalism is true, the circumstances could involve facts about language and linguistic objects. If Class Nominalism is true, then the circumstances could involve facts about what objects belong to which classes. But these circumstances, whatever they are, do not involve multiply exemplifiable entities known as colors, over and above the particular entities (painting, patches of paint, perhaps classes, etc.) involved.

Later I will have more to say about nominalistic explanations. But for now, I just want to use this example to draw attention once again to a fact about philosophical language. When philosophers discuss questions about what really exists, they often use existential statements in a way *markedly different* from the way in which non-philosophers use such statements. An artist or an art critic feels free to assert statements about colors, even if those statements directly entail that colors exist. Philosophers also make such assertions, but for the philosopher, these assertions are much more tentative and problematical. They are assertions that may need to be defended. And some philosophers (at least when they speak in their capacity as philosophers) do not dare to make such statements and mean them.

One can think up many other examples similar to that of our artist. One can imagine a philosophically minded biologist who claims that, contrary to common experience,

composite flowers *do not* exhibit spiral patterns — because patterns are abstract objects, and there are no abstract objects. Or, one can imagine a nominalist lawyer who argues in court that his client had no legal duty to avoid stealing, because a legal duty is an abstract object and there are no abstract objects. Such arguments obviously do not constitute good biology or good law.

Real nominalists are more sensible than the nominalists in the above examples. If real nominalists want to explain the silliness of the preceding biological and legal arguments, they must do so in one of two ways. One way is to offer an account of how statements that seem to be about abstract objects can be true even if, strictly speaking, there are no abstract objects. The other way is to go ahead and deny that statements about spiral patterns or legal duties are literally true — but perhaps to allow that these statements nevertheless are legitimate for use in certain kinds of discourse. I do not intend, quite yet, to accept or reject either of these approaches. I am using the preceding examples from art, biology and law only to point up the important fact that the *use* of existence statements in philosophical discourse tends to be markedly different from the use of such statements in the discourse of other fields. Only an ontologist would ever try to deny that some flowers exhibit spiral patterns. A botanist or a mathematician — the experts most directly concerned with flowers and with spirals — would not.

III. Objects and Situations

The above examples show that existence statements can be used in two different ways. One of these ways corresponds to the way that existence statements actually are used by scientists, mathematicians, lawyers, and others. When existence statements are used in this way, it is appropriate to assert an existence statement when a situation of a specific kind obtains in the world. When speaking in this way, one can assert "There exist colors" if the physical world is in part colored; one can assert "There are numbers" if it is possible to think of numbers and to engage in ordinary numerical reasoning. The other way corresponds to the way in which metaphysicians typically use existence statements. When

one follows this way, it is correct to assert "There exist X's" only if there is an entity which is an X. Those who use existence statements in this second way are forced, in many cases, to regard the first kind of usage as strained, figurative, or otherwise not quite literal.

The relationship between these two ways of speaking is more complex than it might at first seem. It appears that the second way actually is a variant of the first way. According to the second way, "There are X's" can correctly be asserted if a certain *situation* obtains in the world — namely, the situation of the existence of an entity that is an X. Thus, a language user following the second way actually is following a special variant of the first way — a variant in which the only appropriate situation for asserting "There is an X" is the situation of there being an entity that is an X.

A more serious complication arises from a kind of circularity involved in the second way of using existence statements. Suppose that a color nominalist is trying to convince an artist (who, in this new example, is *not* a nominalist) that it is a mistake to claim that there really are colors. The artist, we will suppose, reads "There are colors" in the first way, and therefore implicitly holds "There are colors" to be true if and only if a situation of some particular kind obtains. It does not matter exactly what the artist takes this kind of situation to be; let us simply call situations of the appropriate kind "C-situations." (The situation of there being a yellow patch on a painting is an example of a C-situation.) The nominalist argues that the artist's position is wrong, and that "There are colors" is true if and only if there is a real entity that is a color. This tells us that the nominalist doesn't view C-situations as situations which can ensure that "There is a color" is true. According to the nominalist, the only situations that can do that are situations of the existence of an entity that is a color. Let us call these latter situations "E-situations." According to the nominalist, E-situations do not occur, so it follows that "There is a color" is false, and that there are no colors. However, the artist doesn't see the point of this argument — because the artist believes that it is the C-situations that make "There are colors" true, and therefore make it true that there is a color! In the artist's opinion, it is only the C-situations that count in determining whether there is a color. Hence as long as some C-situation obtains, the nominalist's argument can have no force for the artist.

By accepting that C-situations are what makes “There is a color” true, the artist implicitly accepts that that a C-situation *is* a situation of the existence of a color. On the artist's view of things, the situations in the world which make "There are colors" true are precisely the C-situations. Therefore, on the artist's view of things, the situations which can correctly be described as situations of the existence of a color are, in fact, just the C-situations. The nominalist claims that a situation of the existence of a color, rather than a C-situation, is necessary for “There is a color” to be true. But for the artist, a C-situation *is* a situation of the existence of a color.

The upshot is that the nominalist's argument cannot be convincing for the artist. The nominalist claims that "There are colors" is literally true only if there really are colors — the mere obtaining of a C-situation will not suffice. The artist, on the other hand, thinks that a C-situation *is* a situation of the existence of a color. Thus, the nominalist's argument that only color existence situations will do cannot have any force for the artist.

The real difference between the positions of the artist and the nominalist in this example is a disagreement about what kinds of situations must obtain in order for "There are colors" to be true. The artist thinks it is the C-situations that are required. The nominalist thinks it is another kind of situations, which he might label "situations of real existence of a color." However, the artist *also* holds (implicitly or explicitly) that the C-situations are exactly the situations of real existence of a color. Both the artist and the nominalist hold, at least implicitly, that "There is a color" is true only if there really is a color. (Certainly, when the artist sincerely asserts "There is a color," he means what he says.) The real locus of the disagreement between the artist and the nominalist is the issue of whether C-situations are situations of the existence of a color. The nominalist's claim, that "There are colors" is true only if there are entities that are colors, cannot convince the artist that C-situations are insufficient to ensure that "There are colors" is literally true.

The nominalist might try to convert the artist by an alternate strategy: conceding that "There are colors" is true, but claiming that it is true for reasons other than the real existence of a color. (For example, it could be true because of resemblances among objects, or for some other reason that can be described without mention of colors.) But

this nominalist argument, like the previous one, should have no force for the artist, because the artist believes that a situation which apparently does not involve a color can nevertheless be a situation of the existence of a color. Even if the artist conceded that "There are colors" is true for reasons not apparently involving colors, that would not be enough to convince the artist that there are no colors. At most, the artist might have to concede that the situation of the existence of a color (a C-situation) also can be described as a situation of another kind (say, a situation in which certain resemblances hold among objects). The nominalist might be able to force the artist to admit that there is another way of describing a color-existence situation — a way that does not involve any mention of colors. But the availability of such a description does *not* imply that the supposed color-existence situation really is not a color-existence situation.

The preceding paragraphs may lead us to wonder whether situations involving the existence of a color can be described in ways that do not involve any mention of colors, but that are *equally correct* from an ontological standpoint. The idea that this can be done is not a new idea. H. H. Price ([1969], pp. 30-31) once proposed that a certain kind of realism and a certain kind of nominalism might be merely alternative "terminologies" for "saying the same thing" (p. 30). What I am proposing here is not quite the same as Price's idea; the difference between my position and Price's will become evident later. But Price's suggestion at least implies that one can describe a situation of the existence of a property in other ways not involving properties, without entirely abolishing the ontological import of the description in terms of properties.

Non-philosophical discourse provides examples of situations that can be redescribed in ways that do not mention certain things that are really and genuinely involved in those situations. One such example comes from physics [2]. Consider the situation of the existence of an electric charge inside a closed surface in space. One can describe this situation alternatively as a situation in which a certain quantitative property of the closed surface (known as the “surface integral of the electric field“) is not zero. (The fact that one can do this is a consequence of Gauss's Law of electromagnetism.) This new description does not mention electric charge at all — yet it describes *precisely the same*

physical situation. Analogously, it may be possible to redescribe the situation of the existence of a color in some way that makes no mention of colors. Yet the existence of such a description does not imply that there really are no colors — any more than our ability to describe the presence of electric charge without mentioning electric charge implies that there really is no electric charge anywhere in the universe.

The preceding argument is not a head-on refutation of nominalism. That argument does not rule out the possibility that there are conditions K not explicitly involving colors, such that "There are colors" is true when and only when those conditions K hold. However, the argument does show that even if this is the case, we are not *forced* to accept that colors do *not* exist. Instead, we are free to regard K as the conditions for the real existence of a color. If we wish to regard K in this way, then the nominalistic counterargument, based on the claim that only the real existence of a color can make "There are colors" true, cannot stop us from believing that colors exist whenever the conditions K hold.

These arguments teach us the following very general lesson. Suppose that we regard "There are X's" (where "X's" is a placeholder for "colors," or "numbers," or the like) as being true if and only if some circumstances (call them "X-circumstances") obtain in the world. After doing this, we are free to assert that the X-circumstances are the circumstances under which there exists an X. Once we have made this move, we are immune to any skeptic who tries to persuade us that "There are X's" is false by arguing that only the actual existence of an X, and not mere X-circumstances, can make this sentence true. The skeptic's claim that the circumstance of existence of an X will do, but X-circumstances will not, is a claim that no longer makes sense to us. Also, we are immune to any skeptic who allows that "There are X's" is true while claiming that this statement is true for other reasons besides the existence of an X. This skeptic cannot rule out the possibility that the circumstances which make "There are X's" true are in fact circumstances of existence of an X, described in some way that does not seem to involve an X.

IV. Objects and Situations Again

The above argument shows that certain lines of nominalistic argument, taken by themselves, cannot succeed in refuting the claim that abstract objects exist. Even if it turns out that our reason for regarding "There are colors" as true is a set of circumstances that doesn't involve colors at all, we still can maintain that colors really exist — provided that we identify a color's existence with the set of circumstances that makes "There exist colors" true. In other words, we must regard colors as things which exist if and only if certain circumstances obtain — circumstances which also can be described in a way not involving colors. We must not regard the latter circumstances as merely implying the existence of a color; we must regard them as *being* the existence of a color [3]. The existence of a color is a situation of a particular sort. It is the situation of a color's existing. This situation, we must suppose, *is* the situation which makes "There are colors" true — even if that situation is one that can also be described *without* any reference to colors.

Let me try to make the ideas in the preceding two paragraphs a little clearer. Suppose that a nominalist claims that "There are colors" is true only because resemblances of certain kinds hold among particular physical objects. (Resemblance Nominalism treats color statements in just this way [4].) Suppose, further, that a person named John who thinks colors really exist hears this argument and become convinced of its conclusion. Then John has two choices. He can stop believing that colors exist, on the grounds that "There are colors" is true only because of circumstances that do not involve colors. Or, he might conclude that the circumstances which make "There are colors" true are the very circumstances that *constitute* the existence of colors. If he takes the latter path, then he must accept that the circumstance of some physical objects' resembling each other in a certain way is precisely the circumstance of the real existence of a color. A color, in other words, is an entity that can exist by virtue of the fact that physical objects resemble each other in certain ways. A color is the kind of thing that can exist by virtue of circumstances that can be described without any reference to colors. The existence of a color is, after all,

a state of affairs; that state of affairs is identical to the state of affairs of physical objects' resembling each other in certain specified ways. This position of John's amounts to the following: If one wants to describe the state of affairs of the existence of colors, one doesn't have to describe it as "the existence of entities which are colors," or something like that. One also can describe it as "the resembling of physical objects by one another in such-and-such ways" (where "such-and-such ways" should be replaced by an appropriate listing of the required resemblances). Both of these descriptions point to the same state(s) of affairs.

This position that John must take is not as strange as it seems. Abstract items like colors are not the only items whose existence is equivalent to states of affairs that seemingly do not involve them. Consider our earlier example about electric charges [2]. According to Gauss's Law of electromagnetism, there is a net electric charge inside a closed surface in space if and only if the electric field on that surface meets a certain condition — namely, the surface integral Φ of the electric field over that surface must be nonzero. If Gauss's Law is true, then it is a physically necessary truth that there is a net electric charge inside a surface if and only if Φ is nonzero for that surface. If Gauss's Law is taken as a definition of electric charge (physicists sometimes speak as if it were [5]), then this truth is even logically or conceptually necessary. By saying "The electric field on some surface has a nonzero surface integral," we are specifying exactly the same state of affairs that we would specify by saying "There exist electric charges." Thus, we can describe the state of affairs of the existence of electric charges without even mentioning entities or things called electric charges. Yet the state of affairs that we describe in this way *is* just the state of affairs of existence of an electric charge.

No physicist should argue that electric charges are utterly unreal just because we can specify the state of affairs of the existence of electric charges by talking about electric fields instead of about electric charges. Similarly, no philosopher should argue that colors are unreal just because we can specify the state of affairs of the existence of colors by talking about resemblances instead of about colors. For colors as for electric charge, the existence of the thing in question is a state of affairs that can also be described without

reference to the thing itself.

One can think of less technical examples in which the existence of something is a state of affairs which also can be described without reference to that thing. This often happens in cases involving wholes and parts. For example, the state of affairs of the existence of a book can also be described as the state of affairs in which pages meeting certain criteria are united in a certain way. But this fact does not force us to deny that books exist, or to claim that books are somehow less real than pages. Some philosophers may want to make such claims, but these claims are not forced upon us by the identities that hold between the relevant states of affairs.

Aside from these particular examples, one might even argue that the existence of any material object is redescribable in this way. It sometimes is the case that there exists a material object occupying a spatial region R. We assert that this is the case only when observations of the region R reveal results of certain kinds. Thus we should be able, in principle, to state conditions which are necessary and sufficient for the existence of a material object occupying R, such that these conditions do not mention material objects occupying R. If there are such conditions, then the existence of a material object occupying R is a state of affairs which has an alternative description making no mention of material objects that occupy R. But even if we knew of such conditions, this would not be good reason to deny the reality of material objects occupying R — especially if we feel sure of the existence of material objects in other places. (Some idealists might want to make this denial, but the mere existence of the conditions just described does not force us into idealism.)

The existence of an object is a situation — or, if one prefers, a state of affairs or a set of circumstances. In many cases, one can think of the existence of an object as a situation describable in terms that have nothing to do with the object in question. But this fact does not imply that the object is unreal.

The arguments in this section show that most nominalistic critiques of the reality of abstract objects are doomed to be ineffectual. It is impossible to show that abstract objects of a certain kind do not exist, either by claiming that there are abstract-object-free

truth conditions for abstract object statements, or by arguing that abstract object statements are not literally true because the situations prompting their assertion can be described without abstract objects. Both of these principal lines of nominalistic critique fail, because both of them tacitly depend upon the same erroneous assumption: that a situation described without recourse to objects of a certain kind cannot also be a situation of the existence of an object of that kind. If one denies this assumption, then one can get around the nominalistic critique. We have seen that this assumption fails in a number of interesting cases.

By now it should be a bit clearer how my position differs from Price's suggestion which I mentioned earlier. Price suggested that a nominalistic and a realistic description of the same facts may be merely descriptions in alternative languages, with no final fact of the matter about which description is correct ([Price 1969], pp. 30-32). Like Price, I am proposing that certain situations can correctly be described either in a way involving abstract objects or in a way not involving them. However, unlike Price, I am arguing that the two descriptions are fully compatible with one another, and can be true simultaneously. There is no question of either description being true merely in one language or from one standpoint. Both descriptions simply are right, and can be stated in the same natural or formalized language, where one cannot derive a contradiction from them.

It is worth noting that our new view of the existence of abstract objects leaves open the question of whether properties and relations can exist uninstantiated. For all we know, some abstract objects may exist thanks to situations that do not require the existence of physical objects. Some philosophers have held that a property or relation exists if its exemplification is possible (see for example [Armstrong 1989], pp. 80-81, where Armstrong takes the opposing view). One can ask whether the situation of the existence of some abstract object might be a situation of the mere *possibility* of certain physical circumstances. The fact that these circumstances are possible might remain the case even if there are no physical objects, thus ensuring the existence of the abstract object. I will not pursue this idea further here. I mention it only to show that our view of abstract

objects does not automatically rule out uninstantiated properties and relations.

V. Some Remaining Qualms

Even after reading the above arguments, one may feel some intuitive qualms about the claim that a situation of the existence of an object may also be a situation of a seemingly different kind. There seems to be a palpable difference between (say) the situation of the existence of a stone, and the situation of the existence of a color. It seems as though the situation of existence of a stone obtains because the stone is just *there* — present in the world — while the situation of the existence of a color obtains for a subtler reason: because some physical objects are colored. The existence of a stone seems to be a situation centered on a *thing*. The existence of a color seems different; it seems to be more diffuse, based on a variety of scattered facts (that particular physical objects resemble each other, say), and not centered on one particular item like a stone. Could it be that the first of these situations (with the stone) is a situation of existence of an object, and the second situation (with the color) is not?

My answer to this line of criticism is simple: If one is asking the question, then one has not taken the argument of the previous section seriously. It is incorrect to say that the existence of a stone cannot be described as a diffuse, "non-centered" situation incorporating a variety of facts that are not about stones. One can regard the existence of a stone as a complex situation based on facts about what goes on within and outside of a certain region of space. Also, it is circular to argue that situations of color existence are not really situations of object existence on the grounds that situations of color existence do not center on one object. To deny that a situation of color existence centers on one object is to deny that colors are entities — which is to assume what this line of criticism would claim to prove.

Another qualm about my argument arises from Occam's Razor. Many people would prefer not to believe in abstract objects because such objects are supposed not to be necessary for our understanding of the world. Abstract objects are thought of as excess

metaphysical baggage. But we already have the reply to this argument. Take colors as an example again. If we suppose that colors are not necessary to explain the facts about the world, then we are tacitly assuming that the fact of the existence of colors is not among the facts about the world. However, if the situation of existence of a color can be redescribed in terms not involving colors but involving physical objects, then the fact that there is a color is a fact of the physical world. We cannot escape from this fact by refusing to postulate further objects. Whenever we postulate colored objects, we let colors in through the back door. To kick them out again would require a denial of the real physical situations that constitute the existence of colors.

VI. Why Nominalism Seems So Plausible

Despite its ability to dodge Occam's Razor, the thesis I have presented in this paper may still seem intuitively implausible to some. The reasons why it seems implausible may shed light on the reasons why some people want so badly to deny the existence of abstract objects.

One intuitive feeling that can impede our understanding of abstract objects is what might be called the "physiomorphic illusion." This is the deep-seated feeling that if something exists, then it must exist in much the same way that a concrete physical object exists. We tend to picture abstract objects (if we picture them at all) as extra items sitting there in the ontology of the world, alongside or above the physical objects, acting as *additional* building blocks from which the world is constructed. If we try to draw a diagram of our ontology with pencil and paper, the abstract objects typically come out as additional dots, somewhere near the dots representing concrete physical objects. We do not normally think of an abstract object as an entity whose existence is *not* just a matter of adding one more item to the universe of physical objects.

Various philosophers already have warned us of the danger of this illusion. Price ([1969], p. 31) wrote that realist language "may mislead us into supposing that 'there are' characteristics in the sense in which 'there are' dogs, or planets." Russell ([1976], pp. 98-

100) once acknowledged that abstract and concrete objects exist in fundamentally different ways. (In that same passage, Russell even declined to apply the word "exist" to universals, preferring a different terminology instead (p. 100).) But despite these warnings, the physiomorphic illusion persists. When we try to think about abstract objects, we typically end up thinking of them as additional items added to an otherwise concrete world.

This intuitive illusion appears to have afflicted friends and foes of abstract objects alike throughout the history of metaphysics. For example, some philosophers tend to think that if we allow abstract objects into our picture of the world, then we are allowing *extra things* into our picture — and nowadays that is supposed to be a naughty thing to do. Allowing abstract objects into our world picture is anathematized in much the same way that admitting gremlins would be anathematized. Some other philosophers, who believe in abstract objects, tend to picture those objects as items which must be postulated separately from the concrete objects in the world. By relying on this picture, these philosophers open themselves up to attacks from the gremlin-phobes. All these troubles would be alleviated if people would realize that abstract objects are not simply "extra things" that inhabit the world alongside the physical objects that science recognizes. Abstract objects exist *in a different way* from physical objects. We encounter abstract objects as the forms, patterns, qualities and relationships in the concrete world — not as extra things or extra pieces of stuff. We claim that we know that abstract objects are not concrete objects — yet we stubbornly persist in envisioning them as being so much like concrete physical objects that postulating them would be tantamount to adding additional bits of stuff to the world that we already know.

Philosophers who admit the existence of a physical world exhibiting forms, patterns and relationships, but who still look askance at abstract objects, may in some cases be in the grip of the physiomorphic illusion that anything that exists must exist *like a thing* — in much the same way that things exist. According to the view of abstract objects that I am defending here, this illusion is indeed only an illusion.

VII. Possible and Fictional Objects

If abstract objects exist, then there are multiple kinds of existence in the world. Concrete, actual physical objects exist in one particular way — a way that presumably involves (among other things) persistence through time and interaction with other physical objects. Properties and relations exist in a different way; central to their existence is the fact that they exist if and only if certain situations obtain, or could obtain, in the world of concrete entities.

There may be still other kinds of existence corresponding to other kinds of entities in which metaphysicians traditionally have been interested. Such entities include (among others) possibilia and fictional characters. Consider Sherlock Holmes, who sometimes comes up in philosophical discussions of fictional characters. Does Sherlock Holmes exist? If you mean Sherlock Holmes the actual man, then no. If you mean Sherlock Holmes the fictional character, then to deny his existence is to deny the blatant fact that there is a Holmes character in several novels. Perhaps we can extend the ideas in this paper to fictional characters as well as abstract objects, and then claim that the situation of the existence of a fictional character is a situation that can be redescribed in terms of facts about *stories*.

A similar suggestion can be made with respect to possibilia, such as possible worlds. Intuitively, a possible world of a given kind exists if and only if our world could have been a certain given way. Thus, the situation of the existence of a possible world might be taken to be the situation (perhaps only possible) of the world's being a certain way. To make this idea more rigorous, one might claim that the situation of the existence of a possible world is the situation of its being possible that all the propositions in a certain class are true. This class would then be the class of propositions that are true in that possible world. (Not every class of propositions could determine a possible world in this way.) If the conditions for existence of a possible world are like this, then different kinds of possible worlds (logically possible, physically possible, etc.) might have different kinds of existence situations. For example, the situation of the existence of a logically possible

world might be the situation of its being logically possible that all the propositions in a certain class are true. I am not claiming that this is the correct or the best account of the existence situations of possible worlds. I mention this account only to suggest that the existence of possible worlds might, in principle, be understood along the lines that I have laid out in this paper.

VIII. Concluding Remarks

In this paper, I have tried to call into question some familiar ideas about the relationship between nominalism and realism. While attempting this, I arrived at a new view of abstract objects. This view is more liberal than nominalism but is more conservative than some forms of realism. In a sense, this view is a hybrid of realism and nominalism. According to this new view, we are safe in assuming that abstract objects exist even if the nominalists' reductions of abstract object statements are correct. However, when we assert the existence of abstract objects, we are not asserting as much as we might think. This is the case because the existence of abstract objects is intertwined in a certain way with situations in the world of concrete entities. Abstract objects, conceived in this way, do not threaten the scientific outlook; Occam's Razor cannot succeed in cutting them off. Abstract and concrete objects exist in strikingly different ways, but both occupy important places in the totality that we call the world.

Notes

[1] See [Armstrong 1989] for an introduction to various versions of nominalism. I am following Armstrong's nomenclature for these versions.

[2] This physics example was inspired by, and is based on information from, remarks in [Misner *et al.* 1973], pp. 367-368.

[3] Note that we do not have to suppose that this situation *is* the color — only that the situation is a *situation of the existence* of a color. Armstrong ([1989], p. 95) has suggested that a "thick particular" actually is a state of affairs. In view of this, one might be tempted to make a similar claim with regard to properties, and say that a property is a situation of a particular kind. I will remain neutral regarding this further claim; the argument of this paper does not depend upon this claim, but appears to be compatible with it. My argument also does not commit us to the view that a particular of any sort is a state of affairs or a situation.

[4] See [Armstrong 1989], chapter 3, for details.

[5] Some remarks on p. 368 of [Misner *et al.* 1973] almost, but do not quite, amount to such a definition.

General note: In writing this paper I have presupposed some general background information about realism, nominalism, and the problem of universals. Readers unfamiliar with these topics might consult introductory works on the problem of universals, such as [Armstrong 1989], or works on general metaphysics.

References

Armstrong, D.M. 1989. *Universals: An Opinionated Introduction.* Boulder, CO: Westview Press.

Misner, C.W., Thorne, K.S., and Wheeler, J.A. 1973. *Gravitation.* San Francisco: W.H. Freeman and Co.

Price, H.H., 1969. *Thinking and Experience.* (2nd ed.) London: Hutchinson & Co.

Russell, B. 1976. *The Problems of Philosophy.* (Reprint ed.) London, Oxford, N.Y.: Oxford Univ. Press.

Platonizing the Abstract Self

Mark F. Sharlow

ABSTRACT

In this note I examine the two main differences between Plato's and Dennett's views of the self as an abstract object. I point out that in the presence of certain forms of ontological realism, abstract-object theories of the self are compatible with the full reality of the self. I conclude with some remarks on the relationship between ontology and ethics.

The idea that the self is an abstract object is not at all new. One finds this view, or something close to it, in Plato's *Phaedo*, where the highest part of the soul is regarded as akin to the Forms [Plato, 79-80]. Interestingly, the view that the self is an abstract object is compatible, not only with Plato's view of the soul, but also with extreme materialistic positions in the philosophy of mind. Dennett, whose position in *Consciousness Explained* certainly falls into the latter category, argued in that book that the self is an abstraction [Dennett 2, especially ch. 13]. Thus, thinkers as different as Plato (a mystically inclined rationalist) and Dennett (an advocate of science-based materialism) have held that a self is, at least in part, an abstract entity of some sort.

The most significant difference between Plato's and Dennett's versions of the abstract self is a difference in the epistemological foundations of the two viewpoints. Dennett's theory of mind in *Consciousness Explained* is rooted in the empiricism of science, and not in the sorts of intuitive insights, honed by logic and mathematics, upon which Platonism relies. The second most important difference between Plato's and Dennett's conceptions of the abstract self arises from the differences in the two authors' views on the reality of abstract objects. Plato regarded abstract objects as real and very important entities. As far as I can tell, Dennett, in *Consciousness Explained*, did not *unequivocally* regard the self as real.

In *Consciousness Explained*, Dennett states that the self, as an abstract object, is a "fiction" [Dennett 2, pp. 411 and 429]. In other places, he seems to be saying that other abstract objects are fictions, too [Dennett 2, pp. 95-6, 367]. In yet another place, he seems to be saying that the question of the reality of persons is not worth asking [Dennett 2, p. 460]. In a separate paper, Dennett argued for a "mild realism" [Dennett 1, p. 30] with regard to a particular kind of abstract objects -- namely, patterns. This "mild realism" apparently labels as "real" only those patterns that have some scientific, or other predictive, utility. Thus, this so-called "realism" has little to do with traditional metaphysical questions about the reality of abstract objects of different general kinds (properties, relations, sets, and so forth). Indeed, when discussing the reality of beliefs,

Dennett explicitly sets aside "the 'metaphysical' problem of realism" with regard to beliefs [Dennett 1, p. 50], and develops a science-driven notion of realism that has little to do with the general philosophical problem of the reality of abstract objects (see [Dennett 1, pp. 28-29, 30, 45-46, 50]).

To suppose that the reality of an alleged abstract object is a function of the predictive usefulness of that object is to ignore some serious questions about the ontological status of abstract objects in general. Philosophers have long debated the question of the reality of abstract objects such as properties, relations, and sets. Those who have thought about this question have proposed various kinds of nominalism, which state that there really are no abstract objects (or at least no multiply exemplified abstract objects like properties), along with various kinds of realism, which give different accounts of real abstract objects. (For an introduction to these theories, see [Armstrong].) Plato, of course, is the paradigmatic realist -- though one does not have to be as thoroughgoing a realist as Plato to believe that abstract objects are real.

If we assume unequivocally that abstract objects are real, we stand a chance of being able to reconcile the hypothesis that the self is an abstract object with an even more important thesis: that the self is real. Whether the abstract self is real will, of course, depend upon which kinds of abstract objects are real, and which kind of abstract object the self turns out to be. If the self is an abstract object and is real, then one can believe that the self is an abstract object and still consistently believe in the full-blown, undiluted, and undeflated reality of the self. One can believe that the self is "only" an abstract object, and simultaneously believe in the reality of the self as fervently as any dualist or idealist might. One does not have to adopt a realism as bold as Plato's to get this consequence. A much weaker form of realism might do.

If Dennett, in *Consciousness Explained*, had argued unequivocally that abstract objects are real, then he would not have had to call the self a fiction, and the character of the final chapters of *Consciousness Explained* -- which, in my opinion, have virtually nihilistic implications regarding persons -- might have been different.

In closing, I would like to point out one reason for thinking that the self, whether

concrete or abstract, is real. The doctrine of the nonexistence of the self has a consequence of great moral import: that doctrine implies that, strictly speaking, the harming of a human being really harms no one. After all, if the self does not, strictly speaking, exist, then the killing of a human body does not, strictly speaking, really kill anyone. Note that I am *not* attributing belief in this morally loaded consequence to any actual philosopher. However, I do think that some existing philosophical positions may lead us toward this consequence. From a moral standpoint, this consequence is mind-numbingly bad. I would suggest that we avoid this consequence by maintaining that the self, whether abstract or concrete, is real.

This suggestion leads to a more general thought about the relationship between ontology and ethics. Ultimately, we must decide whether we want ontology to be an afterthought to science, denying the reality of anything that is not scientifically useful, or whether our ontology should be robust enough to underpin our moral beliefs as well as our scientific convictions. If we decide that ontology should support ethics as well as science, then we should assume that the self is real. If we do not care whether our ontology is sufficient to support our ethics, then perhaps we remain too much in thrall to the scientism of a bygone century.

References

[Armstrong] Armstrong, D. M. *Universals: An Opinionated Introduction.* Boulder, CO: Westview Press, 1989.

[Dennett 1] Dennett, Daniel C. "Real Patterns." *The Journal of Philosophy*, vol. 88, no. 1 (Jan. 1991), pp. 27-51.

[Dennett 2] Dennett, Daniel C. *Consciousness Explained*. Boston: Little, Brown, 1991.

[Plato] Plato. *Phaedo*. In *Plato's Phaedo*, trans. R. Hackforth. Reprint ed. (original printing 1955). Indianapolis and N.Y.: Bobbs-Merrill.

I Am an Abstraction, Therefore I Am

Mark F. Sharlow

ABSTRACT

In this paper I examine a new variant of the well-known idea that the self is an abstract object. I propose a simple model of the self as a property of temporal slices of a body's history. I argue that this model, when combined with even a modest realism with regard to properties, implies that the self has many of the chief features traditionally attributed to selves. I conclude that this model allows one to reconcile the full reality of the self with even the most deflationary materialistic theories of consciousness.

1. Is Bob's Self a Property?

Of all philosophical positions on the nature of mind, behaviorism is the most dismissive of the reality of the self. Some non-behavioristic versions of materialism also lead toward skepticism about the self, although this cannot be said of materialism in general. But even if behaviorism or some extreme form of materialism turned out to be right, there still would be something that has many of the characteristics normally attributed to the self. This "something" is neither a physical object nor a Cartesian spirit. Instead, it is a *property*.

Consider the history of a person, whom we shall call Bob. Specifically, consider the set of all of the temporal slices in the history of Bob's body, from birth to death. (To define this set, one need not assume -- contrary to special relativity -- that there is a unique way to slice up the history. One also need not assume that the slices are instantaneous, or that the slices have the same temporal thickness.) All of these slices have a property in common: the property of being temporal slices of the history of Bob's body. For brevity, we will call a slice having this property a "B-slice."

One can ask about what makes a given temporal slice a B-slice. One's answer will depend upon one's position on the problem of personal identity (see [Hirsch, chs. 6-10]). Someone might even argue that it is only linguistic convention that makes a given slice a B-slice. (For one take on this possibility, see [Hirsch, ch. 10]). But regardless of what makes a given slice a B-slice, we can safely assume that there are B-slices. This is the case whether the correct philosophy of mind is behaviorism, materialism, or something else entirely (such as property dualism). In any of these cases, one can define B-slices and a property of being a B-slice. For brevity's sake, let us call this property B.

The property B is something that all B-slices, and only B-slices, have in common. It is something that is connected in an obvious way (instantiation) with all of the temporal slices in Bob's life history. Speaking loosely, we could even say that B is "in" all of the

temporal slices in Bob's life history. (We can say this if we let the word "in" have the same informal meaning that it has when we say that a cannonball has a lot of weight *in* it, or that there is a lot of red *in* a sunset. In these instances, "in" indicates instantiation of a property.)

If someone were to claim that the self is not something physical (such as a brain) and that there is no nonphysical substantial self either, then they might want to argue that the entity B is all that the moments of Bob's life really have in common. In this case, it seems natural (though daring) to ask whether B could be Bob's self. At first sight, this identification seems utterly implausible. I will now try to dispel this implausibility.

2. The Self as an Abstract Object

The intuition that a property is not the kind of item that could be a self is perhaps the main obstacle in the way of seriously asking whether B might be a self. But the feeling that B is "only a property," and hence inevitably not a self, may be quite misleading.

The idea that the self is an abstract object is not at all new. It is a well-established idea, with proponents as diverse as Plato [Plato, 79-80] and Dennett [Dennett 2, especially ch. 13]. (For a comparison of these two views and remarks on their ontological implications, see [Sharlow].) The accuracy of the feeling that a property could not be a self depends upon which view of the ontology of abstract objects is correct. If realism with respect to properties is true, then B is a real entity -- as real as a brain, or as real as Descartes believed mental substances to be, though perhaps belonging to a different ontological category from either brains or mental substances. If nominalism with respect to properties is true, then one cannot say that B is Bob's self without implicitly saying that Bob's self does not exist.

The debate over the ontological status of abstract objects is an old controversy with an extensive literature. (For an entry point, see [Armstrong].) I will not try to enter this debate here. Instead, I will suppose, just for the sake of argument, that properties are real

abstract objects -- not necessarily full-blown Platonic universals, but real and not merely convenient fictions. Then the property B is a real entity, albeit an abstract one. Once we assume that B is an entity and not merely a figure of speech, then the view that B is a self becomes less incredible.

The identification of the real entity B with Bob's self might be rather plausible if B had all of the characteristic features of the self as traditionally conceived. However, we might be able to identify B with Bob's self even if B lacks some of these features. We can consider this move if we are willing to admit the possibility that a self does not have all of the properties that people traditionally expect selves to have. (No present-day philosopher of mind will find this possibility novel.)

In this paper, I will temporarily ignore the possibility that the self is an extremely complex abstract object, such as Dennett's "Center of Narrative Gravity" [Dennett 2, p. 410 and ch. 13]. Instead, I will examine a simpler hypothesis: that Bob's self is the abstract object B. B is not the only abstract object that Bob's self might be. Ultimately, we might want to abandon B as a candidate for the self, and use a more complex abstract object (like Dennett's) instead. But for the time being, we will start with a simple model and see how it fares.

3. Bob's Self and Descartes' Ego

Let us find out how closely the abstract object B resembles a self of a traditional sort. To do this, we will compare B to a typical dualistic conception of the self. We choose a dualistic conception for this purpose, not because we favor dualism, but because dualists tend to allow the self most of the traits that prephilosophical thinking ascribes to the self. Materialism, on the other hand, tends to truncate or deflate these traits. Let us now compare B to the non-material self postulated by Cartesian dualism -- that is, by Descartes' dualism in the *Meditations*, or by brands of dualism close to Descartes' own dualism.

(Before beginning this comparison, I must emphasize once again that I am using

Cartesian dualism only for the purpose just stated. I am *not* arguing in favor of Cartesian dualism. I mention this in case someone reads this paper carelessly and claims that I am defending Cartesian dualism. Such critics are hereby dismissed to make room for the serious critics.)

Cartesian dualism posits a self that is non-material, undetectable to the senses, weightless, and arguably also placeless in the sense that it has no spatial location other than (perhaps) that of its body. Compare this self to the abstract object B. The object B is non-material, undetectable to the senses, weightless, and arguably also placeless in the sense that it has no spatial location other than (perhaps) that of Bob's body. Thus, the abstract object B resembles, in some crucial respects, the self posited by dualism. The chief prima facie differences between B and the Cartesian self are:

(1) The Cartesian self is involved in the causation of the subject's actions. B cannot play this causal role.

(2) The Cartesian self is conscious. B is not conscious.

(3) The Cartesian self is a "mental substance" -- an item somewhat analogous to a piece of physical stuff, but invisible, intangible, and lacking many of the other key properties of matter. A mental substance, thus conceived, seems to be a concrete object rather than an abstract object.

(4) The relationship between B and slices of Bob's history is one of instantiation. The relationship between the Cartesian self and the slices of its body's history is a relationship quite different from instantiation.

We will now cast doubt upon these supposed differences.

Difference 1: No causation?

The most problematic feature of Cartesian dualism is the nonphysical self's causal influence on the brain. The epiphenomenalists, whatever we may think of them, have taught us that one can deny this causal influence without denying dualism. Without the causal influence, humans still could have selves distinct from their bodies. These selves might determine the identities of persons through time, and (as in epiphenomenalism) might even be involved in the having of conscious experiences. Thus, B's failure to cause any effects would not automatically count against B's being a self.

If the self were an abstract entity like B, then the self would not cause the subject's actions as a Cartesian ego would. However, the self still would have a strong bearing on the subject's life. The continued presence of the property B is a *necessary condition* for the continuation of Bob's existence. If B ceases to be exemplified, then Bob is dead. Of course, the absence of B does not *cause* Bob's death, nor does the presence of B cause any of Bob's vital functions. The absence of B simply *implies logically* that Bob is dead. This implication holds because B is a property that belongs only to slices that are slices of Bob's life history. Even if the abstract object B causes no effects, the presence of B still is a necessary condition for the existence of the person Bob.

It seems obvious that B cannot cause any events in Bob's brain, because all of those events are caused by other physical events and not by some property like B. However, there is a way in which B is involved in the causation of Bob's actions, even if Bob's actions can be explained entirely by physical causes. Armstrong once pointed out that "When things act causally, they act in virtue of their properties. The object depresses the scales in virtue of its mass[...]." [Armstrong, p. 28]. This is the case even though the movement of the scales is caused by interactions between atoms in the object and atoms in the scales. Thus, a property of a macroscopic object can be deeply involved in the causation of an event, even if that event is caused entirely by microphysical causes, and even if the macroscopic property does not actually cause the event. Something similar happens with B. For any action A, if A is an action of Bob's, then A is an action of Bob's

in virtue of the property B -- a property exemplified, not directly by A itself, but by slices of Bob's body's history. This follows from the fact that no action of Bob's can originate in a temporal slice of a human body unless that slice possesses B. If a slice does not possess B, then no physical event originating in that slice is an action of Bob's. A temporal slice of a human body can be the locus of an act of Bob's only by virtue of that slice's having the property B, together with whatever other properties are necessary to make that slice the locus of an action.

This conception of the causal role of B has connections with the concept of top-down causation in the philosophy of mind. I will not explore these connections further here.

Difference 2: No consciousness?

At first glance, it seems quite plausible to suppose that B cannot play any role in consciousness. Certainly B is not the seat of the physical processes which give rise to conscious experiences. That honor must go to Bob's brain. However, B plays a role in Bob's consciousness that is much like B's role in Bob's actions. The property B is not a cause of consciousness, but certainly the presence of B is a necessary condition for the existence of *Bob's* consciousness. If a slice of the history of Bob's body lacks B, then Bob is not conscious at the time of that slice. (Indeed, Bob is dead at that time.) Also, if a brain is conscious, then the consciousness of that brain is not *Bob's* consciousness unless the temporal slices of the history of the body containing that brain instantiate B. Slices that do not instantiate B are not slices of Bob's life at all. Hence the presence of B, though not necessary for the existence of consciousness as such, is necessary for the existence of Bob's consciousness. If Bob is to have a conscious life, then B is a necessary ingredient of that life -- even if B causes nothing and the brain is the seat of consciousness.

One can adapt our earlier argument about the indirect causal role of B to build a case for an indirect role of B in the production of Bob's consciousness. A particular brain is the seat of Bob's consciousness in virtue of the presence of B together with other properties -- in much the same way that (in Armstrong's example) an object is able to tip the scales in

virtue of the object's mass.

Difference 3: No mental substance

Normally, one tends to think of a Cartesian ego as a "mental substance" -- a thing that is like a material object in some respects, but is undetectable by the senses, weightless, and devoid of many other attributes of matter. One tends to think of the dualist's world-picture as containing two kinds of stuff: matter and mind-stuff. However, the example of property dualism teaches us that dualism does not actually require this familiar mental picture. Suppose that instead of thinking of the dualist's self as a hunk of mind-stuff, we think of it as just an *entity* -- a real, existing item, but not necessarily made of any kind of "stuff." Switching to this new mental picture will not destroy dualism; it will only strip dualism of a nonessential feature. Once we make this switch, we are left with a dualistic self that is weightless, immaterial, undetectable by the senses, and arguably placeless, but that should not be thought of as being made of any kind of stuff at all.

A dualistic self of this sort is dangerously close to being an abstract object. Once we make the move from a traditional Cartesian ego to a dualistic self of this kind, the contrast between the dualistic and the abstract-object conceptions of the self begins to fade. Both the dualist and the proponent of an abstract self believe in a self that is nonmaterial, weightless, undetectable by the senses, arguably placeless, and so forth. Where is the real difference between these two conceptions of the self?

Difference 4: Instantiation, not influence

Someone might object that B cannot be a self because B is only instantiated by slices of Bob's life, instead of being connected to those slices in some other, more substantial way. However, once we have abandoned the Cartesian concept of the self as the cause of actions and of consciousness, this distinction becomes much less important. If the self cannot causally influence the body, then the connection between self and body becomes

rather thin anyhow. Quite possibly, instantiation could do the same ontological work as this connection.

4. From Dualism to the Abstract Self

It appears that the abstract object B has most of the central features of the Cartesian self. The exceptions are the Cartesian self's direct causal roles in the production of actions and of experiences. What, if any, are the other *important* differences between the dualistic self and the abstract "self" B? Is there any *important* contrast between a dualistic self and an abstract self, once we stop thinking of the dualistic self as able to cause events?

At the risk of taking an overly speculative position, I would suggest that the answer to the preceding question is "no." It appears that any truly significant contrast between the abstract self and the dualistic self has been lost. The thesis that the self is an abstract object gives a version of the self that is practically the same as a truncated version of Descartes' immaterial ego. If we are willing to abandon the highly problematical causal characteristics of the Cartesian ego, then we find that the view that the self is an abstract object amounts to a moderate, post-Cartesian form of dualism.

This suggested view of the self can be regarded as materialistic, because it allows for a materialistic explanation of mind and requires nothing but physical objects and physical properties in its ontology. On the other hand, this view could fairly be regarded as dualistic, because it portrays personal existence as a phenomenon involving a linkage between a person's body and an immaterial entity. Hence this view is, in a sense, both materialistic and dualistic -- but without contradiction. One could call this view either "abstract object materialism" or "abstract object dualism" if one wanted. In a moment I will suggest a less committal name.

Central to this view is the doctrine that abstract objects are real -- that properties and similar items actually exist, and cannot be reduced entirely to figures of speech. Without this, this view of the self would collapse into a nihilism with regard to the self. Thus, the

new view depends upon ontological realism, although it certainly does not require a strong kind of ontological realism like Plato's. Because the view that the self is a really existing, non-fictional, abstract object is a special case of ontological realism, I would suggest the name "abstract-self realism" for this view.

This view may seem radical at first glance. Actually, it is no more radical than any other philosophical idea that depends upon the reality of abstract objects -- for example, Fregean semantics. The view that the self is an abstract object has extensive precedents in the philosophy of mind. As I mentioned earlier, thinkers as diverse as Plato and Dennett have either embraced it or come close to it. The view proposed here is merely a further development in this familiar direction.

In closing, I wish to suggest a variation on Descartes' famous dictum "I think, therefore I am." If the self is an abstract object and the right form of ontological realism is true, then each of us (including Bob) has the right to declare "I think, therefore I am an abstraction -- and therefore I am."

References

[Armstrong] Armstrong, D. M. *Universals: An Opinionated Introduction.* Boulder, CO: Westview Press, 1989.

[Dennett 1] Dennett, Daniel C. "Real Patterns." *The Journal of Philosophy*, vol. 88, no. 1 (Jan. 1991), pp. 27-51.

[Dennett 2] Dennett, Daniel C. *Consciousness Explained*. Boston: Little, Brown, 1991.

[Descartes] Descartes, René. *Meditations.* In *Discourse on Method and Meditations*, trans. Laurence J. Lafleur. Indianapolis and N.Y.: Bobbs-Merrill, 1960.

[Hirsch] Hirsch, Eli. *The Concept of Identity*. N.Y. & Oxford: Oxford University Press, 1982.

[Plato] Plato. *Phaedo*. In *Plato's Phaedo*, trans. R. Hackforth. Reprint ed. (original printing 1955). Indianapolis and N.Y.: Bobbs-Merrill.

[Sharlow] Sharlow, M. F. "Platonizing the Abstract Self." Preprint, 2004.

Mind Is to Brain as Digestion Is to Digestive Tract. Oh, Really?

There is an old philosophical chestnut that says that the mind is to the brain as digestion is to the digestive tract. The underlying thought is clear: why should we regard the mind as something "special," over and above the brain, when we wouldn't regard digestion as something over and above the digestive organs?

The best reply to this chestnut is simple but surprising: digestion *is* something over and above the digestive tract. Your digestion - what you refer to when you say things like "I have a slow digestion" or "my digestion is good today" - is not merely part of your digestive tract. Instead, it is a *feature* of your digestive tract. It is what philosophers call an abstract entity. A feature of a thing is not identical to the thing. Thus, your digestion is not identical to your digestive tract - for the same reason that the mass of an electron is not the same as an electron, or that the shape of a window is not the same item as the window.

The reason the digestion-digestive tract difference is unlike the mind-brain difference is that nothing interesting follows from the digestion-digestive tract difference. The fact that the digestion is different from the digestive tract doesn't tell us anything new about the nature of digestion or of ourselves. It tells us no more than we already know when we admit that the shape of a window is not identical to the window. It is a near-trivial logical fact.

However, in the case of the mind (which is a feature or set of features of the brain), the difference between mind and brain *does* imply something interesting. Unlike digestion, the mind is associated in a distinctive way with a large domain of other abstract entities. These other entities are the contents of consciousness, which make up what we think of as our inner world. The fact that we possess this inner, abstract "world" has a drastic bearing on who we are as individuals and as a species. It makes the difference between a conscious observer and a mere nonconscious thing. Once we face the fact that this inner world exists, we realize that minds and selves are not just lumps of matter, even if they are only features of the brain. What is more, we cannot understand the mind without taking the inner world into account. If we ignore the contents of consciousness, we miss what is most essential to the mind.

With digestion it is different. Once we know the physical mechanisms of digestion, there is essentially nothing left to understand about the nature of digestion. Even if we admit that digestion is something distinct from the digestive tract, this fact doesn't help us understand digestion. We learn no more that way than we already knew when we realized that the mass of an electron is not an electron, or the color of a stone is not the same as the stone. The distinctness of digestion from digestive tract is, as I have said, a near-trivial logical fact. However, if we don't pay attention to the complex abstract features of the brain (specifically mental contents), then we don't really have any idea of what a mind is. We miss the important aspects of the mind.

This, in brief, is why the old analogy between digestion and mind fails.

The same argument works against any analogy that says "Why should I think my mind is distinct from my brain, when my [fill in name of body function] isn't distinct from my [fill in name of organ]?" The analogy fails for the same reasons.

Qualia and the Problem of Universals

Mark F. Sharlow

ABSTRACT

In this paper I explore the logical relationship between the question of the reality of qualia and the problem of universals. I argue that nominalism is inconsistent with the existence of qualia, and that realism either implies or makes plausible the existence of qualia. Thus, one's position on the existence of qualia is strongly constrained by one's answer to the problem of universals.

1. Introduction

The question of the reality of qualia is one of the most contentious issues in the philosophy of mind today. This question has many angles and facets; I will not try to summarize all aspects of the question here. Instead, I will point out a little-noticed feature of this question: its dependence upon another, seemingly unrelated, philosophical problem. Specifically, I will contend that the answer to the question "Do qualia exist?" depends upon the answer to the *problem of universals*. This latter problem is a very old philosophical issue which, at first glance, would seem to have little to do with present-day debates about the nature of consciousness.[1] I will argue that the solution that one adopts for the problem of universals constrains, and may even completely determine, the positions that one consistently can take with regard to the existence of qualia. One's answer to the problem of universals may force one (logically speaking) to deny the existence of qualia, to accept the existence of qualia across the board, or to hold that some kinds of qualia exist and others do not. Since most arguments for or against qualia make no contact with the problem of universals, it is likely that most such arguments leave out something important.

2. Qualia and the Nature of Properties

If qualia are anything, they are properties. Some philosophers tend to speak as though qualia are features of experiences, or else features of mental states or of brain states. (See, for example, [Dennett 2, pp. 17 and 373].) Other philosophers [Lewis, pp. 121-3] speak as though qualia are properties that exist *within* experience; this amounts to saying that qualia are properties that things seem to have. In either case, qualia are portrayed, either explicitly or implicitly, as properties. Some proponents of qualia might find that their conceptions of qualia do not quite fit into this mold. In particular, there is a tendency to

speak of qualia as if they were much like sense-data, traditionally conceived, and hence more like things or events than properties. But even if one speaks this way, one can think of qualia as features that things seem to have, and hence as properties of sorts. If this view is correct (and I will assume for now that it is), then the answer to the question "Do qualia exist?" depends in part upon the answer to the question "Do properties exist?"

The problem of universals, in simplest terms, is the question of whether multiply exemplifiable abstract entities, such as properties and relations, are real. Since qualia are properties, the set of positions that one consistently can hold with regard to the existence of qualia depends upon one's position on the existence of properties.

If you are a strict nominalist, and do not believe that properties exist at all, then your ontology cannot include qualia -- no matter how much you might want to believe in qualia for other reasons. Your picture of reality cannot include items like qualia, for the same reason that it cannot include items like squareness and tallness. Of course, you can speak of all these items, but you must understand your statements about them as mere figures of speech. When you speak of qualia, you cannot really mean to imply that qualia exist. This puts you in the same position as the most ardent opponents of qualia, who, while denying qualia, sometimes concede that one can speak as if qualia existed. If you are a nominalist, and if qualia are properties, then you must be an opponent of real qualia, on pain of inconsistency.

Someone might object to the preceding argument on the grounds that I have misconstrued the problem about qualia. According to this objection, the claim that there are qualia does not really commit one to the existence of qualia as real abstract objects. Instead, the objection goes, a philosopher of mind who claims that qualia exist is using "exist" in an ontologically noncommittal sense, much as a physicist speaks of mass without taking a position as to the reality of properties. But this objection ignores the nature of the controversy over the existence of qualia. Philosophers of mind who debate the existence of qualia are trying to determine what *really exists* in the machinery of the mind. They are not merely arguing about whether we can speak *as if* there are qualia (though that question may concern them as a side issue). They are not like the physicist, who

speaks freely of specific properties and relations and leaves it to others to determine whether properties and relations in general are mere figures of speech. Determining what entities are involved in consciousness is part of the project of the philosophy of mind -- and this determination is an ontological project. A philosopher of mind who says that there is no Cartesian ego *really means* that there is no Cartesian ego -- that there is nothing in reality that answers to the concept of a Cartesian ego. A philosopher who says that there is no Cartesian ego is not using "is" prephilosophically, as is a physicist who says that there is mass and there is time. There is no way around the fact that the philosopher really is denying the existence of a purported entity. Similarly, opponents of qualia really are trying to say that there are no qualia -- that if we inventoried everything in reality, we would not find qualia in stock. (At least this is the way I read those authors.) And this is exactly what a nominalist must say about qualia. Thus, a nominalist must be an opponent of the existence of qualia.

If, on the other hand, you are a realist with respect to properties, then it is consistent for you to hold that qualia exist. Being a realist does not, in itself, force you to believe in qualia. However, there are many different kinds of realism -- and if you believe the right kind of realism, it may be obligatory for you to believe that qualia exist, on pain of inconsistency. I will argue for this last claim in later sections of this paper.

Before going further, I wish to point out that there are two different kinds of property that a quale might be. I alluded to these two kinds a few paragraphs ago. Now I will make them more explicit:

(1) One might regard qualia as features of conscious experiences. For example, one might think of an experience of seeing green as possessing a special, subjective, phenomenal feature; this feature is the quale of green. Philosophers of mind sometimes talk about qualia in this way, as features of experiences -- or, alternatively, of mental states or brain states. (See, for example, [Dennett 2, pp. 17 and 373].) If one thinks about qualia in this way, then qualia are properties that conscious experiences (or mental states or brain states) can have.

(2) On the other hand, one might think of qualia as properties *found in* experience -- as qualities that are presented to us in our conscious experiences. (See, for example, [Lewis, pp. 121-3].) People often talk as if qualia were like this -- for example, when people speak of a perceived shade of green as a quale. According to this way of thinking about qualia, a quale is not simply a property that belongs to conscious experiences. Instead, a quale is a property that *seems to be exemplified*. For example, if one has a subjective experience of a shade of green, then this shade of green is a property that part of one's visual field seems to have. The old distinction between "physical color" (in the world) and "psychological color" (in the mind) reflects this way of thinking about the phenomenal aspects of experience. This way of thinking about qualia amounts to the categorization of the subjective qualities found in experience as properties that seem to be exemplified.

(I wish to emphasize that these two descriptions of qualia represent nothing more than two different conceptions of qualia. I am *not* claiming that there are two different kinds of qualia. Presumably, if there are any qualia at all, then all qualia are of only one of these two kinds.)

The impact of the problem of universals on the question of qualia depends upon which of these two understandings of qualia is correct. I will explore this impact separately for each of the two versions of qualia.

3. Qualia as Contents of Experience

Let us begin by examining the second conception of qualia: that of qualia as properties that seem to be exemplified. If qualia are like this, then the answer to the question "Do qualia exist?" hangs on the answer to the question: "Can a property that only seems to be exemplified exist?"

If you are a realist, then your answer to this question depends upon what kind of

realist you are. Some realists hold that a property exists only if it is exemplified. Other realists maintain that there exist properties that are only *possibly* exemplified. (See [Armstrong, pp. 80-81] for a brief discussion, and rejection, of the latter view.) According to this latter view, not every property P satisfies "For some x, x has P." Instead, there are properties which satisfy "Possibly, for some x, x has P," but which do not satisfy "For some x, x has P." Proponents of this view holds that the actual instantiation of P is not necessary for the existence of P; instead, the existence of P only requires the instantiation of P within (so to speak) the scope of a specific modal operator. Now take note of the following fact: "It seems that" is a modal operator every bit as legitimate as "Possibly." (See [Sharlow 1, pp. 53-54].) The phrase "It seems that" introduces an intensional (with an "s") context, just as do other modal operators like those of possibility, necessity, belief and knowledge. Once one accepts that a property may be only *possibly* exemplified and still exist, then it is not too implausible to suppose that a property may be only *seemingly* exemplified and still exist. In both cases, the property "exists" only within a modal context of some sort -- within a merely possible world (in the case of possibility), or a merely apparent situation (in the case of seeming). Indeed, in many cases, a property that *seems* to be exemplified is a property that *possibly* is exemplified. (One obvious example would be the property of being a pink elephant -- assuming that someone actually is hallucinating such an animal.)

Of course, a realist who believes that possibly exemplified properties are real does not have to believe, across the board, that seemingly exemplified properties are real. However, it may be difficult to successfully defend one of these beliefs without inadvertently providing at least some support for the other.

A realist who believes that only actually exemplified properties are real is stuck with the conclusion that there are no qualia of kind (2) as defined in Section 2 -- with the potential exception of qualia that are really, not just seemingly, exemplified by physical objects. In view of the fact that qualia are phenomenal, psychological qualities, we can safely assume that this exception is empty. (A physical object can exemplify physical green; how can a physical object, apart from our experiences of it, actually exemplify

psychological green, or the *feel* of green?) Thus, a realist of this kind can only believe in qualia of kind (1). As I will show in the next section, such a realist may actually be compelled to believe in qualia of kind (1).

4. Qualia as Properties of States or Experiences

Qualia of kind (1) are properties exemplified by conscious experiences -- or, depending upon the details of one's account of qualia, by mental states or by brain states. In this section, I will argue that some kinds of realism require us to believe in properties that possess the most important features of these qualia. I will begin by defining a particular property which has some of the salient features of a quale of kind (1). I will define this property as a property of brain states. Those who prefer to think of qualia as properties of experiences, mental states, or something else can change the definition and the argument accordingly.

Let us call a state of a human brain a *b-state* if any person whose brain is in precisely that state would experience the color blue. If the phenomenal aspects of a person's experiences are uniquely determined by the goings-on in that person's brain (as materialists, at least, probably must believe), then the set of b-states is well-defined.[2] One can foresee certain potential objections to the concept of a b-state. I will try to dispose of two of these objections before I proceed.

(Objection I) Many materialists (especially followers of Dennett) might argue against the concept of a b-state on the grounds that the content of experience is not determined by a single, temporally sharp state of the brain. (Dennett's theory in *Consciousness Explained* implies there generally is no fact of the matter about the content of an experience until after the putative time of the experience has passed; see [Dennett 2, pp. 134-6].) This objection misses the mark because our definition of b-state does not require a b-state to be a state of an instantaneous temporal slice. To accommodate Dennett, we might have to take the b-states to be non-instantaneous

states, defined at stretches of history instead of at single slices. There is nothing suspect in this idea of a non-instantaneous state. This idea of a state makes sense for the same reason that a news commentator can make sense when speaking of "the state of the world over the past year."

(Objection II) A strict behaviorist or an eliminativist, who does not believe in experiences at all, might argue that we should not say that humans experience blue. However, such a person still could say that most human brains sometimes go through states which, in everyday language, are called states of experiencing blue. No harm is done to anyone's position if we adopt these states as our b-states. Thus, we can define b-states in such a way that even behaviorists and eliminativists must accept that humans sometimes are in b-states.

With these two objections out of the way, we continue the argument. Realists typically hold that things exemplify a common abstract object if those things have what prephilosophical discourse calls a "common feature". (For example, all square things exemplify the abstract object which can be called squareness.) Certainly, all b-states have a common feature; they have something definite, and most significant, in common. One might wonder what, if anything, the b-states have in common with one another at the neurophysiological level. However, this last question does not matter to our present argument, since the b-states, as we defined them, certainly have something in common at a behavioral level at very least. Hence if realism is right, then all of the b-states exemplify a common abstract object. Call this abstract object Q.

According to the view (1) of qualia (with brain states as the items that exemplify qualia), the quale of blue is a property shared in common by all and only those brain states that are associated with the experiencing of blue. Q is a property of this sort. What is the difference between Q and the alleged quale of blue?

One apparent difference is that the quale of blue is "given" in experience; that is, the subject has access to the quale. However, the subject also has a kind of access to Q.

Certainly the subject, if able to see blue and if possessed of ordinary self-awareness, can respond selectively to those states of his/her brain that have Q. (People who see and talk about blue things do this all the time.) If the subject learns the definition of Q, then the subject can say when his/her brain is in a state having Q. Thus, the subject has a degree of access to Q, as to the quale.

A major philosophical issue related to qualia is the issue of the first-person character of consciousness. It is worth noting that the property Q has a kind of first-person accessibility. If you have first-person access to the fact that you are seeing blue, then you also have access, in a slightly less direct way, to the fact that your brain states have Q. Of course, you cannot tell exactly what your brain is doing, but if you already know that your brain is in a state with Q if and only if you are experiencing blue, then you can tell at once that your brain is in a state with Q. Also, if there are facts about your experiences of blue that are third-person *in*accessible (as some philosophers would claim), then there are facts about Q that are third-person inaccessible as well. For example, if no one but you can tell how it feels to you to see blue, then no one but you can tell how it feels to have a brain state that has Q. Thus, Q has a kind of first-person accessibility, and also has a kind of third-person inaccessibility provided that experiences of blue have third-person inaccessibility.

Another objection to identification of Q with the quale of blue arises from the conviction that a quale is something "immediate" for consciousness -- not an abstract, and abstractly defined, property like Q. But this feeling is groundless, since a quale, as ordinarily understood, already is an abstract object, and in fact is defined in a rather abstract way. (To convince yourself of the last point, think carefully about the meaning of the familiar phrase "phenomenal qualities of experience.") If we refuse to accept that Q is a quale, then we have simply refused to identify one abstractly defined, first-person accessible property with another.

Still another objection to the identification of Q with the quale arises from the view that qualia must be something in addition to the physical substance of the brain. One is most likely to hear objections along these lines from dualists, although advocates of

emergent properties might pose similar objections. However, these objections lack force for a realist, for the following reason. Properties, on the realist view, have an existence of their own, distinct from the existence of the things that exemplify them. This is the case even for properties that have reductive scientific explanations. (For example, metallic objects exemplify the abstract property of metallicness, even though the metallic behavior of matter has a reductive scientific explanation.) If realism is right, then Q is an abstract entity distinct from the material of the brain -- for the same reason that metallicness is an entity distinct from a piece of metal. Thus, Q is, in a sense, "above and beyond" the physical substance of the brain. (One emphatically should *not* read this last statement as an assertion of substance dualism.)

Does the property Q have all of the features that qualia should have? The answer depends upon the details of one's understanding of qualia; some people may have to give up some of their ideas about qualia to see Q as a quale. However, Q has enough of the central features of qualia to make the identification of Q as a quale relatively strain-free.

We conclude that a realist should believe in qualia, or at least in properties having the most central features of qualia, if qualia are properties of brain states. The same conclusion would follow if we had taken qualia to be properties of mental states or of experiences.

5. Concluding Remarks

The arguments in this paper suggest that the answer to the question "Do qualia exist?" is strongly constrained by the answer to the problem of universals. Whether you can believe consistently in qualia depends upon what you believe about the ontological status of abstract objects.

In a sense, the problem of the ontological status of qualia is *parasitic* upon the more general problem of the ontological status of properties. This statement is not meant to demean the problem of the status of qualia, but simply to emphasize the fact that the answer to the first problem depends upon the answer to the second. If certain kinds of

realism are true, then qualia are real, period -- regardless of how much the qualophobes may inveigh against them. If nominalism is true, then qualia simply are not part of the ontology of the world, regardless of any arguments put forth by qualophiles. Certain kinds of realism allow for qualia, but place constraints on what qualia can be. Most importantly, if you refuse to take a position on the ontological status of properties, then you will have no firm grounds for claiming either that qualia exist, or that they do not exist.

These conclusions have broad implications for the current debate over the ontological status of qualia. In particular, our conclusions suggests that certain lines of argument about qualia miss the point. Particularly suspect are those anti-qualia arguments that deploy baskets of facts about brain function, computation, perceptual oddities, etc. to debunk qualia (for example, in [Dennett 2]). If certain kinds of realism are true, then qualia are real, period, regardless of the details of how the nervous system works. The status of abstract objects has a much greater bearing on the reality of qualia than do the details of neuroscience; the question of the scientific utility of qualia has very little bearing here. The problem of the existence of qualia is irreducibly an ontological problem -- or perhaps even a logical problem. To find a solution to this problem, one must rely, at least in part, upon old-fashioned ontological analysis. Scientific facts, even when supplemented with some science-based philosophical reasoning, are not sufficient to do the job.

Notes

1. Within the context of the debate over consciousness, Dennett (in [Dennett 1]) has addressed one aspect of this problem and has proposed what he calls a "mild realism" with regard to patterns. Despite appearances, this has little to do with the traditional problem of universals; see my critique in [Sharlow 2].

2. I realize that I am ignoring a number of issues about supervenience here. What I am asserting here does not, I think, require a resolution of these issues.

References

[Armstrong] Armstrong, D. M. *Universals: An Opinionated Introduction.* Boulder, CO: Westview Press, 1989.

[Dennett 1] Dennett, Daniel C. "Real Patterns." *The Journal of Philosophy*, vol. 88, no. 1 (Jan. 1991), pp. 27-51.

[Dennett 2] Dennett, Daniel C. *Consciousness Explained*. Boston: Little, Brown, 1991.

[Lewis] Lewis, C. I. *Mind and the World-Order.* London: Charles Scribner's Sons, 1929.

[Sharlow 1] Sharlow, M. F. *From Brain to Cosmos.* Parkland, FL: Universal Publishers, 2001.

[Sharlow 2] Sharlow, M. F. "Platonizing the Abstract Self." Preprint, 2004.

Rethinking Wholes and Parts:

Reflections on Reduction, Holism, and Mereology

Mark F. Sharlow

Rethinking Wholes and Parts: Reflections on Reduction, Holism, and Mereology

The excerpts in this document are from the ebook version of the author's book *God, Son of Quark* (© 2008 Mark F. Sharlow). A preliminary print edition of the book was released in 2006.

About This Ebook

This ebook is a set of excerpts from one of my early books. In these excerpts I discuss the relationship between whole objects and their parts, with special attention to the type of reductionism that claims objects are “nothing but” their parts.

There are two page numbers at the bottoms of most pages. The upper number is the page number in the original book; the lower number is the page number in the present document.

Comments added after the publication of the original book are in blue and in a different font.

- MFS, 2011

[First paragraph omitted]

Much of the work I will do in this book consists of developing, and supporting with evidence, some new ideas about wholes and parts. These ideas differ from the ones that we usually use when thinking about objects and their parts. The new conception of whole and part that I will present requires us to think about material objects in a way that is slightly unfamiliar. This does *not* mean that I will propose any new scientific theories about the nature of matter. Everything I say will be compatible with existing scientific theories and facts, and with any reasonable new theories that may someday replace the existing ones. What I am going to propose is not a theory of matter or of ultimate particles, but a new view of the *logic of the whole-part relationship*. This view will not be a sweeping theory about what material objects "really are." Instead, it will be an attempt to overthrow certain long-standing ways of thinking about wholes and parts, and to replace those ways with new concepts that may lead to less confusion.

The new view I am proposing may seem unfamiliar. It does not always agree with our everyday thinking about wholes and parts. Yet despite its novelty, this view has many points of contact with previous philosophy. Some of the pieces of the new view already exist in the philosophical literature. A few of the most important ideas

appeared earlier in the work of Donald L.M. Baxter[1] and of David Lewis[2]. Some of the ideas that I will explore in later chapters began in the writings of ancient Greek philosophers, especially Aristotle. Most of my credits to previous authors are in the book's many numbered endnotes, though a few of these credits are in the text. (The spirit of Baxter's and Lewis's approaches to whole and part has influenced the book more than specific credits can show.)

To begin the project, I will point out some of the intuitive beliefs that people normally hold about objects and their parts. Usually we do not think about these beliefs. These usual ideas may play an important role as background to our actions, but they seem so transparent and obvious that we do not reflect on them consciously. Here I will try to bring these ideas into the light, and will suggest that some of them are wrong despite their "obviousness." While doing this, I will lay the groundwork for a new view of the relation between wholes and parts. Although I will offer arguments for this view, the main argument in its favor is its impact on other topics. Once this view is in place, several extremely knotty philosophical issues will become much less tangled.

[1] Baxter, "Identity in the Loose and Popular Sense."

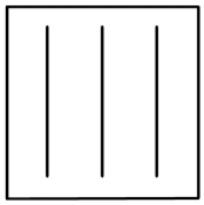
[2] Lewis, *Parts of Classes*.

Three Thought Experiments[3]

To kick off this project, I will present three *thought experiments*—that is, experiments performed in thought instead of in a laboratory. These experiments contain nothing dramatically new. They use familiar objects and actions; they even lead to the outcomes that you would expect. (Only the last experiment needs any scientific background, and I will try to provide that on the spot.) But despite the ordinariness of these experiments, when you think about their outcomes in the right way, you will see that these "ordinary" results are not so ordinary after all. The outcomes of these experiments run counter to some of our commonsense views about wholes and parts—and suggest that those views leave out something important.

Experiment 1. The Interloping Triangle

Think about this box with some line segments inside:



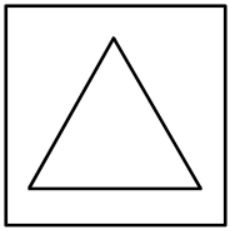
[3] This section informally introduces several new ideas, including some general ideas about the whole-part relationship discussed (either favorably or unfavorably) in the previous literature. See in particular: Lewis, *Parts of Classes*, pp. 81-87, and Baxter, "Identity in the Loose and Popular Sense." Also, I should credit Minsky, in *The Society of Mind* (p. 27), for mentioning the question "*What makes a drawing more than just its separate lines?*"

Now answer the question: "How many things are in the box?"

To simplify this task, forget about the fact that the line segments are divisible. Just count things consisting of at least one whole line segment. (A mathematician might say, "Since a line segment contains an infinite number of points, there are an *infinite* number of things in the box." But that isn't the answer we are after here.) Also, don't count sets of detached items, with no point of contact, as "things." (Usually we wouldn't think of a pair of unconnected lines as a thing or object.) Just count the things in the box in a naive, intuitive way—that is, count only complete, internally connected things.

If one follows these precautions, the answer to the question is obvious. It is "three."

Now rearrange the line segments a bit, without adding anything at all to the contents of the box.



How many things are there in the box now?

Again, we ignore partial line segments and count the things in the box. Clearly, the three original line segments are present; they now touch one another, but they haven't

gone away. We also notice that there is a *triangle* in the box. A triangle is as legitimate a geometric figure as a line segment, as anyone who has studied geometry knows. It would be silly to count the line segments as things, and then to refuse to count the triangle (which is just another whole plane geometric figure made of points!) as a thing. To avoid such arbitrariness, we count each line segment, and then count the triangle.

Counting in this way, we decide that *there are four things in the box.*

It would be more correct to say that there are *at least* four things in the box. One can argue that there are things in the box besides the line segments and triangle. For example, any two adjacent sides of the triangle make up a V-shaped figure, and these V's, though parts of a triangle, are themselves legitimate geometric figures. But this is beside the point. The point is that there are *at least* four things in the box. We have gained an object—the triangle—that was not there before the rearrangement. *Yet we put nothing new in the box.* We brought a perfectly legitimate geometric object into existence by arranging the line segments in a suitable way. And we did it using *nothing but* the line segments. The triangle has no parts in it above and beyond the lines.

Of course, there is nothing mysterious about this outcome. No magic trick has happened here; we did not pull the triangle rabbit-style out of a hat. Everyone knows that when you arrange line segments as we did, you get a triangle. This is obvious because a triangle is just a figure formed from three lines arranged in a specific way. We placed nothing new into the box—yet we were able to get a new object in the box. Although this new object has the

line segments as parts, *the new object was not there at the beginning of the experiment.*

Obviously, we did not bring anything new into the box. But this experiment also brings out another fact, equally obvious but less often noticed: *when we arrange parts to make a whole, we don't just end up with the original parts. A new object comes into being.* Normally, we might dismiss the idea that anything really is "created" here. We might do this by saying that the triangle is only an assembly of line segments. And this statement is correct: the triangle indeed has no parts beyond the line segments—except, of course, the three V shapes that the line segments form (and the parts, which we decided to ignore, that we can get by subdividing line segments and V's). But we also can shift the emphasis, and note that by rearranging the line segments, we can create a real geometric object that did not exist before. We can create a real entity—a fourth entity—simply by arranging the entities that already are there.

The triangle is not among the entities that we had at the beginning of the experiment. It can be counted separately and distinguished from the line segments; it has its own unique properties. It is a new item created by the assembly of the lines. The fact that the triangle is made up entirely of line segments does not change the fact that the triangle is *new*. It is not any of the line segments. Nor is it all the line segments together (note that all the line segments existed together *before* there was a triangle). By rearranging the line segments, we have managed to create a whole new object, without having to add any new "stuff" to the box!

People sometimes have a feeling that a composite

object, like the triangle, is "nothing but" its parts in a certain arrangement.[4] In a way, this is true; the triangle has no parts *but* the line segments, the V shapes (made of line segments), and the parts that we get by dividing up and combining these parts. Also, the properties and relations of the line segments may, for all we know, completely determine and explain the properties of the triangle. But this "explainability" of the triangle in terms of its parts does not do away with the arithmetical fact that there is a *thing* in the box that was not there at the beginning. If you don't believe it, count. Arithmetic and logic tell us that *there is something in the box besides the line segments.* This conclusion is inescapable once we grant some rather simple facts of plane geometry.

The lesson of this experiment is that when parts come together to form a whole, that whole is an object distinct from the parts.[5] The parts may explain the whole, yet one

[4] David Lewis has taken this position, or one close to it, in his philosophical writings. In *Parts of Classes* (pp. 81-87), Lewis argues that a whole just *is* its parts—in a slightly extended sense of the word "is." Donald L. M. Baxter, in "Identity in the Loose and Popular Sense," has discussed the view that the whole is identical to the parts—which he calls "the Identity view"—and has compared it to other competing views of whole and part. The position that Baxter calls "the Non-Identity view" is essentially the view of whole and part that I am advocating in this book, though I will take this position several steps further. Later in the book I will argue against some of the "whole-is-parts" ideas.

[5] Bertrand Russell stated a similar thesis about part and whole, though he was thinking of certain mathematical and logical senses of "whole." See Russell's *The Principles of Mathematics,* par. 137 (p. 141).

can count the whole separately from the parts. Normally we think of a whole as being, in some vague sense, "nothing but" the parts that make it up. That is, when we arrange the parts and hook them together properly, those parts are the whole. But this is not quite true. It is more correct to say that there are all the parts, and also there is the whole. To borrow a comparison from Baxter, if there are N parts, then once we build the whole from those parts there are N+1 things, not just N things.[6]

Are we right to think of the whole as nothing but its parts? Would it not be better to think of the whole as a *new* object, whose existence depends on the existence of the parts but which is not the same as the parts? Shouldn't we think of the whole as an object *brought into being* when the parts are hooked together the right way?

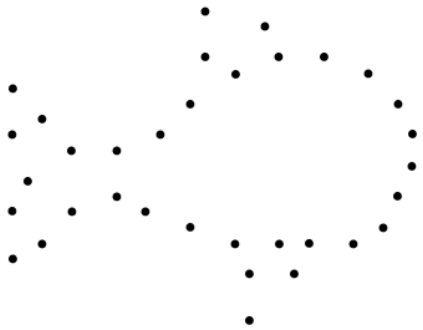
Or should we just look for a new way to count?

Experiment 2. Follow the Dots

Many readers will remember the "follow the dots" pictures on which they worked as children. Despite their simplicity, follow-the-dots pictures can teach us an important lesson about wholes and parts.

Examine the follow-the-dots picture on the next page.

[6] This observation is discussed in Baxter's article "Identity in the Loose and Popular Sense." See p. 579 in that article.



The objective of the puzzle is to draw lines connecting the dots, and end up with a drawing of a recognizable object. (Don't do it yet, though.) Normally, follow-the-dots pictures have numerals that show which lines to draw first. Here I have omitted the numbers to avoid cluttering the picture—and because the lines to draw are pretty obvious.

I want to ask a question about this picture. The question is: "Is there an outline of a fish there?" The answer is obvious: Yes, there is. Most people can see the fish immediately.

Even without thinking about which creature the diagram resembles, you can say that the diagram shows a geometric figure, or shape. This much is clear. But when you look more closely at the picture, *there's nothing there but dots.* Indeed, if you viewed the picture in the right way (from very, very close in), you would find that there are only dots there, and no fish. An ant crawling on the page could not see the fish. At any given time, it would see only a dot. Even if the ant somehow got the ability to think and reason, it would not be able to see the fish. One can imagine what the picture would look like to a human

observer with a severe case of tunnel vision, whose visual field is only big enough to scan one dot at a time. Such an observer would not see the fish, but would be perfectly capable of seeing all the dots (one at a time).

It seems clear that there is nothing on the page above and beyond the dots. What does this say about the fish?

"There are only dots on the page. Since there are *only* dots on the page, and nothing else, it follows that there is no fishlike pattern there at all. Therefore, the fish design is not really there." Are you willing to stand up for this argument? If so, are you willing to stand up for it *in public*?

The geometric design—which appears fishlike to most of us but is just a geometric figure—really is there on the page. To say that it is not there is to say something plainly false. Yet at the same time, it seems correct to say "there is nothing on the page other than the dots." We have arrived at an intriguing pair of seeming truths. There is a fish design on the page—yet there are only dots on the page. Is there a contradiction here?

The obvious answer to this "contradiction" is that the fish is made entirely of dots, so there is nothing special about the fish being on the page even though nothing is there besides the dots. Of course, this answer is right. But despite being right, this answer is rather fishy. This trivially true answer serves to hide an important fact about the fish and the dots. This is the fact that when the dots are put together to form a fish, there *really is* a fish design on the page. There is another recognizable physical object, made of ink, *that is not the same as any of the dots.* By putting together 34 dots, which are simple objects, we have created a new, more complex object—*a thirty-fifth thing.*

This thirty-fifth thing is made entirely of dots. Its presence on the page can be explained by the presence of the dots at certain positions on the page. Despite all this, the thirty-fifth thing really is there—and it is not one of the dots. Before the dots were drawn on the picture, there were no things in the picture. But after someone drew the thirty-four dots, there were thirty-five things in the picture (or even more things, if one counts the fish's fins and other such pieces of the drawing, which are things made of dots).

Of course, the fish is made solely of dots. Once the dots come together in the right pattern, the fish is there. Nothing else is needed to make the fish come into existence. No extra spark, no imposed property of "fishiness" or of "paternness," must be added to make the dots into a geometric pattern that looks like a fish. The dots can give rise to those properties by themselves, without any outside help. Nor does the fish have any mysterious extra parts, such as extra dots hidden on the other side of the page. Nevertheless, once the dots come together, a fish design begins to exist. If the dots were separated and scattered, the fish design would cease to be. And as long as the picture on the page remains intact, *there is something in that picture besides a dot.* There is the fish. If you don't believe that, just count.

It is possible, of course, to claim that this argument is misleading because, after all, the fish is only the composite of the dots. Why should we worry about the fact that there is a thirty-fifth object on the page when that object is *only* the sum total of all the neatly arranged dots?

If you feel an urge to argue in this way, think carefully about what you are saying. You say that the fish is only the sum of the dots—or, to use other words, the composite, or

assembly, or whole, formed by the dots. And what does "the sum of the dots" mean? If it just means all the dots, then you are stating a falsehood. The fish is not any of the dots, nor is it simply all the dots collectively (a scissors can make the dots exist without the fish). But if "the sum of the dots" means something *besides* "the dots" or "all the dots together," then you are admitting that there is *something else* on the page besides the dots. In that case, what you are calling "the sum of the dots" is the same thing that I have been calling "the thirty-fifth object." You have simply given the extra object another name without making it go away.

This experiment, like the previous one with the triangle, reveals a fact we already know. This is the fact that when things are put together to form a larger whole, the whole is itself a thing. No fact seems more trivial and less noteworthy. Yet if we begin to think about this fact instead of just taking it for granted, we also begin to see how puzzling this fact really is. Here are three lines; rearrange them, and now there are four things. Here are thirty-four widely separated dots; rearrange them, and now there are thirty-five things. By rearranging existing things, *we bring new things into being.* We literally create new objects. And those new things exist *in addition to* the things we started with. *The act of assembling parts is a genuine act of creation.*

Normally, we might feel that because the fish is made of dots, we do not need to assume the existence of anything but dots to understand what the fish is. This example suggests a different view: we cannot fully understand what the fish really is without assuming the existence of *the fish*

itself, as well as the dots. If we took an inventory of all things that really exist, we would find both the dots and the fish on our list. Listing the dots would not excuse us from listing the fish itself on a separate line of the list—for the existence of the dots is not equivalent to the existence of the fish.

Of course, the fish is made of the dots, and the fact that the fish exists is "explained" entirely by the existence and arrangement of the dots. Normally, we take this to mean that the fish is, in some sense, "just dots." But should we think this way? "All the dots" means thirty-four things. The fish is the thirty-fifth thing. When we arranged the dots, the universe proved to be big enough to make room for one more thing.

Perhaps our thinking about wholes and parts needs enlargement too.

Experiment 3. The Philosophy of the Surf

Ocean waves are examples of an interesting and beautiful natural phenomenon. They are of interest to the physicist, the marine biologist, the geologist, the surfer, and the artist. Philosophers also have much to learn from ocean waves, though they do not always realize this. The following experiment shows that ocean waves can give us an important clue about the nature of the whole-part relationship.

A water wave results from the motion of matter on and near the surface of a body of water. Water waves happen when some force (usually the wind) pushes the molecules of the water around and starts them moving back and forth. The moving molecules push against other water molecules

near them, making the disturbance move across the water. Wherever the wave goes, molecules in the water move to and fro. If the wave is not too large and meets no obstacles, the pattern in which the molecules move is roughly circular.[7]

A water wave has energy. It can transfer energy to objects in the water, making them rock and bob, or to objects on the shore, causing erosion and other geological effects. The sound of crashing surf comes from the release of some of the waves' energy as sound waves.

Physical science has shown that molecules of water carry the energy of a water wave. Much of this energy is the *kinetic energy* associated with the motion of the molecules. A moving molecule, like a moving train, has energy that it can pass on to other things. The molecules also have *potential energy* because of the Earth's gravity (an interaction between the molecules and the Earth) and because of the molecules' interactions with other molecules through electrical forces. All the energy in the wave results from the motion and interactions of the wave's molecules.

[7] This idea is well-illustrated in Serway's text, *Physics for Scientists and Engineers with Modern Physics* (p. 347).

The fact that the wave has energy is explained by the fact that the molecules moving within the wave have energy. We do not need any mysterious energy source, besides the energies of moving molecules and of the forces that connect them, to explain the effects of the surf on the shore. The energy of the impact *is* the energy of moving molecules—that is all.

Physicists have established these facts about water waves. Yet they have not, to my knowledge, fully addressed a certain subtle point about the nature of the energy in the wave. This is the fact that the energy in the wave is the energy of water molecules in motion, but also is the energy of the wave as a whole.

We know that physical objects have energy. If you and I throw baseballs at the same moment, your baseball will have a certain amount of energy and so will mine. Each of the baseballs has its own quantity of energy, which (according to a well-known principle of physics) can be lost to other objects but never can be destroyed. Once we choose a scale for measuring energy, we can assign every material object a number that is a measure of the total energy of that object. A water wave has a certain amount of energy. Yet the energy of the wave, it seems, is not *just* the energy of the wave. It also is the energy of the molecules that move inside the wave.

There is a simple reason why the energy can belong to the molecules and still belong to the wave. The reason is that the molecules are in the wave. The argument about water waves is a lot like the fish experiment. I am pointing out that the wave is a real, physically significant thing, with its own energy, despite being "just" a product of molecules. But this is not all that I am driving at, for the wave is *not*

simply a composite of molecules. The molecules are not permanent parts of the wave; the wave can move from one area of the water to another, leaving behind one set of molecules and picking up another set. The wave is more like a *motion*, first of this water here, later of that water there. It is best to think of the wave as a process, or prolonged event, that happens to water molecules.

The point of this water-wave argument is not that the wave as a whole is a new item (though that is true). Rather, it is a point about the wave's *substance.*

Einstein's special theory of relativity, with its famous equation $E = mc^2$, implies that energy can be converted to matter and vice versa. This implies that energy is a form of substance; matter and energy are the two expressions of the substance of the physical world. Some philosophers have argued that a single substance which manifests as matter and energy, rather than matter or energy standing alone, is *the* true substance of the physical world.[8] One often hears the alternative suggestion that matter is simply a form of energy. But it would be arbitrary to regard matter as a substance while failing to regard energy as a substance. Indeed, the special theory of relativity implies that all energy has mass, just as matter does.

The water wave, then, carries some substance with it as it goes. When the wave moves from one part of the water surface to another, it might not carry along one single molecule of the water in which it originally traveled. Yet it carries along much of its original energy—its original substance. This energy is the kinetic and potential energy

[8] For example, Haeckel. See Reck, p. 123.

of water molecules. The wave is a real, concrete, substantial item with its own energy and mass—yet all of that energy and mass *also belongs to other things*, namely the water molecules currently inside the wave. The wave is a physical phenomenon that *shares the substance* of other physical entities. It gets its substance only through this sharing; it has no energy apart from the energy of water molecules.

In a certain sense, the existence of the wave is independent of the existence of the water molecules that revolve within it. Although the wave cannot exist without *some* water molecules, those molecules don't need to be the particular ones that now happen to be in the wave. Other water molecules, elsewhere in the sea, would do as well. It does not quite make sense to say the wave is "just" the motions of the molecules within it, since other motions of other molecules can make the wave exist just as well.

The lesson we learn from this is that something may have a real existence, and its own substance, even though all of its substance belongs to something else. The wave has no energy that is solely its own—yet as any surfer knows, it has loads of energy. The wave "lives on credit," as it were—smashing into the shore, or rocking the boat, with energy that also belongs to a bunch of tiny, invisible molecules.

Another lesson is that when many parts begin to act together in an organized way, this can create real wholes of a kind *fundamentally different* from the parts. In the triangle and fish experiments, we created more complex geometric figures from simpler geometric figures. In other words, simple what's-its gave rise to complicated what's-its of the same kind. But in the wave experiment, *objects* gave

rise to a *process*. The molecules gave rise to the wave, yet a wave is an item of a sort fundamentally different from water molecules. The wave is not really a "thing" at all, but a prolonged event. In this case, we could say the what's-its didn't just result in fancier what's-its; they resulted in thingamabobs instead.

This type of creation, which creates fundamentally new *kinds* of items, happens all the time. Most processes and events in the physical universe are "non-objects" that result from the activity of objects.

Common sense about wholes and parts says that when we put together bits of stuff, the most we will get is a bigger piece of stuff of the same kind. According to this view, the worst we can possibly get is a bigger piece with shockingly fancy properties—like a computer chip, made of silicon atoms but having the ability to perform calculations. But the wave example suggests there are exceptions to this rule. If we put together enough water molecules in the right arrangement, we may get a wave—but a wave is a process instead of a proper object. Like the molecules, it exists, but it exists in a way different from the way that molecules exist. A water wave differs from a molecule in other respects besides its physical properties—though the differences in some of these properties (especially size) are obvious. Apart from these differences in physical properties, the wave has a strikingly different kind of being or existence. Using some long-standing philosophical jargon, we can say that a water wave and water molecules belong to two different *ontological categories*. (Later I will have more to say about ontological categories.)

The thought experiments in this chapter do not prove anything rigorously. They do not pretend to be formal philosophical arguments. These experiments only point out some features of the part-whole relation that people (including scientists and philosophers) don't often think deeply about. These features run counter to some of our usual intuitions about wholes and parts. The three thought experiments presented here make these features seem more intuitively reasonable, and challenge the commonsense view that the whole is in some sense "just its parts." In the next few chapters I will challenge this view more systematically, and will begin to lay the foundations for a new understanding of the wholes and parts that we find in the natural world.

Chapter 2.
Is Reality Holistic?

The urge to think about the connection between whole and part is nothing new. Scientists usually use the ideas of whole and part without analyzing them, but philosophers have tried to understand these ideas in a more general and penetrating way. The best-known philosophical problem about whole-part relations is the famous question "Is the whole more than the sum of its parts?"

People have written a great deal about this question over the centuries, and have proposed several answers. Most, if not all, of these answers belong to one of two main groups. Some thinkers have said that an object with parts is, in some sense, nothing more than all of its parts. This line of thought is called *reductionism.* (There also are other ideas called "reductionism," but I won't discuss all of them here.) Other thinkers, equally qualified, have argued that an object with parts is something more than just its parts—that bringing the parts together results in a whole that is something more than just the parts. This line of thought is known as *holism.*

Note added later: The type of reductionism that I critique here is the ontological view that an object is "nothing but" its parts. I am not critiquing intertheoretic reduction or other notions of reduction. Similarly, the type of holism I am discussing here is only one variety of holism.

Reductionism: Method or Ideology?

In science, the reductionist approach has long been in favor. Scientists try to explain the behavior of a complex object in terms of the behavior of its parts. The supreme example of scientific reductionism is the biologists' explanation of life in terms of chemical and physical events. Modern biologists believe the chief features of living organisms result from the behavior of large numbers of physical particles (such as atoms and electrons) organized in a mind-bogglingly complex way. Today, the physicochemical view of life is the one that scientists accept.

If scientists ever explain the human mind in terms of atoms, molecules, and electrons, that would be a reductionistic accomplishment even greater than the physical explanation of life. We do not yet have a full physical explanation of mind, though scientists have made progress in that direction. Some of the simpler features of the mind (and even some complex ones) can be simulated by computers. This suggests that those simpler features may have physical explanations that are not too hard to find.

Many scientists think that a physical explanation of mind is possible. This confidence comes partly from evidence that some features of the mind are physical. But the chief motive for this belief might be other than scientific. Some scientists and philosophers seem to believe that if we cannot explain the mind in terms of the brain, then we will have to leave the mind without a

rational explanation.[9] The idea that there is something unexplainable is taboo to many thinkers, who believe this idea amounts to accepting superstition. Those who think this way trust that a physical explanation of mind will be discovered, because the absence of an explanation threatens the scientific worldview. Thus, although the proponents of reductionism sometimes hold themselves out as advocates of reason, reductionist belief often is a matter of emotion as much as of science. (Of course, this is true of some holistic belief, too.)

Some thinkers who believe in a brain-based explanation of mind still feel that the mind is more than a machine. To develop this view, such thinkers often turn to a holistic interpretation of mind. According to the holistic view, the mind is a product of the activity of the brain—and yet there is something more to the mind than the simple, mechanical firings of neurons. When the neurons come together into the complex pattern known as the human brain, the whole brain develops properties for which the properties of individual neurons cannot account. This is the gist of the holistic view.

Holism is not only an idea about the human mind. One also can take a holistic view of other happenings and objects in the cosmos. Living organisms form the chief target of holistic theorizing. In living things there are many properties, processes and functions that do not have any counterpart in the tiny material parts that make up living things. Living organisms digest other objects; atoms

[9] Daniel C. Dennett takes a position close to this in *Consciousness Explained* (p. 37).

cannot digest anything, and there is nothing that an atom can do that is much like digestion. Even the simplest "automatic" muscular motions of animals, or the water-pressure-driven movements of plants, are far too complex to be carried out by an atom or a quark.

There are different brands of holism, and some of them are not exactly like what I have described here. A serious holist might regard my description as a mere caricature of holism. Nevertheless, my description captures the essential point of holism: that a complex system has features that are not fully explained by the properties of that system's parts.

Reductionists have offered their own caricatures of the holistic school of thought. One of these caricatures is in Marvin Minsky's book, *The Society of Mind*. Minsky presents what he calls a "parody of a conversation between a Holist and an ordinary Citizen."[10] I will summarize this conversation here (the italics in the quotes have been changed). The Holist sets out to show that "no box can hold a mouse." First the Holist claims that a box really doesn't have a property of "'mousetightness' or 'containment'" at all. To prove this point, the Holist points out that "no single board" in the box "contains any containment," and concludes on this basis that "the box can have no mousetightness at all." Instead, the holist contends, "a good box can 'simulate' [mousetightness] so well that the mouse is fooled and can't figure out how to

[10] Minsky, *The Society of Mind*, p. 28. I should mention that Minsky uses the term "Reductionist" in sense slightly different from mine (see *The Society of Mind*, p. 26).

escape."[11]

It is good that Minsky admits that this conversation is a parody, for the ridiculous doctrine put forth by Minsky's "Holist" has little to do with real holism. A real-life holist would not claim that the mousetightness of the box is only simulated. Such a holist would not disagree with a reductionist over the fact that the box really is mousetight. After all, to say that the box is mousetight is just to say that it is able to keep mice in—and both sides agree that it does that. The disagreement between the holist and the reductionist lies in their accounts of *what this mousetightness is.* A reductionist might say that although the box really is mousetight, there is nothing to this mousetightness besides certain features of the individual boards. In other words, the box doesn't need to have a separate property of mousetightness to keep the mouse in. The boards, when properly arranged, can do it by themselves. (This seems to me to be Minsky's view—that there is nothing to the mousetightness of the box besides the separate abilities of the individual boards to block the mouse's movement.[12]) But a holist might claim that the mousetightness is not quite the same as the impenetrability of the individual boards. Instead, it is a new property *which the box itself has*, and which comes into being when one nails together the boards into a box. Mousetightness is not the same as any property of the boards, or as any set of properties of boards. A holist might concede that the

[11] Minsky, *The Society of Mind*, p. 28 (italics changed in quotes).

[12] See Minsky, *The Society of Mind*, p. 28.

properties and arrangement of the boards are what cause the box to be mousetight. However, the holist would say that the mousetightness itself is a new property which the boards do not have—a property of the box, not of the boards. Since the mousetightness belongs to the box and not to any board, it is not a property of boards at all. No part of the box has mousetightness or anything close to it—yet the whole manages to have this property nonetheless, and therefore is more than the sum of its parts!

The holist in Minsky's story is a dupe. The "parody of a conversation" which Minsky discusses is indeed a parody; it makes holism appear to be silly at best and intellectually dishonest at worst. Actually, it is Minsky's parody that lacks credibility. Obviously, no serious holist would claim that a box that can confine a mouse isn't really mousetight, or would make the ridiculous claim that the reason the mouse can't escape the box is because the mouse is fooled. Minsky's argument substitutes ridicule for reasoned debate. In reality, many first-class thinkers, both ancient and modern, have embraced holism of one kind or another.

The founder of holism as a systematic philosophical outlook was the nineteenth- to twentieth-century philosopher J.C. Smuts. In his book *Holism and Evolution*, Smuts set forth a well-reasoned view of nature as a system of wholes, each of which may have certain properties quite different from those of its parts.[13] My view of whole and part is not the same as Smuts's view. Later I will mention some of the key similarities and differences between these

[13] Smuts, *Holism and Evolution*; see particularly chapters 5 and 6.

two views. (Mostly I will do this in footnotes.)

The reductionist approach to the problems of life and mind is indispensable for scientific work. If we refuse to admit the possibility that the properties of the parts account for the properties of whole, then we have a much harder time understanding the properties of the whole. Worse yet, we will lose the possibility of learning about such an explanation if one happens to exist. The philosopher Daniel C. Dennett has pointed out a problem of this sort with mind-body *dualism*—the commonly held view that there is a nonphysical mind apart from the brain. In a book describing his reductionistic theory of consciousness, Dennett once wrote that "*accepting dualism is giving up*"[14] (italics in original). Part of what this means, I think, is that if we assume that the mind does not have a physical explanation, then we are stopping inquiry before we know whether such an explanation is possible. Unlike many current philosophers, I do not believe that a dualistic explanation of mind has to be antiscientific. (That does not imply that I am a dualist; more on that topic later.) But Dennett's remark can just as well be applied to the refusal to try to explain the properties of a whole in terms of those of the parts. If such an explanation is possible, then the seeker of truth wants to know it, and the only way to find out whether there is such an explanation is to try to make one. Assuming in advance that a reductionistic explanation is impossible cuts us off from the possibility of learning something potentially interesting. Therefore we should try to find such explanations, whether or not we have faith in

[14] Dennett, *Consciousness Explained*, p. 37.

the scientific worldview.

Reductionism is useful as a methodological assumption for scientific inquiry. Its usefulness, however, does not settle the question of the truth of the reductionist view of the whole-part relation. It is logically possible that the whole is not reducible to the parts, and that there is something to a whole object that is not encompassed in any of that object's parts. The fact that scientists must pretend to be reductionists while working does not remove this possibility. The well-established usefulness of reductionistic methods in science does not prove the reductionist viewpoint in philosophy. Nor can philosophical holism be used to attack the use of reductionist methods in science. It is important not to confuse the reductionist *method* that scientists follow, with the reductionist *worldview* that certain philosophers and scientists embrace. It is possible to follow the method without buying into the ideology.

Emergent Properties

One criticism often leveled against holism is that it is vague. Some reductionists have pointed out that holists say things like "the whole is more than the sum of the parts" without saying exactly *how* the whole is more than the sum of the parts. Holists claim that there is something more to the whole than the individual parts and their properties. Yet often they decline to say what this "something more" is. This coyness comes from the fact that the holists don't always know what the "something more" is. They have found clear signs that there is something to the whole besides the parts, but they do not know exactly what those

signs point to—what the difference is between a whole and a "mere" sum of parts.

The holists' inability to say exactly how the whole differs from the parts has led some reductionists to claim that holism is unscientific or that it embraces mystification.[15] There are two glaring mistakes in this reductionistic claim. First, there is nothing unscientific about claiming to know something is real without knowing exactly what that "something" is. Most natural phenomena, including radioactivity, meteorites, and life itself, were known, and even studied scientifically, before their true nature was understood. The holist is in much the same position as an early scientist studying meteorites. Such a scientist might have said "The evidence points to the existence of stones that fall out of the sky—but we don't know where those stones come from." Similarly, the holist notes that there is a difference between whole and parts, but does not yet know what that difference is. The reductionist, on the other hand, is more like those early scientists who believed that reports of meteorites simply *must* be wrong.[16] Of course, this loose analogy doesn't prove reductionism wrong. But it should teach the reductionist a lesson in caution.

The other reason that holism is not mystifying is that we *do* know, at least in part, how an object can differ from

[15] Minsky, in my opinion, comes close to this view in *The Society of Mind*, where he suggests that the word "holistic" acts "to anesthetize a sense of ignorance" (p. 27).

[16] Even some very smart scientists once held this view. See Pearl, *Rocks and Minerals*, p. 165.

the sum of the parts. Philosophers (holistic or not) who have thought about these issues have come up with one definite answer to the question "What is the difference?" That answer is *emergent properties.*

Philosophers have thought of emergent properties in different ways,[17] but at bottom, the idea of an emergent property is simple. An emergent property is simply a property that an object has, but that the parts of that object do not have if the object is divided into sufficiently small parts. The mousetightness of Minsky's box (which I discussed earlier) is one example of an emergent property. No piece of wood used to make the box can confine a mouse—yet the box, taken as a whole, can. The shape of the triangle I discussed in Chapter 1 is another example. None of the parts used to make the triangle is triangular—yet the triangle, as a whole, is triangular. The triangle also has the emergent property of *closure*—it is a closed figure; one cannot get out of the figure without crossing a line or leaving the plane of the page. No part of the triangle has this property. Removing any part of the triangle will make the triangle lose the property of closure. Thus, the property of closure depends on the "cooperation" of *all* the parts of the triangle.

Emergent properties are called "emergent" because they emerge when things are put together into larger things.

[17] The way that I define emergent properties does not necessarily agree completely with the way that some other authors have defined these. Also, I should mention that I do not necessarily agree with the philosophical position known as "emergentism"—at least not in all of its forms. In this book I am going to ignore some of the larger issues surrounding emergence and reduction, because these issues are not crucial to my point.

Note added later: The "emergence" I am discussing here is a rather weak version of emergence. This should not be confused with stronger versions found in the literature, which may make emergent properties practically unpredictable or unexplainable.

They are not present in the smaller things—but when the smaller things are assembled into a more complex whole, the emergent properties pop up.

Emergent properties are all around us. You notice this when you begin to look for them. The *page count* of a book is a legitimate property of the book. Yet the individual fibers of paper and splotches of ink that make up the book do not have page counts. The *color* of any colored object is a real physical property. Yet none of the atoms that make up the object is, by itself, colored. An atom by itself is invisible and colorless. The *shapes* of objects are emergent properties; they are results of the arrangements of the parts of objects. The atoms that make up an object have shapes different from that of the object.

Most of the properties that we deal with every day of our lives are emergent properties. We call the world that we perceive with our senses the "physical world." Would it not be almost as correct to call it the *world of emergent properties?*

Some of the most interesting emergent properties occur in the science of chemistry. I am thinking especially of the properties of *solidity* and *liquidity*. Everyone knows intuitively what solids and liquids are. Solids are substances that have definite shapes and do not flow visibly, while liquids assume the shapes of their containers and seem "wet." Physical chemists have more precise definitions of these notions. Scientists know that solids are substances in which the atoms or molecules making up the substance fall into repeating, lattice-like patterns. (Some familiar "solids," such as ordinary window glass, are not true solids but "amorphous solids," which act like solids in many respects.) Liquids are materials in which the atoms

or molecules move freely around one another, yet stick together enough that they do not quickly go flying off into space. Liquids evaporate when this stickiness of the molecules is overcome by something—usually by the energy of heat.

Liquidity, philosophers have noted, is an emergent property.[18] When one looks at the molecules or atoms that make up a liquid, one finds nothing at all that is liquid. Individual atoms do not flow: they can fly through space or be still, but they do not pour or slosh. A sloshing atom is as silly an idea as a flowing baseball. Yet an unimaginable number of atoms, clustered together and stuck to their neighbors by electrical forces, forms a mass of stuff that can flow. We call that a liquid.

It seems clear that emergent properties, as I have defined them, exist. They are as real as any other properties of objects. (Philosophers have long debated the question of whether properties really exist at all; some have argued that properties are mere fictions, and that only plain old objects are real.[19] So perhaps I should say that emergent properties exist insofar as any properties exist.) A box that is mouseproof (to exploit Minsky's example again) really does have the property of being mouseproof, even though all of its small parts lack that property. A pond really is liquid; hence it really does have the property

[18] Searle notes this in *The Rediscovery of the Mind*, pp. 111-112.

[19] I am referring, of course, to the nominalists. For discussions of this and other positions on the problem of universals, see Loux (ed.), *Universals and Particulars.*

of liquidity. To deny that emergent properties are as real as other properties is to deny that a pond is liquid and that a solidly closed box is mouseproof. Do you really want to claim that water, at room temperature and pressure, isn't liquid?

Some people are afflicted with the mistaken view that philosophers don't believe in the physical world. Those who hold this view think that philosophers have somehow denied the existence of the perceptible world around us. Some people find this alleged denial of reality to be amusing. This allegation against philosophers has little basis in fact, but there seems to be a fairly common popular belief that philosophers think that way. Now, would a philosopher who claimed that water isn't wet be in a less ridiculous position than a philosopher who claimed that my chair doesn't exist? If we do not want to fall into skepticism about the existence and basic features of the world around us, then we should not try to deny that water is wet! The case for the reality of emergent properties like wetness is as strong as the case for the existence of tables and chairs. If any properties exist in the world, then emergent properties exist in the world.[20]

Despite all this, the mere existence of emergent properties does not completely settle the holism vs. reductionism debate. To settle that, we still must answer

[20] Often I will speak of a property as *existing* if it is instantiated or exemplified. A Platonic realist might dislike this usage on the grounds that existence and instantiation of an abstract object aren't the same thing. A nominalist might dislike it on the grounds that no properties really exist. I am dodging these questions here and am using "exist" in the more intuitive sense I have just described.

the question I posed at the beginning of this chapter. Is the whole (with all its emergent properties) just the "sum of its parts," or is there something more to the whole than there is to the parts taken together?

If some emergent property of the whole cannot be explained in terms of the properties and relationships of the parts, then there would seem to be grounds for believing in a form of holism. In this chapter I will not ask whether all properties can be explained in this way. Instead, I want to draw attention to a point of logic about emergent properties. The point is this: if we look at sufficiently small parts of the whole, then an emergent property of the whole is not identical with any property found in those parts. Maybe an emergent property can be explained by (or reduced to) simpler properties of the parts. But even if it can, we are stuck with the fact that the emergent property *is not* any of those simpler properties. The mouseproofness of Minsky's box is not the hardness of the box's north wall, or the squareness of the box's ceiling. If it were any of these properties of the parts, then at least one of the smaller parts would itself have the property of mouseproofness—and we know it does not. We know that the mouseproofness of the box is real, and that it is not the same property as any property of a board in the box. To know this, we do not need to know whether the mouseproofness can be explained in terms of the properties of the boards. Even if the mouseproofness can be "explained away," it still is undeniably real. (If you doubt this, ask the mouse!)

The fact that the emergent property is real, and is not identical to the properties of the parts, has an interesting consequence. This is that if we count the properties of the

box and of its parts, we will find at least one more property after we assemble the box than when we started. Before the box is built, each part has a certain set of properties; by uniting all these sets into one big set, we find that the separate parts, taken collectively, have a certain set of properties. After the box is built, another property springs up: that of mouseproofness. Of course, many other properties might come into play too, and some properties of the parts, like their independent movability, are lost as well—so the total number of properties (if one actually counted them!) might go up or down or remain the same. But the important fact is that there is at least one new property, a property that did not exist before the box was built. This property came into being when the box did. By building the box, *we created this property*.[21] The box has a real property that, for all we know, was not present in the world at all before the box was built.

We now see that emergent properties are in much the same position as the triangle and fish depicted in Chapter 1. An emergent property of an object is something that exists *in addition to* all the properties of the object's small parts. To assemble an object having such a property is to *create* that property—to bring it into being, or to bring an example of it into being. To have a full accounting of all the properties involved with the box, we must list the emergent

[21] Some philosophers (a subset of the Platonic realists) might want to maintain that properties really exist always, and are not literally created. If that is so, then we should say that the property was not *exemplified* before the box was built, and began to be exemplified when the box was built.

property as well as the properties of the parts. It may well be that the emergent property can be explained in terms of the properties of the box's parts. But even if it can, this does not change the fact that the emergent property is something real, and something quite apart from any of the properties of sufficiently small parts.

The position I have reached here is similar, though not identical, to certain ideas of the philosopher John R. Searle. Writing about mind and consciousness, Searle has argued that "consciousness is a causally emergent property of systems"[22], and that mental states are caused by physical goings-on in the brain.[23] Taken together, these two claims of Searle's imply that a property (consciousness) of a whole (the brain) can be an *effect* of the presence of certain properties in the parts, instead of *being identical* to properties of the parts.

The implications of Searle's line of thought show up in his discussion of the philosophical idea of *supervenience*. Supervenience is an idea that often surfaces in discussions of complex wholes such as brains. The word itself is somewhat vague; Searle distinguishes more than one meaning for it.[24] On one of these senses, to say that a phenomenon (like thought) supervenes on some other phenomenon (like brain activity) is to say, more or less, that there is nothing to the first phenomenon besides that

[22] Searle, *The Rediscovery of the Mind*, p. 112.

[23] Searle, *The Rediscovery of the Mind*, p. 125.

[24] Searle, *The Rediscovery of the Mind*, p. 125.

other phenomenon. On the other sense (which Searle calls "causal"), the supervenient phenomenon is merely *caused* by the other phenomenon, and is completely controlled by it. Searle points out that "[t]he solidity of the piston is causally supervenient on its molecular structure."[25] This implies that the solidity of the piston is an effect of the state of the molecules in the piston. The solidity isn't a property of any of the molecules themselves, but is something that comes into being when the molecules come to be arranged in a certain way. Searle's views on supervenience come close to, and perhaps imply, my thesis that an emergent property has an existence separate from that of the properties of the parts of the object that has it.

A scientific explanation of the mouseproofness of the box, or of the liquidity of water, may well explain those properties in terms of simpler characteristics. Once the scientific explanation has done this, it is done with the mouseproofness or liquidity; it has nothing more to say about what these properties are. But if we want to understand what the box or the water *really is*, then we *must* count the mouseproofness or the liquidity as real properties, along with the simpler properties used in the scientific explanation. We cannot excuse ourselves from this duty by claiming that mouseproofness or liquidity is not a separate property or has a scientific explanation. Complex properties are real. They are just as real as simple properties. We may know that the emergent properties depend for their existence on other properties, but this does not imply that they are "just" those other properties, or that

[25] Searle, *The Rediscovery of the Mind,* p. 126.

they lack an existence of their own. If you don't believe it, just count!

Emergent properties exist whether or not they are scientifically "reducible" to other properties. Using a bit of philosophical jargon, we can say that emergent properties belong to the *ontology* of the physical world. An ontology is a theory about what exists, or an inventory of the kinds of items that exist. If we want to describe the ontology of a body of water, we must include in our account both the *properties of the molecules* and the *liquidity of the whole*. If we leave any properties at all in our ontology, we must leave in the liquidity too. To do otherwise would be arbitrary and unjustified. The only possible reason for leaving liquidity out would be to support a philosophical prejudice: that things explainable in terms of other things are not quite real. But does anyone—even an intelligent reductionist—want to claim that water is not really wet?

The Triangle and Fish Revisited

The real existence of emergent properties also has another interesting result. It leaves no doubt that a composite object is more than just its parts. If liquidity exists in the physical world, then there must be something to have the liquidity. The property of liquidity that we find in the world is not some free-floating property, exemplified by nothing. Liquidity is a property of *objects*—for example, of certain masses or blobs of water molecules. Such a blob can have a property of its own; therefore, the blob is not "just" the water molecules—that is, it is not all the molecules taken together. Rather, the blob must be a distinct entity—presumably, an entity that comes into

existence when water molecules are arranged in a suitable way. Otherwise the blob could not have a real property.

Take a zillion isolated water molecules, and you just have a lot of water molecules. Put them together in roughly the same place, and let them stick together as they naturally do. Then you have a sample of liquid. Of course, you still have the zillion water molecules. *But now you have a zillion and one things.* The extra thing—the liquid sample—is different from its parts because it has a property—liquidity—that none of the parts can have. Take the beaker in which the water molecules sit, and try to pour from it. If something pours or glugs (instead of just flying through space as a collection of independent molecules, like so many minuscule billiard balls), then it is safe to infer that there is something in the beaker besides the molecules. After all, a molecule can't glug. Of course, what's in the beaker is made of molecules. But that doesn't mean that what's in the beaker *is* just molecules. Without the molecules, the stuff that pours would not be able to pour, would not have any of its other physical characteristics, and would not exist at all. Nevertheless, that sample of stuff is not just molecules.

The water sample, like the wave in Chapter 1, shares all of its substance with the molecules that exist within it. It can have no properties except those to which a zillion molecules, acting together, can give rise. Normally, we would think of the water sample as somehow *being* its molecules; we might think that in some ultimate sense, there is nothing there *but* molecules. As I have argued here and in Chapter 1, it is more correct to say that the water sample is a thing *in addition to* its molecules, but which owes every bit of its substance to the molecules. If we

condense this water sample out of isolated water molecules, *we literally create a new object.* We also create properties that were not there before—properties that we must tally up if we want to count all the real properties of the water sample.

This view of the water sample may seem contrary to the scientific approach to the study of liquids. A little thought will show that it is not. The scientific explanation of liquidity and the molecular model of water will remain verifiable and correct, whether the sample is a new object or is identical to the molecules that make it up. Neither of these two views of the sample can contradict any scientific prediction about the behavior of the water or of its molecules. The question of whether the water sample is an extra object is not a scientific question. We cannot settle this question by doing experiments or making measurements. (This is true of all genuinely philosophical questions.) To settle these questions, one also must worry about the logical consistency and coherence of the different possible answers. As we have seen, some rather simple observations about the logic of properties, and about ways of counting objects, suggest that one view is logically neater than the other.

Elimination vs. Reduction: A Technical Note

Philosophers of science sometimes distinguish between *reduction* and *elimination* in scientific thought.[26] To

[26] The distinction between eliminative and reductive forms of materialism is outlined in Cornman and Lehrer, *Philosophical Problems and Arguments: An Introduction*, p. 282.

reduce an object or phenomenon is to show that it is "nothing but" something else. A classic example of reduction, often mentioned in discussions of the philosophy of science, is the claim that water is nothing but the chemical compound H_2O. To *eliminate* something is to show that we can dispense with it altogether in our thinking, and can get by without claiming that it exists. An example of elimination is the argument that we need not believe a drop of water exists, because if we just assume that the water molecules are there we can explain all the measured properties usually ascribed to the drop.

If we wish to use these terms, we can restate our most recent conclusions as follows. First, it is impossible to eliminate any composite object, if "eliminate" is defined as above. Even if the existence and properties of the object's parts completely explain the existence and properties of the object, it still is the case that the object is there. Also, the object is not the same as any of its parts, or as several of its parts together. Second, a reduction of a composite object may be correct, but only if that reduction does not involve elimination of the object. Instead of saying that reduction of the whole to its parts is possible, we must first be sure we know what we mean by "reduction." If "reduction" means explaining the properties of the whole in terms of those of the parts, then we have not ruled out a reduction. If "reduction" means showing that the whole is nothing but a composite of its parts (that a water drop is nothing but a composite of molecules), then we have not ruled that out. The "composite" of the parts is, after all, just another name

for the whole. But if "reduction" means showing that there really is nothing there except the parts, or that the whole somehow is the parts and nothing more, then such "reduction" is out of the question. It is elimination in disguise, and it is logically untenable.

Of course, it may be convenient to ignore the whole and consider only the parts—for example, in a scientific calculation where we treat a macroscopic thing as a set of atoms. But that way of thinking is a practical convenience, and says nothing about the reality of the whole. If the whole is real but its parts control its properties, then a calculation that substitutes the parts for the whole may yield correct results. But even if we can ignore the whole in our calculations, that whole still is there—and is not the same as its parts.

Some scientists, and other scientifically oriented people, seem unclear about the difference between reduction and elimination. They seem to think that just because physical objects are made of atoms or particles, there really is nothing in the physical world besides atoms or particles. Many scientists would deny that a stone (for example) is unreal—yet when they discuss the nature of physical reality, they state or broadly imply that a stone "really is only atoms." Occasionally one reads statements like this in the literature of science and philosophy.[27] More often, one hears them in conversation—with scientists,

[27] The *locus classicus* is perhaps Democritus' well-known statement that "atoms and Void *(alone)* exist in reality" (p. 93 in Freeman (ed.), *Ancilla to The Pre-Socratic Philosophers*; italics unchanged by me).

academic philosophers, or others. The classic example is the old chestnut about a person really being a few dollars' worth of chemicals. But all such statements rest on a mistake. The universe as portrayed by science does not contain only elementary particles of matter. It contains those particles, *plus the composite objects built up from those particles*. To speak as though the particles are all there is—as though once you've counted the particles, you've counted everything there is—is a grave logical error. The composite objects are not redundant. To get a full inventory of things in the physical universe, you must list not only the electrons, quarks, and so forth, but also the larger things built up from them. If you count all the particles in a stone and then count the stone containing those particles, you haven't counted up the same thing twice.[28] The stone may be made of the particles, but the stone is not the particles. The stone is an additional entity—one more thing, distinct from the particles that make it up.

The view that the physical universe really is only particles, or that human bodies really are only atoms, sometimes gets stated explicitly. Far more often, this view lurks behind other viewpoints as an unstated assumption or an underlying attitude. A psychobiologist might laugh at the view that people don't exist—yet in practice, he might think of the human organism as though its only "real" properties were molecular ones. A physicist might feel that

[28] The concept of "double counting" also is used by Lewis (*Parts of Classes*, p. 81). Lewis, however, draws a conclusion opposite to mine.

the discovery of a "theory of everything"—resulting in a complete physical description of elementary particles—would reveal to us what the physical universe "really is." A materialist philosopher might deny that chairs and tables are illusory—and yet might privately picture the material world as a set of interacting elementary particles. Each of these attitudes rests on an unstated view that the ultimate parts of objects are somehow more important or fundamental to our picture of the world than are the objects themselves.

Epilogue

Nothing I have said in this chapter settles the entire holism-reductionism controversy in its usual form. At every step in my argument, I have allowed for the possibility that the properties of wholes can be explained in terms of the parts. Also, I have allowed for the opposite possibility. I have asserted that when water molecules come together, new properties can appear that do not belong to the individual molecules. However, I have neither asserted nor denied the strict holistic claim that some of these properties *cannot* be explained by the behavior of molecules. I have merely pointed out that when the molecules come together, they may form a new object with new properties. Also I have shown that the new properties that this object has are as real as any other properties in the world. If believing that the whole object exists and has properties is a form of holism, then of course I am arguing for holism. But this label would be unfair, since a reductionist does not have to give up all forms of reductionism to believe that a glass of water exists. (I

suspect that most reductionists do believe this, especially when thirsty.)

Some of my observations about whole objects may seem like small technical points, or like restatements of commonplace truths. One might want to ask whether such modest results have any real relevance to the holism-reductionism controversy. In the coming chapters, I will show that people (including scientists) often forget about these "small" points when thinking about wholes and parts. If we revise our usual thinking about parts and wholes while always keeping in mind that the whole is real, we will arrive at a view of the world and of human existence surprisingly different from our usual views. This new view will make several long-standing philosophical problems much easier to think about. Indeed, we will find that some of the knottiest problems in philosophy were partly illusions, created by our failure to grasp the implications of the separate existence of the whole.

Chapter 3.
Walls, Bricks and Logic

Earlier I said that scientists often ignore the larger questions about the connection between whole and part. This indifference is strange when we consider how much of science is about this connection. As I pointed out earlier, the physicist's search for the final building blocks of matter is just an attempt to answer a question about wholes and parts. But the scientists' unconcern with the general problem of wholes and parts becomes even more ironic when we learn that there already is a precise, "scientific" method for the study of these ideas. This method is called *mereology*—a word of Greek origin meaning the science of parts.[29]

Mereology is both a mathematical and a philosophical subject. Mereology is not a science in the same way that physics and biology are sciences; it does not depend on experiments and scientific observations to prove its conclusions. Like any part of mathematical logic, it is a

[29] Mereology is discussed in a number of sources, including Woodger, *The Technique of Theory Construction*, and Lewis, *Parts of Classes*.

formal science—one that uses the methods of deductive reasoning to analyze old ideas and make up new ones. It is best to regard mereology as a branch of philosophy rather than of science, though it belongs to the more "scientific," or rigorous, end of philosophy.

Mereology as a mathematical discipline was founded by the Polish logician Lésniewski in the first half of the twentieth century.[30] (Mereological thought existed before that time,[31] but earlier mereology did not yet take the rigorous form that Lésniewski gave it.) Other philosophers have extended mereology further and have applied it to various scientific fields. David Lewis, perhaps better known for his innovative ideas about "possible worlds," has shown that one can use mereology to better understand the foundations of the mathematics of the infinite.[32] Joseph H. Woodger used mereology to set up a precise theory of the main ideas of biology, including cell division and even the origin of life.[33]

---

[30] Lésniewski's work is discussed, and references are given, in Lewis, *Parts of Classes*, p. 72. Lésniewski's original papers on mereology are in Polish.

[31] Medieval mereological thought is discussed in Henry, *Medieval Mereology*.

[32] Lewis, *Parts of Classes*. Lewis's ideas about possible worlds are discussed in his books *Counterfactuals* and *On the Plurality of Worlds*.

[33] For an introduction to Woodger's ideas, see his book *The Technique of Theory Construction*. See p. 64 for a mereological version of the idea of abiogenesis (the origin of life from nonliving matter).

In this chapter I am not going to go into the mathematical depths of the subject of mereology. Instead, I am going to examine a few of the guiding ideas that have played important roles in shaping that field. My aim is to review briefly (and non-mathematically!) this study of whole and part, and then to point out some unsolved problems about mereology that may point the way to a new understanding of the entire part-whole business.

The Crucial Relation

The central idea of mereology is that of the relation between a part and a whole. When we say something like "This brick is part of the wall," we refer to two things—a brick and a wall—but not only to two things. We also refer to a *relation*—the abstract object or concept to which the phrase "is part of" refers.[34] The first trick of mereology is to treat this relation in the same way that mathematicians and logicians treat all other relations. Arithmetic deals with relations between numbers, especially the relations represented by the phrases "is greater than," "is less than," and "is equal to" (or "equals"). Mereology deals with a relation between *objects*—the relation referred to by the phrase "is part of."

Like any mathematical discipline, mereology uses symbols for its basic notions. I will not use symbols here,

[34] Some philosophers of language will question my use of "refer" in this sentence and elsewhere. Their point, though worthy of consideration, does not affect the subsequent argument.

since I am not going to set up any complex mereological proofs. I will be able to do what I want to do using words alone, plus a few stray letters. Mereology also uses axioms, or basic principles, as starting points for proving more complex results. These axioms do not legislate in advance the answers to any questions; one always can revise the axioms if they do not hold true in the real world. Two principles of mereology that are useful as axioms are the following:

Principle of irreflexivity. Let x and y be things. If x is a part of y and x is not the same thing as y, then y is not a part of x.

Principle of transitivity. Let x, y, and z be things. If x is a part of y, and y is a part of z, then x is a part of z.

These axioms are just ways of restating truths that seem obvious in everyday life. If a brick is part of a wall, then the entire wall cannot be part of the brick. If a page is part of a chapter, and a chapter is part of a book, then the page is part of a book. One can, of course, ask whether there could be exceptions to rules like these. But I will not do this here, since my goal is to do something else.

Mereology takes the relation expressed by "is part of" to be another relation, on the same logical level as other relations like those expressed by "is greater than," "is longer than," and "existed earlier than." The relation of *being greater than* can hold between two numbers. The relation of *being longer than* can hold between two physical objects. The relation of *being a part of* also can hold between two physical objects, and perhaps (as Lewis

has argued[35]) between two mathematical objects as well. When we do mereology, there are two things we have to think about: the objects that make up the world, and the whole-part relation that links some of them together.

Mereology is about the realm of things—a realm that contains at least physical objects, and (for all we know) perhaps other items as well. Mereology begins with a domain of things or entities, and describes a relation—that of *being a part of*—which holds between some pairs of things and not others. I take it for granted that this general view of the world is correct, at least for most practical purposes. There really are a lot of things in this world (unless one wants to be an utter skeptic and claim that things are illusory). Philosophical accounts of what things really are cannot change this practical fact. Further, it is true that some of those things are related to each other in the way that we call the relation between whole and part. As long as one accepts the existence of a physical world and the fact that some things have parts, one should have no trouble with the basic way that mereology describes the world. One might doubt the particular assumptions that mereologists sometimes use, but there would be no reason to doubt that the world contains things, and that things stand to one another in the relation of part to whole.

A Shift of Viewpoint

My aim here is not to go into mereological theory in detail. Instead, I want to use mereology as a jumping-off

[35] Lewis, *Parts of Classes*; see especially pp. 3-4.

point for an argument about the nature of wholes and parts.

Consider the statements I made two paragraphs ago about the world as portrayed by mereology. This view of the world—which is the view almost everyone uses without thinking about it—is logically sound. However, this view does not fully agree with another view that some scientists seem to use. The conventional scientific picture of the world regards the world as a world of parts. According to that view, the smallest parts of the world explain everything in the world; once we have a description of the ultimate particles, we have, in principle, an explanation of everything in the universe. All else is almost incidental; since a galaxy is nothing but elementary particles, we do not need a theory of what galaxies "really are," apart from our views on elementary particles and their forces. But common sense and mereology both portray the world as a world of *objects*, not just of invisible particles. The ultimate particles may be among the objects, but are not the only objects in the world. The world of objects is not the world of ultimate particles, for although everything is made of ultimate particles, the larger objects in the world are neither more nor less real than the particles.

This last view is the one that actually underlies science—if we consider science as scientists really do it, instead of confusing science with some philosophical attitude that is supposed to be "scientific." Scientists working on problems of complex physical systems (like crystals or liquid drops) treat those systems as real objects. Such objects contain their own mysteries, perhaps as deep and difficult in their own way as the mysteries of subatomic particles. Some scientists may *say* that the world is nothing but quarks and the like, and that a theory of quarks and

similar particles would explain everything. But in their work, scientists *act* as if larger objects were exactly as real as quarks—that is, as if the physical world were a world full of *objects,* and not just a set of tiny pieces.

Apart from mereology's correct picture of objects, there is something else about mereology that is equally right. This is mereology's treatment of the whole-part connection as a *relation.* The physical world is full of relations that link one physical object to another. Among these relations are the spatial relations, such as relations of distance. The phrase "is one mile away from" expresses a relation that can hold between two ordinary material objects, and perhaps even between two atoms or quarks. Relations like these are important in scientific theories. The relations of distance between two objects control the ability of those objects to collide with each other, or to push or pull on each other through gravity or other forces. Mereology forces us to recognize that the link between a part and the whole also is a relation. From a logical and mathematical standpoint, whenever we say "A is a part of B," we are expressing the same general kind of fact as when we say "A is a mile away from B."[36] We are saying that A and B stand in a certain *relation* to one another.

The observation that the whole-part connection is a relation seems obvious if you think about it long enough.

[36] The philosophers who hold that the whole is somehow identical to the parts must deny this, and instead must hold that the whole-part relation is different from ordinary relations like that of being a mile away from. See Lewis, *Parts of Classes,* pp. 84-85, for a position like the latter.

Yet in some respects, this observation runs counter to the usual ways of thinking about parts and wholes. Normally, we do not think of the bricks in a wall as simply items *related* to the wall, in the same way that Chicago is related to Atlanta by the relation "is north of." We think of the bricks (together with any other wall-parts, like mortar) as somehow *being* the wall. We do not think of the wall and the bricks as separate objects. Instead, we think of the wall on the one hand, and the bricks and other wall-parts on the other hand, as the same piece of stuff—the same substance. The suggestion that the bricks are simply objects related to the wall seems to leave something out—the fact that the bricks and mortar are the same piece of stuff that is the wall.

Now I am going to suggest a slight change in our way of thinking about material objects. Normally, we think of the brick as related to the wall in a certain way, and we also feel that all the bricks and other parts in the wall, taken together, somehow *are* the wall. The relation of "being a part of" seems different from all the other usual physical relations. This is the way it seems: if A is a part of B, then A, together with all the other parts of B, just *is* B. This does not hold true for other physical relations, like "is north of" or "weighs more than."

In place of this usual view, let us think of the brick as being related to the wall—*period.* That is, once we have said that the brick stands in the is-part-of relation to the wall, there is nothing more to say about the relationship between the brick and the wall. Of course, there still are a lot of details to settle, like exactly where the brick is located in the wall or how much mortar was used to attach it to the wall. But there is nothing fundamental left over. It

is not necessary to state the additional fact that the brick isn't merely related to the wall, but also somehow makes up the wall—because *there is no such additional fact.* To say that the brick is a part of the wall is to say that this object, the brick, stands in a specific relation to this other object, the wall. That is all.

According to the normal, intuitive view of the relation between part and whole, the bricks in the wall are what the wall is. On this view, each of the bricks shares part of the existence of the wall, as it were, and there is nothing to the wall besides the bricks hooked together in a certain way. According to the new view I am proposing, the bricks in the wall simply are related to the wall, just as a tree *to the north* of the wall is related to the wall. Of course, the relations involved are different; the tree is linked to the wall by the relation *is north of*, and the brick is linked to the wall by the relation *is part of.* But to pretend that the bricks *are* the wall in some way, while the tree *is not* the wall, is to miss the point. Once we have said that the wall and the brick are objects, and that the relation is-part-of holds between them, we have said all there is to say about the relation between brick and wall, except for incidental details. It is unnecessary to add something like, "But the wall really is just the bricks; it isn't a different object." Such a statement would not merely be redundant; it would be false.

This conclusion is the one to which the thought experiments in Chapter 1 pointed us. There I pointed out that a whole must be an object logically distinct from its parts. When we reflected on some simple wholes, we found that it does not make sense to regard the whole as being nothing but the parts. Even the arithmetic told us

that! Someone might have misunderstood my purpose in Chapter 1. The triangle, fish, and wave examples could be taken as arguments for a conventional holism which says that the behavior of the whole has no explanation in terms of the parts. But those examples are not arguments for that doctrine. Instead, they support the milder view that the whole is an object that exists in the world *in addition to the parts.* The question of whether the parts explain all the properties of the whole remains open. But the question of what kind of object the whole really is—an object in its own right, not identical to the parts that make it up—has been answered. The whole, whether or not it has a scientific explanation, is something other than its parts.

This new view of wholes and parts does not beg the question of the reducibility of the whole's behaviors to those of its parts. It is not a thesis about the behavior of the whole, but about what philosophers would call the *ontology* of the whole—that is, what kind of an entity, or being, the whole is. Regardless of whether the parts explain the whole, the whole is not the same object or being as any of its parts, or as all of its parts collectively.

There is a possible technical exception to this conclusion. This exception will not affect any of my future arguments, but I should mention it for the sake of completeness. Mereologists sometimes define the word "part" in such a way that an object is a part of itself. That is, the brick wall as a whole is part of the brick wall. If one chooses to define "part" in this way, then of course there is one part of the wall that *is* the wall; that part is the wall itself. But usually, when we speak of "parts" we mean parts that are not the whole. Throughout the book I will use the word "part" in this conventional way. I will not call

the whole a part of itself. I will have only one more occasion, much later, to mention the technical sense of "part" which makes an object its own part.

"The Sum of Its Parts"

This is a good time for some further comments on one traditional form of the holism-reductionism question: "Is the whole more than the sum of its parts?" The argument of the last section underscores the well-known fact that this question is too unclear to be answered as it stands. It is not clear what "the sum of its parts" really means.

Sometimes people who claim that the whole is the sum of its parts may mean that the whole is just the parts. In other words, if we have the parts, and arrange them properly, then that's all there is to the whole.[37] If this what we mean by "Is the whole the sum of the parts?", then I already have given the answer: no, the whole is not just the sum of the parts. The view that the whole is all of its parts collectively is simply illogical. But this is not the most reasonable reading of the question. Arithmetic teaches us that the sum of a series of numbers need not be the same as any of the addends that go to make it up. If we take the

[37] J.C. Smuts, the founder of holism whom I mentioned earlier, once wrote "the whole is not something additional to the parts: it *is* the parts in a definite structural arrangement and with mutual activities that constitute the whole" (*Holism and Evolution*, p. 104). Although Smuts assigned the whole a high place in the world, he would have disagreed with the view of part and whole that I am advocating. My view of wholes and parts neither implies nor excludes holism of a Smutsian sort, although my view might be regarded as holistic in a broader sense.

expression "sum of its parts" to mean, not the parts themselves, but the object formed by putting together the parts, then it is no longer implausible to regard the whole as the sum of the parts. Indeed, if the "sum of the parts" means whatever we get when we put the parts together, then the answer to the question is trivial. Of course the whole is the sum of its parts, for "the sum of the parts" is just another way of saying "the whole"!

If we read "the sum of the parts" to mean either just the parts, or what we get when we combine the parts, then the question "Is the whole the sum of the parts?" becomes easy. In one case, the answer is no; in the other case it is trivially yes. But these answers do not add up to holism or reductionism. We know that the whole is not identical to the parts, and we know that the whole is the object formed from the parts. But there still is plenty of room for holists and reductionists to disagree. They can debate whether the properties of the parts fully explain those of the whole. They can ask whether the whole contains any special factor or principle not foreshadowed in the parts. My claim that the whole is not the parts, and that it is an object existing in addition to the parts, may sound holistic. But my position does not rule out reductionistic explanations and does not settle all the pieces of the holism-reductionism controversy.

Substance Sharing

It seems as if the part-whole relation is "special" compared to other relations[38]—that there is something

[38] This intuitively appealing view has been well stated by Lewis. In *Parts of Classes* (p. 84), Lewis characterizes "mereological

radically different between it and, say, the relation of *being north of*. Some of our best established intuitions about reality suggest that the link between part and whole is more than just a relation—that somehow or other, the whole is just the parts. The source of these intuitions might be the fact that the parts contain all the matter contained in the whole. The bricks, together with the mortar (if any), contain all the matter that belongs to the wall; that matter is partitioned among these parts of the wall. Outside the bricks and other parts of the wall, the wall has no matter at all. But one must be careful before deciding that the wall *is* just the bricks and mortar, or that the being of the wall is just the being of the (properly arranged) bricks and mortar. In Chapter 1 I showed that two logically distinct objects can share the same energy and mass. Even if the bricks and mortar contain all the matter that is in the wall, this does not automatically imply that the bricks and mortar are the same as the wall. Instead, this may tell us that the bricks and mortar are distinct from the wall, but share substance with the wall. The wave example in Chapter 1 was one illustration of such *substance sharing.* To get an example of substance sharing which involves two objects of the same sort, visualize two water waves coming together and passing through each other. (Breakers may have trouble doing this without crashing to bits, but smaller water waves, like the ones in boats' wakes, do not.) At the moment of their overlap, the two waves encompass the same matter. Of course, the waves aren't just "things," they are processes. But the same kind of substance sharing happens with the wall and its bricks.

---

relations" as "something special".

The brick wall is not the bricks; the wall's existence is not the existence of the bricks, for the bricks could exist without the wall. The bricks are simply objects that share substance with the wall. This substance sharing is of the same nature as the substance sharing in the wave examples, here and in Chapter 1. Of course, the commingling of substance is more intimate in the wall. *All* the matter of the brick belongs to the wall, and *all* the substance of the wall is shared out among the bricks (plus perhaps a little mortar).

This sharing of substance by the part and the whole is what makes an object seem to be nothing but its parts. Early in life, we learn that every bit of stuff that makes up a physical object belongs to one or another of that object's parts. We learn that if you take away a part, you take away from the substance (and the mass) of the whole. If you take away all the parts, the whole vanishes. But this only shows that all the substance of the whole belongs to the various parts at the same time that it belongs to the whole. It does *not* imply that the whole is nothing more than the parts. The distinction between these two implications may seem subtle, but when one thinks about it, it becomes more and more blatant. The whole is there, and is not the parts—yet all the whole's substance happens to be the substance of the parts. (Perhaps this is part of the meaning of the idea of a part. To be a part of a thing X is, at very least, to "own" no substance except some of what X also "owns.")

The Stonemason's Argument

For an object to be a part, it is enough for that object to stand in a certain relation to a whole. No added equation of

the whole to the parts is necessary or possible. To make this claim more credible, I will point out that such an "added equation" could not be verifiable through experience. That is, once we know that the brick is a part of the wall, no extra observation could confirm that the wall is, or is not, just the bricks.

Consider a stonemason trying to build a section of a brick wall from a pile of bricks. Suppose that the mason is trying to restore a damaged wall that originally contained a single green brick as well as many of the usual red bricks. The mason asks himself "Is the green brick still in the wall?" He looks at the wall, and finds that the green brick still is there. Now he knows that the green brick is *part of* the wall. This knowledge enables the mason to do many things he could not do before. He can avoid building more green bricks into the wall if he wants to restore the damaged wall to its original color scheme. He can remove the green brick if he wants to make the wall more purely red.

Now suppose a reductionist philosopher comes along and tells the mason, "You already know that the green brick is part of the wall, and that the other bricks in the wall are parts of the wall. But there is another fact you should know: the wall is just the bricks and mortar. Strictly speaking, there is nothing there besides the bricks and mortar." Would this information enable the mason to do anything that he could not do before, when he only knew that the green brick was a part? Of course it would not! Once the mason knows that the green brick is a part of the wall, he understands the practical results of this fact (for example, that the wall will get more uniform in color if he removes the green brick). He does not need to worry about

the philosopher's claim that the wall is just the bricks. He can do the same things to the brick and to the wall, whether that claim is true or false. Nothing that he experiences will tell him whether the wall is just the bricks. All he ever needs to know is that the brick stands in a certain *relation* to the wall. And of course, there is no doubt for him that the wall and the brick both are legitimate objects—that both bricks and wall really exist. To work as a mason, he has to believe in the existence of the bricks and of the wall—or at least to behave as if those two facts were true.

Reflection on this example, and on other examples like it, will reveal that no sensory evidence can tell us whether the wall is or is not just the bricks. One can generalize this observation to all objects made of matter. All possible observations of parts of material objects are compatible with the belief that parts are objects related to the whole and sharing substance with it, instead of objects constituting the being of the whole. Even if the wall were just the bricks and mortar and nothing more, the mason would never find this out by doing masonry work. An experimental scientist studying the wall would not find this out either. Observations and experiments simply cannot answer the question of which belief is best.

I do not want to take up the old and well-known philosophical questions about verifiability and meaning here. For those who care, I will say that I am not a verificationist in any ordinary sense of that word. There are significant questions that sense experience cannot answer. Philosophy is full of questions of this kind. The argument about the stonemason shows that the question "Is the wall just the bricks?" is just such a question. We know, by inference from our observations, that the brick is a part

of the wall. We know, by inference from our observations, that the brick shares substance with the wall. If we assert in addition that the wall in some sense *is* its bricks and other parts, then we are asserting a metaphysical thesis that science cannot confirm or challenge. Also, we are adding a new relation—the identity relation between the whole and its parts—to the picture of what is happening.[39] We will never need this new relation to explain the observable behavior of the wall, since the existence of the ordinary whole-part relation, plus substance sharing, can do that. Neither the mason nor the scientist can bump into this identity relation, and the logic and arithmetic of whole and part (recall Chapter 1) suggest that this relation does not hold between whole and parts. It appears that there is no good reason to believe that this added relation of identity holds—and there are some good reasons not to believe it.

[39] Lewis holds that there is an identity of this sort; see *Parts of Classes*, pp. 81-85. D.M. Armstrong argues for a view in which the whole-part relation is a kind of partial identity (see *A Theory of Universals*, pp. 37-38).

Chapter 4.
Wholes or Just Parts?

Not everyone who thinks about wholes and parts arrives at the conclusions that I reached in the previous chapter. The philosopher David Lewis[40] has suggested a philosophical interpretation of mereology which is, in some respects, opposite to mine. Another philosopher, Donald L.M. Baxter, has discussed another version of the view that the whole is the parts.[41] Baxter also has discussed an opposing "Non-Identity view" of whole and part[42]. My conception of whole and part is a variation of what Baxter calls the Non-Identity view.

In this chapter I will discuss some of these philosophers' ideas about whole and part, and some of the arguments that philosophers have used to attack and defend these ideas. I will devote special attention to Lewis's view, as I understand that view. Then I will show where Lewis's interpretation of mereology goes wrong, and why his

[40] In *Parts of Classes*.

[41] Baxter, "Identity in the Loose and Popular Sense," pp. 578-581.

[42] Baxter, "Identity in the Loose and Popular Sense," pp. 578-579.

objections to the opposite view do not hit my interpretation at all.

Lewis claims that the relation between whole and parts is one of *identity.* To understand this claim we must know what philosophers mean by "identity." Identity is the relation that holds between things that are the same thing. If Antarctica is the southernmost continent on Earth, then we can say that Antarctica is *identical to* the southernmost continent on Earth. We also can say that Antarctica stands in the relation of identity to the southernmost continent on Earth. The expressions "Antarctica" and "the southernmost continent on Earth" name the same object, so the object named by one of these expressions is related by identity to the object named by the other.

In mathematics, the relation of identity is called *equality.* It is the relation that mathematicians represent by the equals sign =. If 2+2 is the same number as 4, then the number 2+2 is identical to the number 4.

These examples point up the fact that identity is a relation that relates every object to itself. Unlike other relations, which can relate one object to a different object, the relation of identity can only connect objects that are the same. The fact that identity can never relate two *different* objects makes it an unusual relation. Some philosophers have doubted that identity is a relation at all.[43] But even if identity were not a genuine relation, this would not change the fact that identity acts like a relation and can be treated

[43] These doubts are mentioned by Armstrong in *A Theory of Universals*, pp. 37-38. Russell discusses identity as a relation in *The Principles of Mathematics*, par. 95 (p. 96).

as one in formal reasoning. Mathematicians and logicians usually represent identity with the equals sign, =.

Philosophers have thought a great deal about the relation of identity. One philosophical question about identity is whether there is any such thing as *partial identity*—that is, whether two objects can be distinct or different in some respects, and yet somehow or other be the same thing. Some philosophers, including the philosopher of religion Charles Hartshorne, have argued that partial identity not only is possible, but also plays an important part in the world. Hartshorne suggested that the notion of a partial identity among beings provides a fruitful way to think about the moral unity or interconnectedness which, according to some religious traditions, exists among persons.[44] Another philosopher, D. M. Armstrong, has argued that the relation of part to whole is a kind of partial identity.[45] Armstrong has used this conception of partial identity in the study of a classic philosophical problem, the problem of universals.[46]

Mathematicians often use a trick like partial identity when they need to equate things that are not identical but only resemble each other in some respects. The mathematical device known as "equivalence classes"[47] lets

[44] Hartshorne, *Omnipotence and Other Theological Mistakes*, pp. 99-110.

[45] Armstrong, *A Theory of Universals*, pp. 37-38.

[46] Armstrong, *A Theory of Universals*, especially p. 38.

[47] This device is discussed in introductory texts on abstract algebra.

mathematicians make such thinking rigorous. As the philosopher W. V. O. Quine has pointed out, in the foundations of mathematics it sometimes is practical to regard things as being equal even if those things actually are similar only in certain respects.[48] But one does not have to believe in partial identity to do this.

In this book, I am not going to argue for or against the reality of partial identity. I am discussing these relations mostly to point out that one can ask serious philosophical questions about the seemingly simple idea of *being the same*. I will look into a different problem about identity: the question of whether a *single thing* can be identical to *several things together*. This is the kind of identity that Lewis claims to exist between any whole object and its parts.[49] (I should mention that Lewis does not seem to deny the reality of composite wholes.[50])

In mathematical logic, statements about what exists are couched in the language of *quantifiers*. A quantifier is a phrase like "There is" or "For all" which tells us how many objects or entities have some property. For example, in the sentence "There is a brown dog," the phrase "There is" acts as a quantifier. Because of the presence of the phrase "there is," that sentence tells us that there is at least one

---

[48] See Quine's remarks on equality and identity in his book *Set Theory and Its Logic*, pp. 14-15.

[49] See Lewis, *Parts of Classes*, pp. 81-85.

[50] I think this is clear from his remarks in the footnote on p. 70 of *Parts of Classes*.

brown dog. In the (false) sentence "All dogs have tails," the word "All" acts as a quantifier. It tells us how many dogs have tails: they all do. (Of course, "All dogs have tails" does not tell us the exact number of dogs that have tails. Depending on how many dogs currently exist, there may be one tailed dog or a million. But it does tell us how many dogs do *not* have tails: zero. Thus it is a statement about quantity.) There are other quantifiers that are more complex, but I will not deal with them here. My aim is not to provide a lesson in mathematical logic, but to say enough about quantifiers to make clear the central idea of Lewis's argument.

Lewis points out that not all quantifiers say something solely about individual objects.[51] Quantifier phrases like "There is" and "All" say that there is an individual object of a certain sort, or that all individual objects have a certain property. But people often reason about groups of objects as well as about individual objects. For example, one could say "In this field, some dogs formed a pack." In this sentence (which is not Lewis's example), the word "some" acts as a quantifier. However, that word does not only say that individual objects exist. Instead, it says something about a *group* of dogs—a group that formed a pack. In this particular sentence, "some dogs formed a pack" means this: there are *several* dogs, which *together* have the property of having formed a pack. It does not mean that there is at least one individual dog that formed a pack. (After all, no individual dog, considered alone, can form a pack.)

In "Some dogs formed a pack," the word "some" acts

[51] Lewis, *Parts of Classes*, pp. 62-71.

as a *plural quantifier*. That is, it is a word or phrase that declares the existence of a plurality of things, or of several things, which *together* have some property. This word or phrase does not simply declare the existence of individual things, *each* of which has some property.

Lewis's arguments imply that plural quantification is a legitimate part of logic that does not pose any fatal philosophical difficulties.[52] According to this view, it is legitimate to speak of *some* dogs having the property of having formed a pack, just as it is legitimate to speak of *a* dog having the property of having *joined* a pack. This subtle and technical thesis in philosophical logic has a great impact on our view of the relation between wholes and parts. As Lewis knew, it suggests that there may be a logical way to regard a whole as being nothing more than its parts.

Consider the claim that a particular dog is just some dog-parts in a certain arrangement—that there's nothing else to the dog beyond that. This is a claim that a reductionist might love. But just what could a reductionist mean by this? Mainly that a dog is identical to its parts. This claim implies that once we have listed the dog-parts, we do not also have to list the dogs themselves to get a thorough inventory of everything alive in the kennel. However, we cannot truthfully say that the dog *is* just the parts unless the parts can have the property of *being a dog*. This particular reductionist claim cannot possibly be true unless the parts can literally be the whole dog.

It is plenty clear that the dog cannot simply be identical

[52] Lewis, *Parts of Classes*, pp. 62-71.

to any one of its parts.[53] But it is equally clear that identity of the ordinary sort, which is mentioned in sentences like "2+2 is identical to 4," cannot hold between the dog and all of its parts together. For the dog-parts (as distinguished from the dog they make up) are many things. The dog is one thing. Ordinary identity or equality links objects which are the same—or, more correctly, it links an object to itself. But the relation between the dog and its parts does not link an object to itself. It links an object which is a dog to many objects, none of which are dogs. Therefore, the relation between the dog and its parts cannot be the ordinary relation of identity.

Some philosophers already have made this objection to the identity of whole and parts. The objection takes various forms in their writings.[54] The objection seems airtight, but Lewis's position shows a possible way around it. Lewis has proposed that we think of the relation of whole to parts as a genuine relation of identity, but one different from the simple relation of identity that holds between 2 + 2 and 4. The relation between the dog and the parts is one of plural identity. This relation relates a thing to *some things,* not simply to another thing. And this is where Lewis's ideas about plural quantification come to bear on the problems of whole and part. If plural quantification is a legitimate part

[53] Except itself, if one counts the dog itself as a part of the dog. As I said earlier, I elect not to count an object as a part of itself, though the opposite choice is commonly made and is just as logically sound.

[54] Baxter describes a similar, though different, objection in "Identity in the Loose and Popular Sense," pp. 578-579. Lewis (*Parts of Classes*, p. 84) mentions another similar-though-different objection.

of logic, then statements like “Some parts are a dog” make sense. It is possible to say “Some parts are a dog” and mean that the dog is *all* the parts, instead of meaning (absurdly) that each part is a dog. If we use plural quantification to describe the world, then we can describe the relation of plural identity between a whole and all its parts. Therefore, if we allow logic to include plural quantification (as Lewis suggests that we do), we can easily extend it slightly further to allow for plural identity (as Lewis does).

The use of plural quantifiers makes it easier to state the claim that the whole is just the parts. If we can say that all the dog-parts collectively have a certain property, then we can say that those parts collectively are the whole.

Of course, the fact that it is possible to make such a statement without contradicting oneself does not imply the that statement is true. The statement “The earth is a cube” is not obviously self-contradictory, but it happens to be false. And even if we accept Lewis’s views on the nature of plural quantification, this does not automatically imply that the dog really is its parts. To see whether this further conclusion is true, we must examine in more detail Lewis’s views about mereology.

On Lewis’s view, the whole-part relation of mereology is a relation of partial identity. In other words, the dog-parts taken together are identical to the dog. Each of the parts is, as it were, credited toward the being of the dog; the dog’s being is nothing more than the being of all its parts, considered simultaneously. On this view, the whole-part relation is a kind of identity.

Lewis’s work shows that it is possible, using plural quantification, to speak of the whole-part relation *as if* it

were a kind of identity. But this, alone, does not imply that the relation *really is* a kind of identity. Lewis himself recognized this; he gave examples of ways to speak of other relations, which clearly are not identity, as if they were kinds of identity.[55] Lewis knew that even if we can talk about a relation in a way that makes it *sound like* identity, we cannot be sure that the relation really *is* a kind of identity. But Lewis also decided that the whole-part relation is a kind of identity. I am arguing for the opposite conclusion. Plural quantification may let us speak as if the dog were its parts, but it leaves open the possibility that the "are" in the sentence "These parts are this dog" does not express genuine identity. An opponent of Lewis could say "Yes, these parts 'are' this dog—but the 'are' in that sentence doesn't stand for identity."

On Lewis's view, the whole-part relation is one of plural identity, and is a legitimate kind of identity. But an opponent of this view remains free to argue that the relation of plural identity is not really a relation of identity at all, and should be called something else. Perhaps *plural* identity is not a kind of *identity*, just as a full house (in the poker sense) is not a kind of house. Perhaps so-called partial identity only resembles identity in certain key respects. (Lewis noted such a resemblance, but used it to support the view that partial identity is a type of identity.[56])

Mathematicians sometimes use relations that closely resemble identity but are not genuine examples of identity.

[55] Lewis, *Parts of Classes*, p. 84.

[56] Lewis, *Parts of Classes*, p. 84.

This lends weight to our suspicion about Lewis's view. As I mentioned earlier, mathematicians often treat objects that are not identical at all as if they were identical—and mathematicians do this without the slightest threat to the consistency of their reasonings. The trick is the method of equivalence classes, which uses the idea of an *equivalence relation.*[57] An equivalence relation has many of the algebraic properties of the identity relation—for example, it relates each object to itself. It also comes close to many of the logical properties of identity—for example, if two objects are equivalent, then they share some of their properties (though not necessarily all of their properties, as would happen with genuine identity). But there is no question that most equivalence relations are not relations of identity. Lewis has shown that the whole-part relation resembles identity more closely than we previously had suspected. But does that prove that it is identity?

This, then, is the first part of my objection to Lewis's position: our ability to treat the whole-part relation as an identity relation does not imply that it actually is an identity relation. But Lewis's argument for the identity of whole and parts does not rest solely on the fact that the whole-part relation is formally like identity. Rather, it rests on weightier philosophical considerations. What I take to be the crux of Lewis's argument is summarized in the following quotes from Lewis's book *Parts of Classes*. Speaking of "cat-fusions" (wholes built up from cats), Lewis argues that a cat-fusion is just identical to the cats it

[57] Equivalence relations are discussed in various texts on abstract algebra.

contains: "Take them together or take them separately, the cats are the same portion of Reality either way[...]." Lewis then goes on to say that in an accounting of all there is, "it would be double counting to list the cats and then also list their fusion."[58] These two quotes together express the view of wholes and parts that I have been trying to undermine throughout this book. By asserting that one does not need to count the cat-fusion as well as the parts, Lewis presupposes that the cat-fusion is its parts, in some sense of "is." But more importantly, the quote reveals a central intuition that appears to underlie Lewis's position. This is the feeling that the cat-fusion must be the cats because it is made of the *same stuff* as the cats. Add up all the cats, and you have the same portion of substance—or "portion of Reality" as Lewis puts it—that you find in the cat-fusion.

This intuition seems to support the view that the cat-fusion is just the cats that make it up. But this evidence is not so weighty if one recalls the idea of substance sharing, which I set forth in Chapter 1. Lewis used the phrase "portion of Reality," but another way of putting it is that the cats *share the same substance* as the cat-fusion. The substance of the cat-fusion is exactly the same stuff as the substance of all the cats together. And the fact that two objects share substance does not automatically make them identical. The water-wave experiment in Chapter 1 pointed to this fact.

The main reason Lewis's position seems plausible is, I think, the fact that the whole is made of the same stuff as the parts. Presumably, this is at least part of what it means

[58] Lewis, *Parts of Classes*, p. 81 (for both quotes).

to say that the whole is "the same portion of Reality" as the parts. But this fact does not support the identity of whole and parts. Our intuitions may make us *feel* that it supports this identity—but that feeling only shows that our intuitions about whole and parts are inadequate, as I argued in Chapter 1.

Using Lewis's cat-fusion example as a start, I will now set forth the rest of my objection to Lewis's conception of the whole-part relation. To do this, I will use the relationship between cat-parts and a whole cat rather than that between a cat-fusion (a less familiar object!) and a single cat.

The first point in my second objection is this: the claim that there is a separate cat, besides the cat-parts, is impossible to refute or confirm scientifically. I argued this point in an earlier chapter, using a brick wall instead of a cat as an example. But even if the assumption that there is a separate cat does not help us explain our experiences, it does help us to understand them properly. Indeed, we *must* make this assumption if we want to avoid falling into nihilism—the view that nothing exists at all.

The cat is the whole that exists when certain parts are united in a certain way. Once we have admitted that the cat-parts exist, and have admitted that those parts are hooked together as they are, we no longer can deny there is a cat. If we think that this denial is acceptable, then it would be sheer arbitrariness not to extend it to other composite objects besides cats. And if we do this, we have to deny that any object divisible into parts is real. In the next chapter I will show that this view is self-contradictory; for now I will simply point out that it is not a view that a sensible person should adopt. If there are no composite

objects, then there literally are no things other than the ultimate constituents of matter. This means no atoms, no bricks, and no human bodies. (As I will show later, this view also implies that if the subatomic structure of matter happens to be a certain way, then there is nothing at all!)

Of course, Lewis does not adopt any of these conclusions. But these conclusions do follow from Lewis's view of identity, if we take that view to its logical endpoint. To escape these conclusions, we must stop short of that endpoint by changing Lewis's view to allow a whole *distinct* from the parts into our picture of existence. There are cat-parts—but we cannot fully understand the world until we admit that there is, in addition, a cat. However, once we have admitted the existence of the cat in addition to the existence of its parts, then the whole-part relation cannot be an identity relation of any ordinary sort. Whatever kind of relation this partial identity is, it is not the kind of identity that would let us say, "There—we've counted all the cat-parts. Now we don't need to count anything else to find out which entities just went up into that tree." Even if the parts in some sense *are* the cat, that sense of "are" cannot be one which excludes the additional existence of the cat.

Lewis's part-whole "identity" relation is not like what we usually call "identity." Antarctica is identical to the southernmost continent on Earth; thus, Antarctica does not exist *in addition to* the southernmost continent on Earth. To be complete, our ontology needs to contain only one of these continents. This is the hallmark of real identity: "two" things are identical if they really are one thing. But if our ontology contains all the cat parts, and those parts are arranged in a catly way, we still have left something out

until we let the cat in. The reality of the parts, plus the fact that the parts are connected in a suitable way, *implies* the reality of the cat. Nevertheless, the parts are not the cat.

The cat-as-a-whole helps us to interpret a certain fact in our experience. I am speaking of the simple and obvious fact that *cats exist.* The reductionist view may be "right" when used solely as a rule of scientific method. For practical reasons, we should try to explain the properties of wholes in terms of the properties of their parts, and often (perhaps always) we can do this. But even if the properties (including behavior) of the cat can be "explained away" in terms of cat parts, we still are stuck with the experienced fact that *there is a cat.* Only the existence of a real cat can make this fact true. If strictly speaking there is no real cat, then there simply is no cat at all—and adding the weasel words "strictly speaking" does not change the impact of the conclusion that *Tabby does not exist.* If we do not want to drop cats from our picture of the cosmos, and embrace a skepticism about cats as total as the skepticism with which Descartes contended, then we must admit that "There is a cat" is true. And once we have admitted this truth, we can assume just enough objects to explain why this fact is true. Only one object will do: a real cat.

The preceding arguments show that Lewis's view of the whole-part relation cannot stand up to the facts of experience and to the demands of logic at the same time. Lewis suggests that we regard the relation between the whole and its parts as a type of identity. This suggestion, if followed, would issue in the view that the whole is the parts, and that there is no whole at all beyond the parts. But we now know that we should not embrace that view. If we do decide that parthood is a kind of identity, we also

must concede that this kind of identity is quite different from what we normally call "identity." In particular, this identity relation must leave room for the existence of a new object—the whole—as well as the parts. If we believe that Lewis's whole-part identity relation is like this, then it is not really an identity relation, and Lewis's position loses force. We can call this relation a very strange kind of identity, or if we prefer, we can call it not quite a kind of identity. But no matter what we call it, it must relate the parts collectively to a whole that exists in addition to those parts. If we count everything that's up in the tree, we still must count all the cat-parts plus the cat.

Chapter 5.
The Vital Relation

The "parts as whole" view fails to capture some features of the part-whole relation. This failure supports my proposed view of that relation. Earlier I said that people often think wrongly about the part-whole relation. We do not normally think of that relation as a relation between distinct objects, like the other basic relations of physical science. We can see intuitively that spatial and temporal relations are relations between terms that may be distinct. If object A is one mile due north of object B, then A and B are distinct, but are related in a certain way. If event E is one second earlier than event F, then E and F are distinct but are related in another way.[59] If X is a part of Y, then X and Y are related in still another way—but we tend to feel that this is a "special" relation, different from the rest.[60] The spatial and temporal relations normally connect

[59] These examples ignore certain possibilities suggested by the general theory of relativity. In certain extreme examples of curved spacetimes, an object may be one mile north of itself, and an event may even be one second earlier than itself. The examples here hold good in any reasonably "normal" spacetime.

[60] As I mentioned earlier, Lewis characterized "mereological

distinct objects. For the whole-part relation, we feel that the whole is *not distinct* from the parts, that it is *not separate* from the parts. Sometimes we feel that the part-whole relation is not a relation of an object to another object, as much as it is a relation of an object to itself.

These intuitive feelings are inaccurate, but it is understandable that we have them. There is no point in denying that the part is just a piece of the whole, or that the part is not spatially separate from the whole. Spatially, the part is inside the whole (provided that the whole is an object located in space). And as I pointed out earlier, the substance of the whole is just the substance of its parts. The intuition that the whole-part relation is different from ordinary physical relations is well-motivated in this respect.

However, there is another respect in which our unschooled intuitions about wholes and parts go grievously wrong. We are in error if we feel that a whole object isn't anything but its parts. I have spent the last three chapters trying to demolish this seemingly natural view. But there also is another error—one that creeps into much of our thinking about wholes and parts. This is the feeling that the relation between whole and part somehow is *contained in* the whole or the part.

We are not normally aware that we have a feeling of this sort. Probably not everyone has this feeling. But the following thought experiment will show that it is easy to get this feeling if we think just a little about wholes and parts.

Imagine a ham sandwich. Now mentally lift the bread

---

relations" as "something special" (*Parts of Classes*, p. 84).

off the top of the sandwich, and move the displaced piece of bread to the opposite side of the room. After this operation, the sandwich no longer exists as a whole; that is, the room contains no object that qualifies as a sandwich. (Of course, one could call what remains an "open-faced sandwich" plus a loose piece of bread, but this is not a sandwich in the strictest sense of the word.) Now bring the piece of bread back across the room, and shove it back down on to the rest of the sandwich. Presto—a sandwich is born.

Ask yourself this question about the last step in this experiment: When we brought the bread and the remaining part of the sandwich back together, what did we have to add to make sure they really formed a sandwich? Answer: Nothing! A sandwich consists of bread and other foodstuffs in a certain combination. And when we say that the top slice of bread is *part of* the sandwich, we are admitting that the top slice of bread, with the other materials, is arranged in a way that creates a sandwich. There is no special tie, no special relation or logical "glue," needed to make the bread part of the sandwich. The bread's being part of the sandwich is a consequence of the make-up of the sandwich—the way the sandwich is arranged. Once we have made up the sandwich, then we do not also need to create an extra *relationship* between bread and sandwich to make the bread part of the sandwich.

This simple kitchen experiment leads us to an interesting finding: the relationship between a sandwich and its parts can seem to be an aspect or facet of the sandwich itself. This relationship between bread and sandwich seems different from other relations, such as *being north of.* For a piece of bread to be north of a

sandwich, the bread and the sandwich must be arranged in a certain way with respect to the Earth (the bread must be closer to the North Pole). But for a piece of bread to be part of the sandwich, the Earth is not required, and neither is any other external object. All one needs is the bread and the sandwich. The relationship seems to be "in" the sandwich, not "outside" of it in the form of an extra, added relation.

This feeling that the sandwich-bread relation is "in" the sandwich is a mild version of an old and honorable philosophical idea. I refer to the traditional philosophers' distinction between *internal relations* and *external relations*.[61] Many philosophers have claimed that there are two kinds of relations. Some relations are "external": that is, things can have them, but they are not "built into" the things that have them. *Being north of* is an example of such a relation; it holds between two objects if those objects happen to stand in a certain relation to the Earth, and not simply because of what the objects are. Another example is the relation of *being older than*; this depends on the time elapsed between the moments when different things begin. The "internal" relations, on the other hand, are relations that are facets of the things themselves—relations that link things because of what the things are, or relations that are "built into" the things they relate.[62] An

[61] This distinction is discussed in (for example) Armstrong's book, *A Theory of Universals*, pp. 84-85.

[62] Armstrong gives a more precise and more adequate definition (*A Theory of Universals*, p. 85).

example might be the relationship between a printed road map and a road shown in the map. The map is a map of that particular road because of the characteristics of the map itself—namely, what is shown on the map. The map is related to the road because of the map's own internal characteristics.

Not all philosophers have believed there is a difference between external and internal relations. Indeed, many twentieth-century philosophers have ignored this difference. I am not going to take up this issue here. This much is clear: normally, *we think of the whole-part relation as if it were an internal relation.* That relation has the psychological "feel" of an internal relation. It seems to hold just because of what the whole is, and not because both whole and part are joined by some third factor, some "external" tie or bond.

The naive view that the relationship between whole and part is internal is one of several ideas that I am challenging in this book. I maintain that the whole-part relation is not internal; it is not just a side effect of what wholes and parts are. My earlier arguments against the identity of a whole and its parts should provide a clue to why I am making this claim. One cannot think of the whole-part relation as internal, because one cannot equate the whole to its parts. Instead, the whole is one object, and the parts are other objects; by arranging the parts correctly, one can *cause* the whole to come into being, but that is not the same as saying that the whole *is* the parts. Certainly the behavior of the whole is strictly regulated by that of the parts; if all the parts are moving east, the whole cannot simultaneously

move west.[63] But the whole is not simply the parts; it is a separate, distinct object, whose existence and properties happen to reflect the state of the parts. Thus the relation between whole and parts is more correctly thought of as an "external" relation between distinct objects. That relation is not a built-in facet of the objects' nature, but is a relation into which two objects may enter as a result of physical circumstances, and which ties those objects together in some way. In this respect, the relation *is part of* is like the relations *is north of*, *is older than*, and *is deeper in the ocean than*. If the relation holds between two things, it does so because those two things are joined or placed together in a certain way. The fact that the relation holds is not simply a side effect of the nature of the things involved.

There is another peculiarity of the external-internal relation distinction that suggests that if that distinction has

[63] Baxter ("Identity in the Loose and Popular Sense," p. 579) notes that there is something wrong with the idea that one can sell all the pieces of a parcel of real estate and still claim to own the parcel itself. Baxter suggests that this example supports the identity of the whole with its parts. But the view I am presenting can handle this example just as well as can Baxter's view. According to my view, the parcel is not identical to all of its pieces collectively. However, the *substance* of the parcel—the land, or earth materials, of which the parcel is composed—is shared out among the pieces. Because of this substance sharing, if one gives away all the pieces of the parcel, one has no land left. Hence the scam artist in Baxter's example, who claims to own the whole parcel and not its parts, is wrong in thinking that he still owns any land. Of course, it is thinkable that the laws of the country in which the land is situated might still allow him some kind of formal ownership of the parcel. But even if this were the case, he would not in fact have any *land* at all.

any force at all, we must put the whole-part relation on the external side of the divide. This is the fact that the whole-part relation, at least for physical objects, involves *space* in an essential way.[64] Most relations involving space—such as *is north of, is above*, and the like—seem to be external relations. But the whole-part relation, at least as it applies to physical objects, clearly involves space. A brick cannot be part of a wall unless its position in space is within the boundaries of the wall. This tells us that the whole-part relation cannot hold between two space-filling objects unless the relation of *being spatially within* also holds between those objects. If A is a part of B, then A is inside the spatial boundaries of B. Of course, there is more to being a part than just being on the inside. A brick in a box is inside the box but not a part of it, and a board driven through a tree by a hurricane is not a genuine part of the tree. But a physical object A cannot be a part of another physical object B if A is somewhere else besides where B is. Thus, the whole-part relation for physical objects depends on the presence of an external relation. It cannot be simply a byproduct of the nature of the whole involved, since that external relation is not a byproduct of anything's nature. So the whole-part relation cannot be just an internal relation.

[64] For an argument that this relation does not always involve space, see Armstrong, *A Theory of Universals*, p. 37. But physical objects' parts are related to them spatially in a certain way.

[Pages Deleted Here]

In Conclusion

In the last three chapters, we began with a look at mereology and went on to do some experiments with wholes and parts. Along the way, we drew several conclusions that will be important in the rest of this book.

The first and most important conclusion is that an object made of parts is not identical to those parts. A brick wall is not just a bunch of bricks—though it is made of bricks. Rather, a composite object is a separate object, additional to and quite distinct from its parts. Of course, the whole is not independent of its parts; it comes into being when the parts are arranged and interconnected in the right way. The whole shares the substance of the parts, and its properties are largely (and perhaps completely) fixed by the properties and relationships of its parts. But this does not imply that the whole *is* its parts.

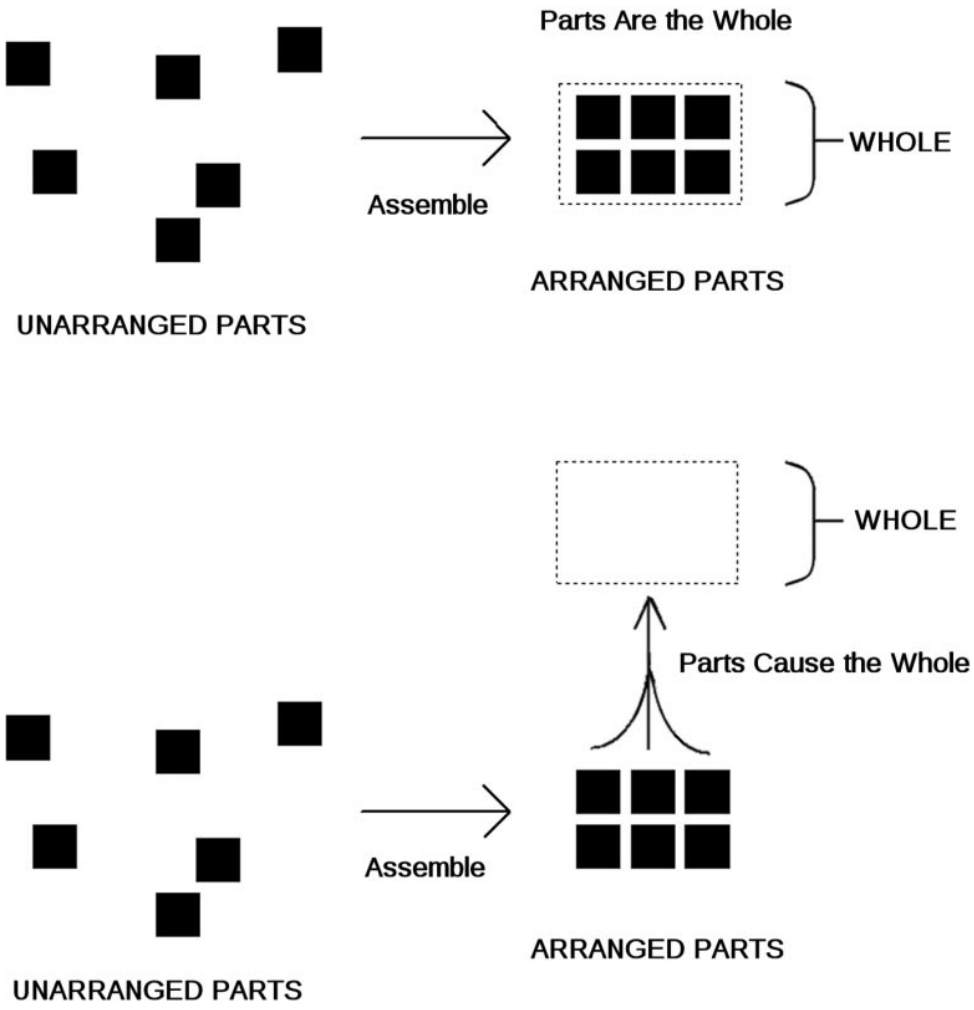
The relationship of parts to whole is not one of identity, but one of *causation*. The parts, by occurring in a particular arrangement and in particular states, cause the whole to come into being. If you start with N parts, and put them together in a suitable way, then (*bam!*) you have N+1 objects on your hands.[70] You create an extra object. This object is not created out of nothing, for all of the substance that makes up the parts also belongs to that extra object. After the extra object comes into being, the substance of a

[70] This is another variation on the comparison of N with N+1 objects that I earlier credited to Baxter ("Identity in the Loose and Popular Sense," p. 579).

part no longer belongs only to that part. Instead, that substance belongs to the part and also to the extra object. The substance is not divided among the part and the extra object. Instead, it belongs in full to both, in an arrangement that a lawyer might call joint ownership.

If one disbands the parts, the whole ceases to exist. As long as the whole exists, the properties of the whole are determined by the properties and relations of its parts. (A holist might amend the last statement to say that many, but not all, of the properties of the whole are determined by the properties of the parts. But the existence of the exceptions would not undo my point.)

To make the new picture of whole and part clearer, I have sketched this idea schematically on the next page, alongside a diagram of the conventional idea of whole and part. Visual metaphors for abstract concepts can be treacherous, but hopefully this drawing may be a little more effective than words in getting my point across. When parts come together to form a whole, something else happens besides the convergence of the parts. Specifically, a new object comes into being—an object that is logically distinct from the parts, and not identical to them in any reasonable sense of identity. This object is the whole. In many respects, it is a product of the parts; the coming together of the parts, in the proper arrangement, gives the whole its existence and its properties. But the whole is not the parts.



Classical (top) and new (bottom) conception of the whole-part relationship. Traditionally, the whole is regarded as being nothing but the parts in a certain arrangement. But it is more logically coherent to suppose that the whole is an object distinct from the parts, which comes into existence when the parts are properly arranged.

Note added later: The lower half of this diagram is meant to suggest that the whole is distinct from its parts, but not that the whole is independent of its parts. Read the text for details.

One consequence of this new picture of the whole is that the relationship between wholes and parts is not what is called an internal relation. The fact that something is a part of some whole is not merely a fact about the make-up of that whole. Instead, it is a fact about a relationship between two distinct objects—analogous to the fact that a certain object is to the north of another, or is older than another.

This picture of the whole-part relationship is radically different from the usual picture. According to the usual picture, the brick wall is nothing but the bricks that make it up, and the human body is nothing but the atoms of which it is composed. Our new view, despite its apparent oddness, is more logically coherent than the traditional view. The view that the whole is the parts raises problems about identity and difference—problems that we cannot solve without some serious fudging of the concept of identity. Also, the usual view makes it difficult to imagine how anything besides tiny particles could exist at all. The new view, that the whole is a separate object and is an effect of the parts, does not have these defects. The new view may seem repulsive to some intuitions and offensive to some philosophical positions. Actually, it comes closer to common sense than does the old view. In ordinary life, everyone assumes that there are brick walls and human bodies, ham sandwiches and planets. The discovery that these items can be broken down into atoms and particles does not lead people of common sense to decide that ham sandwiches don't exist after all. Our "new" view of whole and part agrees with the commonsense attitude that composite objects really exist. The opposing view is surprisingly difficult to reconcile with this attitude.

These considerations support the new picture of a whole distinct from its parts. There also is another, more important consideration: the new picture simplifies some of the major problems of philosophy. Some of the most important traditional philosophical riddles rest on subtle confusions about the whole-part relation. If we reread these problems with the new view of wholes and parts in mind, we will find that the problems look much less puzzling than they did at first. This is what I will do in the rest of the book.

Note added later: Part of this page has been omitted.

[Pages Deleted Here]

The Ultimate Reality of All Things

The second and third important points in our new view

Note added later: Most of this page was irrelevant to this shortened book and hence was blanked out. Read on.

of the universe are ideas that have come up again and again throughout the book. These are the ideas that *wholes are not identical to their parts in any sense*, and that *all wholes are constituents of reality as fundamental and irreducible as the ultimate particles of matter*.

Scientists, philosophers and others sometimes speak as though the universe consisted of elementary particles and nothing else. This idea, which may have existed even before Democritus's view of the universe as "atoms and Void,"[141] crops up again and again in writings about philosophy and science, though usually in disguised or implicit forms. According to our new view of reality, this idea is completely and utterly wrong. If by "the universe" one means "the natural world," then the universe consists of many kinds of objects—not just elementary particles. It includes birds, and icicles, and silver crystals, and galaxies, and human brains—all of which are *just as fundamental*, and just as ultimately real, as the elementary particles. If you want an answer to the age-old question "What kinds of things really exist?", then one obvious answer, "Elementary particles of matter," is wrong. A more correct answer is: *all things*. All objects made of material particles are parts of the irreducible foundation of ultimate reality. All of them are constituents of the world and cannot be equated to simpler or more fundamental constituents. And this is the case *even though all of these things are wholes built up from elementary particles and from nothing else.*

An inventory of the ultimate constituents of reality—of

[141] Democritus, quoted in Freeman (ed.), *Ancilla to The Pre-Socratic Philosophers*, p. 93.

the ontology of the world—would have to include silver crystals, starfish, the Mona Lisa, and galaxies as well as leptons, quarks, and bosons. The doings of the leptons, quarks and bosons may *cause* those other items in the list to come into being. But that does not mean that the other items *are* just leptons, quarks and bosons at bottom, or that the other items are less fundamental or real than the so-called "fundamental" particles.

All this does not lessen the importance of the physicists' quest for the final constituents of matter. This quest, which currently takes the form of the search for a "theory of everything," is an important expression of the human spirit. The success of this magnificent endeavor would immeasurably deepen our understanding of reality, and would provide an answer to another age-old philosophical question: "What is the material world made of?" But no scientific theory can prove that a brain, a flower, or a piece of silver is only elementary particles.

Emergent Properties Again

The fourth major thesis in our new picture of reality is perhaps a corollary of the second and third. It is that *the emergent properties that belong to complex wholes are as real, and as central to reality, as the measurable physical properties of simple material particles.*

This is an important point for several reasons. With the second point, it enables us to find the place of *values* in the natural world. Values, such as goodness and beauty, are emergent properties. A situation or event is good because of its effects on people, and perhaps for other reasons as well. The goodness is not in any one of the elementary

particles that take part in the good situation. Instead, it belongs to the situation as a whole. A flower is beautiful because of the way its parts are arranged. Without that arrangement, there would only be atoms, not a flower. (An atom may be beautiful too, but that is a different kind of beauty, best known to scientists. Presumably, that kind of beauty is emergent as well.) Properties like goodness and beauty depend on the structure of complex wholes. This is not to say that values always are emergent properties; the smallest constituents of matter might, for all we know, have goodness and beauty. (My personal suspicion, which I won't try to defend here, is that they do. At very least, the final mathematical theory that describes these particles is likely to be beautiful.) But the goodness we find in everyday life, or the beauty we find in art and nature, belong to complex wholes. These values, in their present forms, would not be there if the wholes dissolved into particles. Therefore, those values are emergent properties.

One hears it said that the "scientific" view of the world—that of a world built from material particles—leaves no room for values. Our new conception of wholes and parts shows that this claim has no basis in fact. If emergent properties are ultimately real, and values are emergent properties, then values are ultimately real. *The goodness of a compassionate deed is as much a part of fundamental reality as the charge of the electron. The beauty of a wildflower is as much a constituent of the cosmos as the mass of the proton.* A worldview that leaves room for wholes and their emergent properties leaves room for values, which are, at least usually, emergent properties of wholes.

This does not mean that our theory of wholes and parts

can settle any specific moral or aesthetic controversies, either about the natural environment or about anything else. But it can put to rest the old saw that a world made of matter has no room in it for values.

Aesthetic Experience: A Road to Truth

This conception of value has another important consequence for our understanding of human experience and its relation to the world. This has to do with the nature of *aesthetic experience*. By "aesthetic experience," I mean experience centered on emotions and feelings produced by something a person is observing. This type of experience includes perceptions of beauty, such as the feelings produced by art and nature. However, it also includes other feeling-centered experiences that may not fall directly under the heading of "beauty." For example, there is the sensation of mystery that one sometimes has in the presence of a lowering dark sky. There is the distinctive sensation of "farawayness" that accompanies certain summer days. There is the unique emotional tone that one sometimes feels around *trees*—a tone different from the one that one feels around, say, flowers or grass. And there is the feeling, caused by some works of art, that there is more in the artwork than one sees—a feeling that one is about to discover something that is not visible at first glance. All of these subtle and elusive feelings are examples of aesthetic experience.

There is a common belief that aesthetic experiences are subjective—that they are "in the eye of the beholder," as the familiar saying about beauty goes. Many people think of aesthetic experience as something that lies entirely in the

mind of the observer, and does not accurately reflect any feature of the external world. According to this view, experiences of physical properties, like the size and weight of an object, may yield real knowledge about external objects, but experiences of beauty and emotional tone cannot. Those who hold this view believe that aesthetic feelings and experiences can tell us about our own states of mind, but not about things out there in the world. Real knowledge, on this view, comes from science, and perhaps from philosophy and theology, but not from art. The arts, it is believed, can reveal beauty, cause enjoyment, and even communicate ideas, but cannot supply us with any new truths or knowledge, except for some knowledge that we could obtain by other means. This view, at least in its simplest and less reflective forms, seems to be very widely held. Critics and philosophers have proposed more elaborate versions of the idea, arguing for example that poetry can have an "emotive" function but cannot reveal truths about the world as science can do.[142]

If aesthetic qualities are ultimately real, then this last view is wrong. Aesthetic experiences *do* teach us something new about reality—something that neither science nor philosophy can discover or verify. The

[142] I. A. Richards made a claim of this general sort. Richards's views are discussed briefly in the articles "Belief, Problem of" and "Pseudo-statement" in *Encyclopedia of Poetry and Poetics* (Preminger, ed.). Various ideas on poetic truth, and on the question of whether poetry can reveal or convey truth, are discussed in the following articles in that reference volume: "Belief, Problem of"; "Meaning, Problem of"; "Criticism" (especially pp. 161-162); "Poetry, Theories of".

qualities of feeling that we find in our aesthetic experiences are *real features of the world*—emergent features that are not reducible to the physical properties of material particles. And since emergent properties belong to ultimate reality (as we saw in earlier chapters), aesthetic experience is a form of experience of ultimate reality.[143]

There is an obvious rebuttal to the claim that the aesthetic qualities of the world are real. This comes from the fact that different observers get different experiences when observing the same object. This common observation is the root of the saying that "beauty is in the eye of the beholder." But actually, this does not contradict the reality of aesthetic qualities at all. This observation simply shows that an object can have more than one set of aesthetic qualities, and that an observer who interacts with the object may find any one (or more) of these sets of qualities. A poet who sees beauty in an apple tree is finding a particular set of aesthetic qualities in the tree.[144] Another poet may find different aesthetic qualities, seeing

---

[143] Some philosophers, including some Platonists, have regarded art as affording contact with ultimate reality (see, for example, the article "Platonism and Poetry" in Preminger (ed.), *Encyclopedia of Poetry and Poetics*). Most such philosophers seem to regard ultimate reality as something external to, or distinct from, the visible and tangible natural world. On my view, the natural world is ultimately real (whether or not anything else is), and the ultimate reality with which aesthetic experience connects us may lie entirely within that world. [part of footnote omitted]

[144] Those familiar with my previous writings may find this apple tree example familiar too.

the same tree as imposing rather than simply as beautiful. What the poet finds in the tree depends on the poet's state of mind—but that only shows that the tree has a diversity of aesthetic qualities, and that different qualities are obvious to observers in different mental states.

Aesthetic perceptions of external objects arise from the brain's interpretation of sense data. But the fact that aesthetic knowledge of the apple tree depends on sensations of the apple tree does not contradict the fact that the tree's aesthetic qualities are objectively real and extramental. The perception of the rectangular shape of a wooden door depends on the brain's interpretation of sensory information—but that does not change the fact that the door is rectangular. And the fact that this perception depends on the observer's mental state (if one is drunk enough, one may see the door's sides as a rhombus instead of a rectangle) does not imply that the door is not really rectangular. Similarly, aesthetic perception can reveal real qualities in objects, even though such perception depends on one's mental state and on one's reactions to sensations.

The observer-dependence of aesthetic perception is analogous to what happens when one photographs an apple tree through several different kinds of colored filters. What colors one finds depends on what instrument one uses to photograph them—yet from a physicist's perspective, all the colors of light that one can photograph really are there. The poet's mind may filter out some of the emotional "colors" of experience and record only one set of emotional tones. Another poet may record different tones. Yet all of these feeling "colors," or aesthetic qualities of the tree, are

equally real.[145]

Aesthetic experience is not simply a matter of enjoyment. It also is a *cognitive* process—a process of learning, in which we find facts of a special sort. The poet who sees beauty in an apple tree learns something new about the tree. This learning is not merely metaphorical; the poet actually learns *new facts* about the tree—facts that scientific methods cannot disclose. These are facts about certain aesthetic qualities—tones of feeling, or emotional "colors"—latent in the tree. Which of these tones one finds depends on one's state of mind. But these qualities are not "merely psychological," since they depend on the tree as well as on the poet's state of mind. Perhaps we should think of them as *potentialities* of the tree—products of the tree's ability to produce certain inner changes in an observer. Whatever they are, they are real emergent properties (perhaps relational properties) of the tree.

The idea that aesthetic experience yields knowledge of reality is not a new one. I wish to stress that I did not invent this idea. The view that aesthetic experience yields a special knowledge of reality is deeply embedded in the history of human thought. It crops up at many points in that history, from Platonism to the views of some modern poets, artists and critics.[146] Throughout most of humanity's

[145] Popular wisdom recognizes the parallels between emotional perception and seeing through colored glass. This insight lies behind the familiar expression "looking at the world through rose-colored glasses."

[146] Many of these views are mentioned in the above-cited articles in the *Encyclopedia of Poetry and Poetics* (Preminger, ed.). The relations between Platonism and poetry are described particularly in the

existence, people have recognized that art is a path to truth. Only in recent times have we largely forgotten this fact, under the influence of the so-called scientific worldview. Yet this fact is compatible with all that science has discovered. Nothing in the idea of aesthetic truth can contradict any of the claims of science. Aesthetic knowledge is gained through feeling experience, not through sensation and thought alone. Aesthetic knowledge does not deal with the same facts that concern science; therefore it cannot conflict with scientific knowledge. Yet aesthetic experience explores the inner nature of reality as deeply as do the discoveries of modern physics.

Some philosophers, notably some of the Platonists, have regarded art as affording contact, not only with reality, but also with the ultimate reality behind the visible world.[147] This is more or less the view that I am trying to revive here, but I want to give this view a new twist. Proponents of this view often regard ultimate reality as something purely spiritual—that is, something transcending the visible world, or lying outside the world of things made of material particles. On my view, the world made of matter *is* ultimate reality—or at least is one part or sector of that reality. The ultimate reality with which aesthetic experience connects us lies within the natural world, not outside it. Yet despite this, the realities we discover through aesthetic experience do belong to a different order

articles "Poetry, Theories of" and "Platonism and Poetry."

[147] See the article "Platonism and Poetry" in the *Encyclopedia of Poetry and Poetics* (Preminger, ed.).

[Pages Deleted Here]

[Pages Deleted Here]

Note added later: These are the works cited in the original book. Only some of them are cited in this short book.

Works Cited

The following list gives bibliographic information for all works cited in the book. In some instances, information used in the book was available from many sources; in these cases, the works listed are the ones actually consulted.

(No author listed), "Editorial Commentary", *Behavioral and Brain Sciences*, vol. 13 no. 1 (1990), p. 94.

D.M. Armstrong, *Universals: an Opinionated Introduction* (Boulder, Colorado: Westview Press, 1989).

———————, *A Theory of Universals*, vol. 2 of *Universals and Scientific Realism* by D.M. Armstrong (Cambridge, England: Cambridge University Press, 1978).

Jonathan Barnes, *Aristotle* (Oxford: Oxford University Press, 1982).

Donald L. M. Baxter, "Identity in the Loose and Popular Sense" (*Mind* vol. 97 (1988), pp. 575-582).

Henri Bergson, *Creative Evolution*, trans. Arthur Mitchell (N.Y.: Henry Holt & Co., 1911).

Keith Campbell, *Body and Mind* (2nd ed.: Notre Dame: Univ. of Notre Dame Press, 1984).

Donald A. Cooke, *The Life and Death of Stars* (New York: Crown Publishers, 1985).

James W. Cornman and Keith Lehrer, *Philosophical Problems and Arguments: An Introduction* (2nd ed.: New York, Macmillan, 1974).

V. Csányi, "Are Species Gaia's Thoughts?", *Behavioral and Brain Sciences*, vol. 13 no. 1 (1990), p. 76; references on pp. 105-108.

Daniel C. Dennett, *Consciousness Explained* (Boston: Little, Brown & Co., 1991).

———————, *Content and Consciousness* (London: Routledge & Kegan Paul, 1969).

John C. Eccles, *Evolution of the Brain: Creation of the Self* (London: Routledge, 1989).

Robert Eisberg and Robert Resnick, *Quantum Physics of Atoms, Molecules, Solids, Nuclei and Particles* (N.Y.: John Wiley & Sons, 1974).

Paul Fishbane, "Elementary Particle Physics", in Serway (cited below), pp. 1096-1099.

Kathleen Freeman, trans., *Ancilla to The Pre-Socratic Philosophers* (Oxford: Basil Blackwell, 1956).

Etienne Gilson, *The Christian Philosophy of St. Thomas Aquinas* (London: Victor Gollancz Ltd., 1957).

Charles Hartshorne, *Omnipotence and Other Theological Mistakes* (Albany, N.Y.: State University of New York Press, 1984).

Desmond Paul Henry, *Medieval Mereology* (Amsterdam and Philadelphia: B.R. Grüner, 1991).

Eli Hirsch, *The Concept of Identity* (N.Y.: Oxford University Press, 1982).

David Hume, *A Treatise of Human Nature*, L.A. Selby-Bigge, ed.; revised by P.H. Nidditch (2nd ed.; Oxford: Clarendon Press, 1978).

Frank Jackson, "Epiphenomenal Qualia", *Philosophical Quarterly* vol. 32 no. 127 (1982).

Menas Kafatos and Robert Nadeau, *The Conscious Universe: Part and Whole in Modern Physical Theory* (New York: Springer-Verlag, 1990).

David K. Lewis, *Parts of Classes* (Oxford: Basil Blackwell, 1991).

David K. Lewis, *On the Plurality of Worlds* (Oxford: Basil Blackwell, 1986).

David K. Lewis, *Counterfactuals* (Cambridge: Harvard University Press, 1973).

Lawrence Brian Lombard, *Events: a Metaphysical Study* (London: Routledge & Kegan Paul, 1986).

Michael J. Loux (ed.), *Universals and Particulars: Readings in Ontology* (Notre Dame: University of Notre Dame Press, 1976).

———————, "The Existence of Universals". In Loux (ed.), *Universals and Particulars: Readings in Ontology* (cited above).

Arthur O. Lovejoy, *The Great Chain of Being* (Harvard University Press, 1964).

J. E. Lovelock, *Gaia: a New Look at Life on Earth* (Oxford: Oxford University Press, 1979; reprint ed. 1987).

William A. MacKay, "The Way of All Matter", *Behavioral and Brain Sciences*, vol. 13 no. 1 (1990), pp. 82-83; references on pp. 105-108.

D.H. Mellor, *Real Time* (Cambridge, England: Cambridge University Press, 1981).

Marvin Minsky, *The Society of Mind* (New York: Simon & Schuster, 1988).

Harold. J. Morowitz, "Rediscovering the Mind", in *The Mind's I*, ed. D.R. Hofstadter and D.C. Dennett (N.Y.: Basic Books, 1981). Reprinted from *Psychology Today*, vol. 14 no. 3 (1980), pp. 12-18.

Thomas Nagel, *The View from Nowhere* (N.Y.: Oxford University Press, 1986).

————————, "What Is It Like to Be a Bat?", *Philosophical Review*, vol. 83, no. 4 (October 1974).

Terence Parsons, *Nonexistent Objects* (New Haven: Yale University Press, 1980).

Richard M. Pearl, *Rocks and Minerals* (rev. ed.; New York: Barnes and Noble, 1956).

Plato, *Plato's Republic*, trans. G.M.A. Grube (Indianapolis: Hackett Pub. Co., 1974).

——, *Plato's Timaeus*, trans. Francis M. Cornford (New York: Liberal Arts Press, 1959).

Alex Preminger, ed., *Encyclopedia of Poetry and Poetics*, (Princeton, N.J.: Princeton University Press, 1965).

Willard Van Orman Quine, *Set Theory and Its Logic* (rev. ed.; Cambridge, MA: Belknap Press, 1969).

————————————, *Word and Object* (New York: The Technology Press of The Massachusetts Institute of Technology and John Wiley & Sons, 1960).

Andrew J. Reck, *Speculative Philosophy: A Study of Its Nature, Types and Uses* (Albuquerque: University of New Mexico Press, 1972).

Sir David Ross, *Aristotle* (5th ed., reprint; London, Methuen & Co., 1960).

Rudy Rucker, *Infinity and the Mind: The Science and Philosophy of the Infinite* (Boston: Birkhäuser, 1982).

Bertrand Russell, *The Principles of Mathematics* (paperback ed.; New York: W.W. Norton & Co., 1996).

———————, *Logical Atomism*, in *The Philosophy of Logical Atomism*, ed. David Pears (La Salle, Ill.: Open Court, 1985).

———————, *Human Knowledge, Its Scope and Limits* (New York: Simon and Schuster, 1948).

Jonathan Schull, "Are Species Intelligent?: Not a Yes or No Question", *Behavioral and Brain Sciences*, vol. 13 no. 1 (1990), pp. 94-104; references on pp. 105-108.

———————, "Are Species Intelligent?", *Behavioral and Brain Sciences*, vol. 13 no. 1 (1990), pp. 63-75; references on pp. 105-108.

John R. Searle, *The Rediscovery of the Mind* (Cambridge, Mass.: The MIT Press, 1992).

Raymond A. Serway, *Physics for Scientists and Engineers with Modern Physics* (2nd ed.; Philadelphia: Saunders College Publishing, 1986).

J.C. Smuts, *Holism and Evolution* (N.Y.: The Macmillan Co., 1926).

Baruch Spinoza, *The Ethics*, in *The Chief Works of Benedict de Spinoza*, trans. R.H.M. Elwes (reprint ed.; N.Y.: Dover, 1955).

W.T. Stace, *The Philosophy of Hegel* (New York: Dover, 1955).

Kim Sterelny, "Learning, Selection and Species", *Behavioral and Brain Sciences*, vol. 13 no. 1 (1990), pp. 90-91; references on pp. 105-108.

Matt Visser, *Lorentzian Wormholes: From Einstein to Hawking* (Woodbury, N.Y.: AIP Press, 1995).

John Archibald Wheeler, "Law without law", in *Quantum Theory and Measurement*, ed. J. A. Wheeler and W. H. Zurek (Princeton, N.J.: Princeton Univ. Press, 1983).

Thomas Whittaker, *The Neo-Platonists: A Study in the History of Hellenism* (2nd ed.; Cambridge University Press, 1928).

J.H. Woodger, *The Technique of Theory Construction*, vol. 2 no. 5 in *International Encyclopedia of Unified Science* (Chicago: University of Chicago Press, 1939)

A Final Note

This collection of articles does not finish the project I undertook here. These articles support the conclusions I laid out in the Introduction, but they do not give a complete line of argument. In the Introduction I mentioned some other writings of mine that fill in the gaps. If you haven't already read the Introduction, you might want to do so now. It will give you an idea of the purpose of this collection, and it will tell you where to find those other writings.

If you have comments on this collection, please feel free to contact me. My contact information is in the Introduction.