

Bowdoin College

Bowdoin Digital Commons

Honors Projects

Student Scholarship and Creative Work

2023

Searle's Mind: Brains, Subjects, and Systems

Saul Cuevas-Landeros

Bowdoin College

Follow this and additional works at: <https://digitalcommons.bowdoin.edu/honorsprojects>



Part of the [Philosophy of Mind Commons](#)

Recommended Citation

Cuevas-Landeros, Saul, "Searle's Mind: Brains, Subjects, and Systems" (2023). *Honors Projects*. 411.
<https://digitalcommons.bowdoin.edu/honorsprojects/411>

This Open Access Thesis is brought to you for free and open access by the Student Scholarship and Creative Work at Bowdoin Digital Commons. It has been accepted for inclusion in Honors Projects by an authorized administrator of Bowdoin Digital Commons. For more information, please contact mdoyle@bowdoin.edu, a.sauer@bowdoin.edu.

Searle's Mind:
Brains, Subjects, and Systems

An Honors Paper for the Department of Philosophy

By Saul Cuevas-Landeros

Bowdoin College, 2023

©2023 Saul Cuevas-Landeros

Table of Contents

Acknowledgments	3
Introduction	4
Chapter I	
1. Into the Room.....	13
Chapter II	
1. Searle’s Biological Naturalism.....	27
2. The Underlying Assumption of Conceptual Dualism.....	30
3. Thesis One: Consciousness is Irreducible.....	37
4. Thesis Two: Consciousness is Causally Emergent.....	41
5. Thesis Three: Consciousness is a Feature of the Brain.....	44
6. Thesis Four: Consciousness Functions Causally.....	47
Chapter III	
1. Subjectivity and Other Loose Ends.....	56
2. Searle as a Biological-Property Dualist.....	71
3. What Comes Out of the Room.....	80
Conclusion	86
Bibliography	90

Acknowledgments

Attending college from 2019 to 2023 has been messy, to say the least, most notably because of the COVID-19 pandemic. Amid the uncertainty, I'm happy to say I found a place in the Bowdoin Philosophy Department. I am extremely grateful for both the academic and intellectual growth and support the Philosophy Department has fostered within myself and my peers. I am especially indebted to Professor Schon for properly introducing me to the philosophy of mind and supporting this honors project. I could not have completed this without his continuous support and guidance. To my peers in the department, thank you for many novel conversations both in and out of Pols House. I would also like to thank the Digital and Computational Studies Department for supporting my growth as an academic and providing a space to combine my interests in philosophy and technology.

Of course, none of this could have been possible without the support of my family, especially my mom. Mamá, gracias por todo tu apoyo durante toda mi vida. Nunca has dudado de mí y sigues apoyándome, hasta cuando lo que haga no parece tener mucho sentido. Seguiré haciéndote sentir orgullosa como pueda. Thank you to my cousins for their support and openness to talk about these seemingly esoteric philosophical subjects.

I would not be who I am without my friends and their influence on my thinking. I am especially thankful to Nestor for lifelong talks on philosophy and to Leighton for being my local biology expert. Jose, Coltrane, Sebastian, Gerardo, Jason, Harrison, and Timmy thank you for your overwhelming support and willingness to listen about the sordid problems of philosophy – I hope to continue bothering you all with these issues for many years to come.

Introduction

It has been forty-three years since Searle conceived of the Chinese Room Argument, and still it has managed to maintain its relevance as a classic thought experiment in the philosophy of mind. In this thesis, I will be addressing Searle's Chinese Room Argument and how the thought experiment connects to Searle's overarching project of biological naturalism. I believe that underlying Searle's presentation of the Chinese Room Argument is his eagerness to propose biological naturalism. This thought experiment has influenced our inquiries and concerns within the philosophy of mind by establishing the importance of the 'mental' processes that occur in the mind. The mind is more than behavior, more than inputs and outputs. I believe Searle firmly establishes this point among the confusion of his biological naturalism, but he is still left with some unsettling and contradictory truths that stem from his use of the Chinese Room. It is best to historicize the issue and further illustrate its general influence.

In 1950, Alan M. Turing famously "proposed to consider the question, 'Can machines think?'"¹ Turing believed the question was too ambiguous to properly answer, and instead posed a series of sub-questions that are meant to act as substitutes for the original question. These sub-questions come from his variation of an 'imitation game.' The original imitation game considers three agents: The Man (who is a biological man), The Woman (who is a biological woman), and The Interrogator (who can be anyone). In this game, The Interrogator is in a separate room from The Man and The Woman. The Interrogator's goal is to figure out who The Man is and who The Woman is by posing various questions that should help them guess correctly. To assist The Interrogator, the question-answer exchange is text-based. But there is another challenge: The Man is meant to deceive The Interrogator into believing he is The Woman, while The Woman

¹ Alan M. Turing, "Computing Machinery and Intelligence." *Mind* 59, no. 236 (1950): 433.

can help The Interrogator correctly guess who is who. Turing then inserts a computational spin: “We now ask the question, ‘What will happen when a machine takes the part of [The Man] in this game?’ Will [The Interrogator] decide wrongly as often when the game is played like this as [they do] when the game is played between a man and a woman? These questions replace our original, ‘Can machines think?’”²

Turing’s variation of the Imitation Game has become a keystone thought experiment for philosophers, computer scientists, and even cognitive scientists. Although it was once a useful heuristic for our scientific or psychological studies, it is not such a straightforward thought experiment in the philosophy of mind. Perhaps the question, ‘Can machines think?’ is ambiguous and difficult to pin down for good reason – it is not an easy task to define ‘thinking.’ Philosophers have been looking into the nature of knowledge since philosophy’s conception and are still arguing what it even means for a person to think, so how would this deep reduction of such a Herculean question solve our dilemmas on if machines can think? I can appreciate the nature of trying to break a seemingly unanswerable question (‘Can machines think?’) down into more ‘solvable’ questions; however, I do not think leaving the conversation there and acting as if the two sets of questions respond to the same problem is enough. I do not believe this was Turing’s intention either. But from this starting point is where Searle provides what I deem to be more relevant concerns on our cognitive faculties as can be seen by the thought experiment’s longevity and influence.

The Chinese Room Argument goes as follows: Imagine you are an English speaker locked inside a room. You know no Chinese; you cannot distinguish it from other languages you do not know or from other meaningless symbols. Inside the locked room with you, there are

² Turing, “Computing Machinery and Intelligence,” 434

many filing cabinets full of papers that have symbols on them. Along with the filing cabinets and their contents, there is a rule book written in English inside the room. The rule book provides detailed instructions for formally manipulating the symbols or correlating them to other symbols. On the outside of the room there is someone passing strings of unknown symbols to you while you are inside it. You do not know it, but the symbols being passed into the room are questions written in Chinese. Now imagine that, by following the rule book, you can pass out a different set of symbols than those given to you to the person on the outside. Unknown to you, what you passed out were responses to the questions written in Chinese. Although you did not know the meaning to what you produced you are still providing a correct response to the question you were given; you produced a correct enough response to fool a native Chinese speaker. You produced ‘correct’ behaviors or responses to the questions you were given, but the mental state or intentionality – understanding – was left out of the room.

Throughout this project, I will ‘step into the Chinese Room’ and develop the areas where the Chinese Room Argument succeeds. My aim is to pick out what Searle has done well with the Chinese Room Argument and introduce how it fits in with his school of biological naturalism. In Searle’s original presentation of the Argument, he clearly had some kind of conception of his then upcoming biological naturalism. Throughout the process of analyzing what Searle’s Chinese Room Argument has done right, more of Searle’s theory of intentionality will be revealed since he relies on it as a condition of understanding. From here, I will introduce some of the primary arguments used against the Chinese Room Argument because they do not fit with Searle’s overarching theme of biological naturalism. Particularly, Searle’s conception of systems and system features is something he endorses for the biological but immediately labels as silly for the Chinese Room. Following the exposition of systems and system features, I will expand on how

there is a disconnect between Searle's use of system features and his view of the Chinese Room Argument. What is so special about Searle's conception of systems and the systems present in the Chinese Room Argument? Searle should just claim that the Chinese Room is simply not the kind of thing that can think.

While Searle's position on the Chinese Room can sometimes look like a functionalist theory since he focuses on causality in the processes that mold the mind, he does not want to identify himself with this school or any other. They are all stuck on a broken conception of the mind, what he calls conceptual dualism that stems from Descartes. John Searle consistently attempts to escape the grips of Descartes and various forms of dualism throughout his writings on the philosophy of mind. Yet, Searle's philosophy constantly reminds me of Descartes, his method of doubt, and dualism in general. Take these introductory quotes from Descartes and Searle, respectively, for example:

Several years have now passed since I first realized how numerous were the false opinions that in my youth I had taken to be true, and thus how doubtful were all those that I had subsequently built upon them. And thus I realized that once in my life I had to raze everything to the ground and begin again from the original foundations, if I wanted to establish anything firm and lasting in the sciences.³

The different positions then are all taken within a set of mistaken assumptions. The result is that the philosophy of mind is unique among contemporary philosophical subjects, in that all of the most famous and influential theories are false... To make the whole subject even more poignant, many of these theories, especially dualism and materialism, are trying to say something true. One of my aims is to try to rescue the truth from the overwhelming urge to falsehood.⁴

Descartes opens his *Meditations* with a radical method of doubt: doubt in his perceptions and experiences, in the views of modern science, and in the grounding of his religious beliefs.

³ René Descartes, Donald A. Cress, and René Descartes. *Discourse on Method ; and: Meditations on First Philosophy*. 4th ed. (Indianapolis: Hackett Pub, 1998), 59

⁴ John R. Searle, *Mind: A Brief Introduction*. Fundamentals of Philosophy Series. (Oxford; New York: Oxford University Press, 2004), 2

Searle's introduction to the contemporary philosophy of mind starts with a radical doubt of the prominent schools of thought – any kind of '-ism' is likely to have erred in its account of the mind. Both want to work with the obvious truths of their times.

But Descartes concludes from his radical skepticism that the mind and body exist as two different kinds of substances and then runs into issues as to how it is possible for these substances to interact or be located in space. Searle's radical doubt of the contemporary theories of the mind, on the other hand, lands him in a material monist camp that attempts to make room for our subjective, qualitative, first-person experiences, seemingly still running into many of the same problems Descartes faced. The explanatory gap – or how it is possible for the physical to cause or work with our consciousness – runs deep in both philosophers' work. Searle believes he solves this explanatory gap, but he usually seems to simply restate it with an emphasis on biology. Further pressing the similarities and differences between Searle and Descartes, Searle places more faith in science than Descartes, whereas Descartes puts too much faith into his own religious practice. Searle also seems to take his own perceptions of the world as more 'truth-telling,' often taking them for granted more than Descartes does. Searle takes some things to just be obviously true. And these truths should not be denied.

Despite Searle's distaste for '-isms,' Searle advocates for one of his own: biological naturalism. Biological naturalism attempts to place what we would typically think of as the 'mental' into material, more specifically, biological terms. Biological naturalism is a naturalism because it claims that the mind is just another part of the natural world, and it is biological because the correct level to look for the mind is the biological level. Our mind – with all the typical 'mental' features of consciousness, intentionality, the qualitative character of experience, etc. – is a real phenomenon and can be explained at the level of biology, not at the level of

particles, molecules, or a spiritual plane. Searle's position confidently asserts that the 'mental' is grounded in the biological. In the brain, Searle trusts. Although Searle's philosophy of mind faces a slew of problems, I believe his use of contemporary empirical sciences to guide it is helpful and informative. His attempt to 'rescue the truth from the overwhelming urge to falsehood' does not wholly succeed, but it has provided a necessary push in the field of philosophy of mind to make room for what we obviously believe exists: our consciousness and the subjective character of experience. These ideas begin to show themselves in the Chinese Room Argument: Searle's original abstract emphasizes the brain's biological powers and his concerns on proving that other people have minds is biologically backed.

How did Searle arrive to his position of biological naturalism and to conceding so much philosophical power to the brain? His work in the philosophy of mind largely stems from the rise of cognitive science and takes root in his interest in the philosophy of language. In Searle's view, "the philosophy of language is a branch of the philosophy of mind; therefore no theory of language is complete without an account of the relations between mind and language and of how meaning... is grounded in the more biologically basic intrinsic intentionality of the mind/brain."⁵ It is both tempting and simple to group Searle in some existing school of the mind, such as functionalism or property dualism, but Searle always points back to the brain as the foundation of his philosophy of mind. Searle's trust in the brain emphasizes an overarching move in neuroscience, cognitive science, artificial intelligence, and philosophy to take the brain for granted in the equation of what makes a mind. The brain takes a central position in many philosophical stances that accommodate for scientific worldviews, and Searle serves as a landmark in this route.

⁵ John R. Searle, *The Rediscovery of the Mind*. Representation and Mind. (Cambridge, Mass: MIT Press, 1992), xi

A large part of Searle's influence traces back to his Chinese Room Argument. The Chinese Room Argument challenged many prevalent schools of thought in philosophy of mind and artificial intelligence such as cognitivism, connectionism, and, most explicitly, the computational theory of the mind. Many contemporary philosophers may find the thought experiment to be thought-provoking, silly, challenging, and/or, sometimes, question-begging, but no matter their position, it is admittedly an influential and unavoidable thought experiment in the philosophy of mind and many other fields such as cognitive science and artificial intelligence. The Chinese Room Argument brings into question the role of the body, symbols and symbol manipulation, representations, other minds, and more in our conscious states and cognitive faculties such as 'understanding.' Additionally, Searle expands on some of the more general implications of the Chinese Room Argument in further writings, which fit into his overall framework of the mind.

In the first chapter, I will be delving into the Chinese Room Argument, presenting it charitably and arguing for the areas in which it has succeeded. Along the way, I will identify the features John Searle believes are missing from the Chinese Room to be able to attribute mentality to it. I will then present some of the original replies to the Argument and how Searle responds to them. Most notably, I will present the Systems Reply. In brief, the Systems Reply posits that although the individual inside the Chinese Room may not understand Chinese, the entire *system* taken together – the person, the rulebook, and the database – does exhibit understanding. Searle dismisses the case, saying it is embarrassing to respond to it, but, with the additional context of biological naturalism, it starts to look like more reasonable of a stance for Searle's position. Searle employs systems and system features in his own explanations, but instead of using a room-system, the system is the brain and it exhibits its own kind of system

features such as consciousness and intentionality. I want to do as best as I could to analyze what Searle believes makes the brain so special.

An analysis of Searle's Chinese Room serves well as an application of some of Searle's theories of the mind since the thought experiment must obviously fit within his own framework or else Searle would be presenting self-contradictory views. Searle's biological naturalism rests on the conjunction of four theses that I will analyze individually. These theses can give hints of mind-brain identity theories, supervenience, and even dualism, but Searle denies all these other positions. These schools of mind, claims Searle, operate on broken and unnecessary assumptions of how the world works. Just because we have mental activity, mental states, or conscious life, does not mean that we need to posit some other entities to solve our problem. We simply know our mental life is happening and now we just need to fit it into our sciences. The question of if our subjective, qualitative experiences exist is not a good one. We simply need to accept these as true and instead figure out how they fit into our lives.

After an analysis of Searle's biological naturalism throughout the second chapter, I will go into some of the loose ends of Searle's philosophy in the third chapter. Most notable is Searle's conception of subjectivity that will be touched on in when analyzing Searle's theses. We typically think of subjectivity as having to do with our personal tastes, personal descriptions, opinions, etc. Searle's use of subjectivity means that a subject is experiencing something, that an experience is coming from a point of view. Although I do not want to go off the rails with some of the implications of Searle's position on subjectivity, I believe this is the part of biological naturalism that makes it so difficult to see how Searle could be a substance monist and not some kind of dualist. From there stems much of our confusion on Searle and his biological naturalism. Lastly, I will tie some of these ideas back into the Chinese Room, using Searle's biological

naturalism to inform our conception of the Room, instead of having the Room inform where Searle was going with biological naturalism. I will delve into an analysis of Searle as a biological-property dualist and if the charge holds up. I claim that, overall, Searle's philosophy of mind is a variation of a Cartesian dualism because of his conception of subjectivity and the brain's causal powers. I will then expand on some of the merits of the Systems Reply, using some of Searle's own strategies to criticize the Chinese Room Argument. Because of Searle's notion of systems and system features, his philosophy of mind seems to be incohesive, or the Chinese Room Argument simply fails. Searle's special treatment of the brain's causal powers starts to seem like an extra piece in a completed jigsaw puzzle. Ultimately, Searle wants to concede his knowledge to the sciences, but it leaves him with mystical claims. Let's start attempting to understand Searle by entering the Chinese Room.

Chapter I

§1. Into the Room

Following Turing's line of thought regarding thinking computers, Searle formulates his Chinese Room Argument. He first presents the Chinese Room in his 1980 paper, "Minds, Brains, and Programs." He continues to present the argument in his 1984 Reith Lectures and in a 1999 summary for The MIT Encyclopedia for the Cognitive Sciences. In 2010 Searle describes the Chinese Room Argument's conclusion in terms of consciousness and intentionality. But ever since his original 1980 paper was published, the Chinese Room Argument and its implications have been contentious. The "Minds, brains, and programs" article even includes six preemptive replies to objections foreseen by Searle and twenty-six original responses from various other philosophers, neuroscientists, and cognitive scientists to which Searle also replies. The constant reappearance of the thought experiment emphasizes its contentious nature. Whether all the discussion on it has been fruitful is an issue of its own. I would argue the Chinese Room Argument has offered a variety of challenges to many of our classical philosophical notions. If not directly, Searle and the Chinese Room have at least influenced the thought behind approaching these issues.

When first presented in 1980, Searle posed the Chinese Room thought experiment to refute the position of Strong Artificial Intelligence (Strong AI). The proponent of Strong AI defends the position that "the appropriately programmed computer literally has cognitive states and that the programs thereby explain human cognition."⁶ Strong AI is contrasted with Weak Artificial Intelligence (Weak AI), which proposes that the computer and computer programs are

⁶ John R. Searle, "Minds, Brains, and Programs." Behavioral and Brain Sciences 3, no. 3 (September 1980): 417

nothing more than good tools for studying the mind. For example, we may be able to test and examine some of our hypotheses and predictions on how the brain works more rigorously. Searle has no problem with Weak AI and targets Strong AI with the Chinese Room Argument. Searle even agrees that Weak AI is and will be a useful tool for simulating various processes, even the processes of the brain. We can see such computational simulations through the use of contemporary data structures in computer science such as neural networks and variations thereof. But these simulations should not be confused for duplications of neural events and brain causality. The results of the Chinese Room Argument can also be made out to argue against the more prevalent schools of the computational theory of mind (CTM) or, even more broadly, functionalism, but they were not Searle's explicit focus as originally presented.

Let me lay out the explicit points Searle intends to make with the Chinese Room Argument: The person inside the room has no understanding of Chinese. This person in the room lacks *intentionality*, as related to the Chinese symbols, so they cannot understand it. They only have *formal* features of the Chinese symbols. Searle's definition of intentionality is "that property of many mental states and events by which they are directed at or about or of objects and states of affairs in the world,"⁷ while "all that 'formal' means here is that I can identify the symbols entirely by their shapes."⁸ The formal features correlate with syntax, but what is lacking is *semantics*, or meaning. The person inside the Chinese Room can produce the right syntax, but they (or it) lack any semantics. The only semantics given to the symbols are those assigned by the Chinese speaker outside of the room.

Searle believes that although the correct output is being produced by the room, there is no understanding of Chinese going on or being had by the room or anyone in it. The person in the

⁷ John R. Searle, *Intentionality: An Essay in the Philosophy of Mind*. 1st ed. (Cambridge University Press, 1983), 1

⁸ Searle, "Minds, Brains, and Programs," 418

room does not have any representation, or intentionality, behind the production of their Chinese responses, Searle claims. They exhibit correct behavior, yes, but there are no mental aspects as we typically conceive of them. A computer program is simply being instantiated by formally manipulating various symbols. In contrast, proponents of Strong AI would claim:

1. that the [English speaker] can literally be said to *understand* the [Chinese], and provide answers to the questions, and
2. that what the [English speaker and their rule book] do *explains* the human ability to understand the [Chinese] and answer questions about it.⁹

Completing the analogy of the Chinese Room with computation, supporters of Strong AI claim that the English speaker is just a computer, and the rule book is a program. For supporters of Strong AI, the program's instantiation is enough for understanding. Proponents of Strong AI may also believe that Searle's picture of the Chinese Room and understanding is rather simple. Some may claim that a computer program may operate similarly, but the level of complexity of the program will contribute to its having mentality such as understanding. Either way, Searle believes such a claim is misguided because it does not add much in terms of meaning and representation. Searle believes such a claim confuses syntax and semantics – it confuses the formal features of a symbol with its meaning. For instance, if you only speak English, you may be able to describe the word 'quincena' in terms of its formal features; you are likely to describe the individual letters in the word and their relation to each other but would not be able to describe its meaning in Spanish, the language that the word is from. In other words, you may be able to describe that the word has eight letters or shapes, and the first one looks like a 'q' and is next to what looks like a 'u' and so on and so forth, but from this form alone would not be able to ascertain that the word 'quincena' means 'a period of fifteen days.' You are likely to

⁹ Searle, "Minds, Brains, and Programs," 417

understand the word's syntax, but not its semantics. Somehow, for the Strong AI supporter, the syntax gets you to the semantics.

Strong AI's claim may seem to be answering a different question than, "Could machines think?" And that's because it *is* a different question or at least a slightly different configuration. Searle clarifies that the claim he promotes against Strong AI is more pointed toward the question, "Could a digital computer think?" He distinguishes between a 'digital computer' and a 'machine' because he believes that "only a machine could think, and only very special kinds of machines, namely brains and machines with internal causal powers equivalent to those of brains."¹⁰ If it is not yet clear, the underlying intention Searle has in his position for the Chinese Room Argument is to set up his own theory of the mind, biological naturalism, as a (or *the*) solution to the nagging mind-body problem.

Extending some of the Chinese Room Argument's claims and implications against Strong AI to other philosophical schools like functionalism can be even more contentious.

Functionalism claims that what makes something a mental state is the functional role it plays. Instead of looking for an ontological definition of what makes up a mind, the functionalist takes a causal approach by defining the mind by its function. The functionalist claim proposes that a combination of certain sensory experiences, mental states, and behaviors cause some other specific mental state – the correct combination of inputs (i.e., mental states, behaviors) and outputs (i.e., mental state) constitute the mind. So, if the claim of Strong AI is true and "the mind is to the brain, as the program is to the computer hardware," then some correct combination of inputs produces a certain output that can be identified with a mind and at least a variation of functionalism is true. Considering the positions of functionalism and Strong AI, it becomes clear

¹⁰ Searle, "Minds, Brains, and Programs," 417

how a functionalist, much like the proponent of strong AI, may argue that there is understanding going on in the Chinese Room; the understanding of the Chinese symbols just is taking them in and outputting another set of Chinese symbols that a program describes.

Searle starts his Chinese-Room-conceiving article “Minds, brains, and programs,” with the following start to his abstract:

This article can be viewed as an attempt to explore the consequences of two propositions. (1) Intentionality in human beings (and animals) is a product of causal features of the brain. I assume this is an empirical fact about the causal relations between mental processes and brains. It says simply that certain brain processes are sufficient for intentionality. (2) Instantiating a computer program is never by itself a sufficient condition of intentionality. The main argument of this paper is directed at establishing this claim. The form of this argument is to show how a human agent could instantiate the program and still not have relevant intentionality.¹¹

It is important to keep Searle’s initial desires with this article and thought experiment in the back of our minds while panning for gold through his work. His intentions with the Chinese Room Argument can sometimes get lost if we strongly feel one way or another on artificial intelligence. AI Optimists may be keen to point out all the imitations of mentality we have witnessed and are experiencing thus far, while AI Pessimists may blindly stick to their skepticism and declare that nothing inorganic can exhibit mentality. Keeping Searle’s intentions with the Chinese Room Argument clear from the beginning helps bring out what we can learn from it despite our intuitions one way or the other.

Proposition (1) seems like something Searle takes as a matter of fact. It is an assumed premise for him that can be reasoned out from our neurobiology and our experience. We human beings have the ability to represent the world, make predictions about it and in it, hold beliefs, desires, and other mental states, and this can all be attributed to our brain’s ability to produce and sustain conscious states and intentionality. Again, intentionality is “that property of many mental

¹¹ Searle, “Minds, Brains, and Programs,” 417

states and events by which they are directed at or about or of objects and states of affairs in the world.”¹² Intentions are caused and realized by brain activity. It is not as if intentionality is some extra property of the brain since it is just a consequence of neurobiological causal processes. Searle believes intentionality is a key feature of consciousness and that it has a biological reason for existing. It has its own logical properties, but our conscious states usually have intentionality because they are *about* something. Since every other being we intuitively attribute intentionality to is biological, Searle assumes the first proposition’s credibility. In short, Proposition (1) asserts that for biological beings everything that is about something is a result of biological processes. The Proposition does not rule out the possibility of other beings, artifacts, or machines having intentionality without biology, but it does assert that that is how it works for *us*.

Proposition (1) is an assumption that Searle goes into the Chinese Room with, so to speak. What the Chinese Room Argument is really meant to establish is his second proposition: a computer program alone is not enough to have ‘aboutness,’ or intentionality. Searle is not attempting to establish that a computer program *and* a body or anything else are not enough for intentionality. With the Chinese Room Argument, he is focused solely on the role of the computer program in mental states. Searle makes this position clear in the Robot Reply to the Chinese Room Argument, which I will touch on. But for now, one just needs to know that Searle is inferring Proposition (2) through the Chinese Room Argument.

Once he has gone through and supported the Chinese Room Argument, Searle believes that Propositions (1) and (2) will have some simple consequences:

These two propositions have the following consequences: (3) The explanation of how the brain produces intentionality cannot be that it does it by instantiating a computer program. This is a strict logical consequence of (1) and (2). (4) Any mechanism capable of producing intentionality must have causal powers equal to those of the brain. This is meant to be a trivial consequence of 1. (5) Any attempt literally to create intentionality

¹² Searle, *Intentionality*, 1

artificially (strong AI) could not succeed just by designing programs but would have to duplicate the causal powers of the human brain. This follows from 2 and 4.¹³

Now equipped with Searle's overarching goal for the "Minds, brains, and programs" article, his argument can be laid out in a logical structure:

- (1) Intentionality in human beings (and animals) is a product of causal features of the brain. [Premise]
- (2) Instantiating a computer program is never by itself a sufficient condition of intentionality. [Premise – from Chinese Room Argument]
- (3) The explanation of how the brain produces intentionality cannot be that it does it by instantiating a computer program. [1, 2]
- (4) Any mechanism capable of producing intentionality must have causal powers equal to those of the brain. [1]
- (5) Any attempt literally to create intentionality artificially (strong AI) could not succeed just by designing programs but would have to duplicate the causal powers of the human brain. [2, 4]

Given that intentionality is a causal feature we get from our brain's biology (Proposition 1) and that a computer program is not sufficient for intentionality (Proposition 2), then the brain's biological ability to produce and sustain our conscious states must *not* be something that comes from *just* a computer program. To repeat, Searle is not here making the stronger claim that producing intentionality and conscious states is only something that organic beings can do. Searle is just asserting that in the world as we know it, we human beings (and other animals) *do* have intentionality, presumably because of our biology. It could have ended up some other way, but from what we know, biology is what causes conscious states. This is Proposition (1).

The Chinese Room Argument aims to establish Proposition (2). I would be remiss to not pose the Chinese Room in its traditional garbs. Here is the argument as originally presented:

Suppose that I'm locked in a room and given a large batch of Chinese writing. Suppose furthermore (as is indeed the case) that I know no Chinese, either written or spoken, and

¹³ Searle, "Minds, Brains, and Programs," 417

that I'm not even confident that I could recognize Chinese writing as Chinese writing distinct from, say, Japanese writing or meaningless squiggles. To me, Chinese writing is just so many meaningless squiggles. Now suppose further that after the first batch of Chinese writing I am given a second batch of Chinese script together with a set of rules for correlating the second batch with the first batch. The rules are in English, and I understand these rules as well as any other native speaker of English. They enable me to correlate one set of formal symbols with another set of formal symbols, and all that "formal" means here is that I can identify the symbols entirely by their shapes. Now suppose also that I am given a third batch of Chinese symbols together with some instructions, again in English, that enable me to correlate elements of this third batch with the first two batches, and these rules instruct me how to give back certain Chinese symbols with certain sorts of shapes in response to certain sorts of shapes given me in the third batch. Unknown to me, the people who are giving me all of these symbols call the first batch "a script," they call the second batch a "story," and they call the third batch "questions." Furthermore, they call the symbols I give them back in response to the third batch "answers to the questions," and the set of rules in English that they gave me, they call "the program." Now just to complicate the story a little, imagine that these people also give me stories in English, which I understand, and they then ask me questions in English about these stories, and I give them back answers in English. Suppose also that after a while I get so good at following the instructions for manipulating the Chinese symbols and the programmers get so good at writing the programs that from the external point of view – that is, from the point of view of somebody outside the room in which I am locked – my answers to the questions are absolutely indistinguishable from those of native Chinese speakers. Nobody just looking at my answers can tell that I don't speak a word of Chinese. Let us also suppose that my answers to the English questions are, as they no doubt would be, indistinguishable from those of other native English speakers, for the simple reason that I am a native English speaker. From the external point of view – from the point of view of someone reading my "answers" – the answers to the Chinese questions and the English questions are equally good. But in the Chinese case, unlike the English case, I produce the answers by manipulating uninterpreted formal symbols. As far as the Chinese is concerned, I simply behave like a computer; I perform computational operations on formally specified elements. For the purposes of the Chinese, I am simply an instantiation of the computer program.¹⁴

Even with all his presuppositions, Searle's thought experiment generally does well at establishing his claim that the instantiation of a good enough computer program is not the same as exhibiting mental states such as understanding. It seems that the primary project succeeds and the simple summary can be put in two phrases: [1] syntax is not semantics, and [2] simulation is not duplication. Although it *prima facie* succeeds, I will show how Searle's conception of

¹⁴ Searle, "Minds, Brains, and Programs," 417-418

systems and system features leaves Searle needing to admit that either the Chinese Room Argument is incoherent because he does not properly address the Systems Reply or his biological naturalism is incoherent because his use of system features is not actually what matters for mental states.

If the Chinese Room Argument succeeds in supporting Proposition (2), then it puts forth plethora of subsequent claims. For instance, the correct combination of inputs and outputs in a computation does not seem to be sufficient for mental states. Computation only works on syntax, or formal symbols, but it does not follow that from just the syntax we derive semantics, meaning, representations, or in Searle's terms, intentionality. The person in the Chinese Room exhibits no understanding of Chinese, and analogously, a Chinese translation computer program exhibits no understanding of Chinese when it translates. The person in the room and the Chinese translation computer program are performing a function – holding a conversation – but they do not *understand* what it is they are doing. The computation of translating is simply a simulation of translation or Chinese interpretation, but it is not the same thing as duplicating the understanding of Chinese. The person in the room and the program are simply symbol manipulators operating on syntax with no sense of semantics, or meaning. The man in the room is operating on meaningless Chinese characters, while the program is working with binary operations – it's all 0s and 1s.

In Proposition (2), Searle states that the person in the room implementing the Chinese program is not exhibiting 'relevant' intentionality. Here he is implying that the notion of *as-if* intentionality is not sufficient. When we speak of a state of the world, we sometimes attribute various kinds of intentionality to things without any *intrinsic* intentions. For example, we may say a plant is thirsty. We do not *literally* mean that the plant has the desire for water; we are

simply acting as if the plant was thirsty in the same way we are. It is a practical use of intentionality, as it may help us better understand a phenomenon, such as a plant's biological need for water, but it is just a metaphorical description. In the same way, we may say that a computer program *understands* what it is doing. Such a description may work for as-if intentionality, but it is not a literal description. The computer program lacks any sort of *intrinsic* intentionality. These distinctions serve to push the point of Proposition (3) that a computer program instantiation explains the production of intentionality. It is a simple consequence of Propositions (1) and (2), Searle asserts – no brain-like causality, no intentions. Furthermore, the Chinese Room is meant to act as a simulation of conversational responses. But it does not follow that from a simulation you derive all the features of the event itself such as understanding the language being translated. For example, meteorologists often must simulate weather patterns in their line of work. From simulating rain movement, it is not like it will suddenly start raining in the meteorologist's office. At the time of writing, OpenAI – an artificial intelligence research and deployment company – has released a popular AI model that simulates conversations. The model, called ChatGPT, seems to simulate the function of intelligent conversation and the mental state of understanding, but the simulation itself is not sufficient for understanding the conversation. Given the way ChatGPT is programmed, it seems to simulate understanding but we can be pretty sure it has no understanding at all.

Up to Proposition (3), Searle's argument seems to take a reasonable shape. Nothing seems controversial unless we find it implausible that the brain causes intentionality, but I am inclined to agree. His first two Propositions are agreeable if the Chinese Room Argument succeeds:

- (1) Intentionality in human beings (and animals) is a product of causal features of the brain.
[Premise]

- (2) Instantiating a computer program is never by itself a sufficient condition of intentionality.
 [Premise – from Chinese Room Argument]

At this point, Searle does not pose his view as the brain being the *only* kind of thing that causes intentionality. But that is the direction he is heading after Proposition (3) and it is why I believe Propositions (4) and (5) have grander consequences. Proposition (4) is meant to be a “trivial consequence” of Proposition (1), but it could also be an empirically incorrect claim. For now, it is just another one of Searle’s assumptions that is meant to be based on contemporary scientific advancements. Arguably, Propositions (1) and (4) could be substituted with various other biological components that are involved in our conscious states, but they would all drive the underlying point Searle is putting forth: our biology matters for causing our intentionality, and thus, consciousness because Searle believes they are codependent. For instance, instead of the brain, Searle could have posited the brain and nervous system or the brain and body, but it would all need to insinuate our biology in the equation. No matter what fills in the role, though, it is something empirically verifiable.

From Propositions (2) and (4), Searle then deduces Proposition (5) that creating intentionality requires duplicating the causal powers of the brain. Merely simulating the brain’s activity is not enough for conscious states because it is the brain’s ‘causal powers’ that produce consciousness. For Searle, the mind does not rely on some ontology like a Cartesian substance because it comes from causal means. This does not seem too far from the functionalist claim that what it is for something to have a mind is to exhibit the right function as opposed to being made of a certain kind of thing. For instance, a diamond is made from a very particular kind of stuff. There is no diamond-like function. To be a diamond is simply to be carbon in a certain crystal structure. In contrast, a key only has to have a certain function. At college, my room key is a card

I swipe, while at home I have a traditional metal key, and furthermore, my laptop requires a combination of alphanumerical characters as a key to unlock it. The key is simply a functional concept. To be a key is to perform a certain function – for functionalism, being is doing.

Despite Searles's continuous emphasis on the role of the brain's causal powers in creating and sustaining mentality, he does not want to identify himself with functionalism and its claims. Searle wants all the working parts of every existing philosophy of mind but does not want to ascribe to any school that is not his own. That is why Proposition (5) stands out as the most interesting of all his Propositions. It closely resembles a functionalist or identity theory of the brain, yet Searle renounces the labels. Although any attempt to create intentionality would need to duplicate the brain's causal powers, programs will not be sufficient for duplicating these causal powers. Searle's position seems to assert that the brain is the only thing capable of having the brain's causal powers. I believe this is the actual claim Searle should be making, but his longing to fit within the progress of AI makes him take an odd position. Searle's position may seem to exclude beings like Martians from having minds of their own. But Searle would likely point to his own intuitions as enough reasoning for saying they have mental states, likely claiming that they act closely enough to human beings or other animals like dogs and snakes, so may have some degree of consciousness being caused. He wants to leave room for whatever he deems could be empirically possible.

What the Chinese Room Argument does firmly establish is that behaviorism is not a correct approach to the philosophy of mind. As a descendant of Turing's Imitation Game, the Chinese Room serves well as a follow-up exercise to challenging the behavioristic conception of the mind and emphasizing what we should care about regarding mental states. Obviously, it seems silly to say the person in the room is understanding Chinese since they do not know any

Chinese at all. The person following the rule book, or ‘instantiating the computer program,’ just sees meaningless squiggles. But this original presentation of the Chinese Room Argument can also be said to be a weakness because it could be question-begging.

As Searle presents the Chinese Room Argument, he makes it seem like it is an impossibility for the Chinese Room to exhibit or produce any kind of understanding of Chinese, or more broadly, to produce intrinsic intentionality. At best, the room derives intentionality from the Chinese speaker’s meaning or from our third-person, as-if attributions. As the Chinese Room Argument is presented, I believe it is impossible for the person in the room to pick up any meaning from the Chinese they are translating. All that the rule book or program is telling the person in the room is something like “move symbol x here and erase symbol y but replace it with symbol z .” Even if the person in the room was able to get extremely efficient at the process, they are not able to pick up meaning from the relationship because they are still just operating on the formal symbols, or syntax.

The Systems Reply responds by saying that focusing on the person within the room is the wrong level for mentality. Instead of focusing on the person, focus on the mental states of the whole *system*. Searle presents the Systems Reply as follows:

While it is true that the individual person who is locked in the room does not understand the story, the fact is he is merely part of a whole system, and the system does understand the story. The person has a large ledger in front of him in which are written the rules, he has a lot of scratch paper and pencils for doing calculations, he has ‘data banks’ of sets of Chinese symbols. Now, understanding is not being ascribed to the mere individual; rather it is being ascribed to this whole system of which he is a part.¹⁵

I must admit that the Systems Reply seems silly. Searle certainly believes that such a response is silly and even embarrassing to respond to but he does so out of the belief that fanatics of Strong AI will be drawn to it. Yet after reading through Searle’s philosophy of mind, I am not sure it is

¹⁵ Searle, “Minds, Brains, and Programs,” 419

at silly as Searle sets it out to be. Searle refers to systems and features of systems that are not features of the system's parts. For instance, Searle might say that the brain understands, while individual neurons do not. The system of the brain has the feature of intentionality, specifically understanding *of* something, while the system's parts, neurons, do not exhibit understanding. So why is the Chinese Room as a system not capable of producing some causally emergent conscious or intentional state such as understanding? Despite its *prima facie* absurdness as a response, the Systems Reply seems to be the most pressing issue for Searle's stance on the mind. It is ironic how Searle's reliance on putting off the Systems Reply becomes a key component to his own philosophy of mind, but it all goes to show how the seemingly simple and taken-for-granted mental features of our lives are not so simple. The mind-body problem – or how and why our mental and physical activity connect – continues to take a firm position in the philosophy of mind and Searle's biological naturalism is an attempt to make consciousness something simpler than we make it out to be.

Chapter II

§1. Searle's Biological Naturalism

The mind-body problem has been with philosophy of mind since the field's beginnings. It calls attention to how it is that our mind and body could possibly be connected and give rise to our rich mental activity. And Searle adamantly believes he has the solution to the tyrannical mind-body problem: "biological naturalism." The basic view defends the position that "mental phenomena are caused by neurophysiological processes in the brain and are themselves features of the brain."¹⁶ For Searle, our mental phenomena are as biologically real as digestion. The simple version of the thesis already seems puzzling, insofar as it suggests that mental phenomena *causally emerge* from the brain while also simply being *features of* the brain. Yet Searle does not want to do a third-person ontological reduction of consciousness because that would leave out our first-person experiences and their qualitative character. Although it is the 'basic view,' it holds a lot of content worthy of analysis within it. The summary statement of biological naturalism already seems to have a smell of reduction of the mental to the physical, but somehow, Searle wants to distinguish the first-person ontology of our lived experiences from the underlying biological, third-person ontology that sustains it.

Searle proposes four main theses as definitive of biological naturalism. In contrast to Descartes's doubt, the brain's power seems so obvious to Searle, and he thinks it should be obvious to us too. The theses, when taken together, are meant to solve the mind-body problem and address some of the immediate issues we may have with his summarized thesis stated earlier

¹⁶ Searle, *Rediscovery*, 1

that “mental phenomena are caused by neurophysiological processes in the brain and are themselves features of the brain.” Word for word, the theses of biological naturalism are:

1. Conscious states, with their subjective, first-person ontology, are real phenomena in the real world. We cannot do an eliminative reduction of consciousness, showing that it is just an illusion. Nor can we reduce consciousness to its neurobiological basis, because such a third-person reduction would leave out the first-person ontology of consciousness.
2. Conscious states are entirely caused by lower level neurobiological processes in the brain. Conscious states are thus *causally reducible* to neurobiological processes. They have absolutely no life of their own, independent of the neurobiology. Causally speaking, they are not something “over and above” neurobiological processes.
3. Conscious states are realized in the brain as features of the brain system, and thus exist at a higher level than that of neurons and synapses. Individual neurons are not conscious, but portions of the brain system composed of neurons are conscious.
4. Because conscious states are real features of the real world, they function causally. My conscious thirst causes me to drink water for example.¹⁷

The theses solve the mind-body problem for Searle, but the rest of us may have a harder time accepting them. They may seem reasonable individually but can pose problems in conjunction. For instance, despite Searle’s denial of materialism and any kind of dualism, he has been labeled both in different instances. Labels of property dualism, functionalism, eliminativism, and more have been attached to Searle by others – perhaps the labels stem from some of Searle’s lack of self-doubt that even Descartes expressed in his own thoughts. No matter the reason, Searle’s philosophy of mind is not as clear as he expects. It is worth charitably exploring each thesis in turn, with a critical eye for where confusions typically arise. After this analysis, I will follow Kevin Corcoran’s line of thought in positing Searle as a “biological-property dualist” and determine if the description is apt. I’ll also delve into some of Corcoran’s objections to biological naturalism, as they implicate some of the biological properties he describes for biological-property dualism. Before delving into such properties, I will explain a critical underlying part of

¹⁷ Searle, *Mind*, 113-114

Searle's general framework – conceptual dualism – as it is an assumption of his towards most philosophers of mind.

§2. The Underlying Assumption of Conceptual Dualism

Searle accuses most philosophers of mind of acting under the scheme of ‘conceptual dualism’ – a term he uses to describe the underlying thinking in the traditional philosophy of mind. He claims, “this view consists in taking the dualistic concepts very seriously, that is, it consists in the view that in some important sense ‘physical’ implies ‘nonmental’ and ‘mental’ implies ‘nonphysical.’”¹⁸ According to Searle, due to a combination of holding onto Descartes’s dualist inquiries and our modern scientific advancements, we find it difficult to progress within the field because we are stuck like gum on conceptual dualism’s shoe.

In his *Rediscovery of the Mind*, Searle states four main reasons we are under the mark of conceptual dualism.¹⁹ The first reason is ‘the terror of Cartesian dualism.’ Searle says that ever since Descartes our progression in the history of philosophy and of science has drawn us to stick with observable, material progress, i.e., progress from empiricism. Straying from this path will inevitably lead us to a sort of Cartesian dualism and we are left with two irreconcilable ontological realms or with issues of privileged access, neovitalism, mysticism, and the like. Because of our difficulties in reconciling a mental property or substance with the material world we believe to be true, we immediately try to keep our distance from anything dualistic. The second assumption of conceptual dualism is that the history of philosophy and science have left us a vocabulary with its own implicit categories, namely the buckets of ‘mental’ and ‘physical.’ The mental *qua* mental is not physical and the physical *qua* physical is not mental. The stark and implicit opposition present in our language limits our thinking and leaves us under the framework of conceptual dualism. Our thinking is clouded by implicit dualities in language that

¹⁸ Searle, *Rediscovery*, 26

¹⁹ Searle, *Rediscovery*, 12-18

make us think of the mental and physical as mutually exclusive. In addition to these implicit dualities, a lot of language is tossed around and is implied to be equally understood by everyone – “mind,” “self,” and “introspection” are some prevalent examples.”²⁰ Although we may each have our own notion of what each word means, everybody’s conception of these abstract words tends to be variable enough to bring about confusion.

The third reason for conceptual dualism, according to Searle, is that modern science gives us the impression that whatever is true or real must be equally observable by anyone. We should be able to get *anything* down to an objective, third-person science if it is real. However, consciousness is not equally accessible in the way most things of science are – “the third-person character of the epistemology should not blind us to the fact that the actual ontology of mental states is a first-person ontology.”²¹ Lastly, Searle proposes we are under an *ivresse des grand profondeurs* – a drunkenness of the great depths. He believes we are not satisfied with the truth, despite its being held out in front of us. We keep looking for something deeper, but there is simply nothing there, says Searle. We simply need to accept the truth that Searle keeps repeating: As a result of biological processes, consciousness exists, is physical, ineliminable, and has causal potency. In other words, Searle suggests we sober up.

Searle provides the below chart to point out the implicit duality that the philosophy of mind operates under, specifically pertaining to consciousness. He extends his claim of conceptual dualism to say that we typically think of items one through three of the ‘Mental’ column as implying items four through seven of the ‘Mental’ column.²²

²⁰ Searle, *Rediscovery*, 15

²¹ Searle, *Rediscovery*, 16

²² Searle, *Mind*, 116

The Distinction between the Mental and the Physical

Mental	Physical
1) Subjective	1) Objective
2) Qualitative	2) Quantitative
3) Intentional	3) Nonintentional
4) Not spatially located	4) Spatially located
5) Nonextended in space	5) Spatially extended
6) Not explainable by physical processes	6) Causally explainable by microphysics
7) Incapable of acting causally on the physical	7) Acts causally and as a system is causally closed

© Oxford University Press. Source: John R. Searle, *Mind: A Brief Introduction*, Fundamentals of Philosophy Series (Oxford; New York: Oxford University Press, 2004), 116.

To put it clearly, Searle believes we are drawn to think that subjective, qualitative, and intentional features (first-person features) of consciousness imply that they are nonspatial, nonextended, inexplicable by physical processes, and causally impotent. But this would be a mistaken assumption since we *know* that items four through seven of the ‘Physical’ column (i.e., spatial location, spatial extension, microphysical causal explanation, and causal potency) must be true about our conscious states based on our most up-to-date advancements in science, and, most obviously, because of our lived conscious experiences. “As far as we know anything about how the world works, all [conscious] states are spatially located in the brain. (Indeed, with current brain imaging technology we are starting to discover something about their locations and spatial dimensions within the brain.) And they are entirely caused by brain processes. Furthermore, like any other higher-level feature of a physical system, they are capable of functioning causally.”²³ Because Searle thinks it is so obvious that the brain causes conscious states, he takes it for

²³ Susan Schneider, and Max Velmans, eds. *The Blackwell Companion to Consciousness*. 1st ed. (Wiley, 2017). Chapter 23: Biological Naturalism.

granted. Whether Searle's claim is empirically correct is yet to be proven beyond doubt. It seems Searle has simply identified neural correlates to conscious activity, but instead of hypothesizing on their 'mental powers,' Searle has simply posited the correlates as the cause of the mental. Unlike Descartes's radical doubt, Searle expresses no doubt in his brain.

Throughout all of Searle's attempts to clear the muddy waters of philosophy of mind, he still seems to have a lingering Cartesian smell to his work. First, he says that the subjective and qualitative aspects of what we consider to be the 'mental' are part of what bring out the first-person ontology of consciousness. Whereas most of the world is describable in the third-person, objective ontology, consciousness stands alone in the realm of first-person, subjective ontology. This point is the most worrisome for Searle's position on the mind as it has the most difficulty sitting neatly within his philosophy of mind, but it seems to be one of his most significant and pressing claims. Searle's position that all conscious states are spatially located in the brain can also be confusing because of his desire to distinguish himself from reductionists and/or eliminativists. Searle wants to imply that conscious states are simply *features* of the brain. The conscious states cannot be identified as *nothing but* neuron firings, but neuron firings do play some kind of role in shaping consciousness.

It is worth clarifying Searle's notion(s) of reduction at this point because he is constantly using the terms, especially in his theses on biological naturalism. Searle's explanation of reductions includes five different kinds of reductions: 1) ontological, 2) property ontological, 3) theoretical, 4) logical or definitional, and 5) causal reductions. Reductions are often employed to observe the identity relations between objects. Most important to Searle are ontological and causal reductions, but they are all worth a brief overview.

Ontological reductions are the most common kind of reduction where a certain type of object is shown to be consisting of nothing but objects of another type. For instance, “material objects in general can be shown to be nothing but collections of molecules.”²⁴ Next, property ontological reductions are a form of ontological reduction concerning properties and are closely tied to theoretical reductions since properties often correspond to theoretical terms (i.e., heat, light). Searle states that a “theoretical reduction is primarily a relation between theories, where the laws of the reduced theory can (more or less) be deduced from the laws of the reducing theory.”²⁵ Therefore “the reduced theory is nothing but a special case of the reducing theory.”²⁶ A theoretical reduction seems to find most of its use in science, as it may allow for a streamlined explanation of theories. Searle’s notion of theoretical reductions seems like it can hold a deeper place within Searle’s philosophical framework, but its murky nature and difficulty to distinguish from property ontological reductions can be difficult.

Logical or definitional reductions were once very popular in philosophy of mind as they would relate words and/or entities to each other by using logic or definitions. “Since the words and sentences are *logically* or *definitionally* reducible, the corresponding entities referred to by the words and sentences are *ontologically* reducible.”²⁷ This form of reduction is often employed by identity theorists. For instance, an identity theorist may claim that numbers are nothing but sets of sets. Lastly, and most importantly for Searle, are causal reductions: “this is a relation between any two types of things that can have causal powers, where the existence and a fortiori the causal powers of the reduced entity are shown to be entirely explainable in terms of the

²⁴ Searle, *Rediscovery*, 113

²⁵ Searle, *Rediscovery*, 114

²⁶ Searle, *Rediscovery*, 114

²⁷ Searle, *Rediscovery*, 114

causal powers of the reducing phenomena.”²⁸ What is confusing about Searle’s definition of causal reduction is that he uses the example of solidity to show how causal reductions work. Searle mentions that solid objects have causal consequences such as impenetrability by other solids or pressure resistance and that these causal powers are “causally explained by the causal powers of vibratory movements of molecules in lattice structures.”²⁹ But this explanation still seems to be hanging loosely between an ontological, property ontological, and causal reduction.

I see Searle’s explanation of reduction as an attempt for him to start separating consciousness and the experience of consciousness into another ontological realm. Searle is trying to bridge the gap that he thinks conceptual dualism creates in our understanding of the mind by appealing to causal reduction, but even he notes that consciousness is alone in being the only thing that is *solely* causally reducible. For instance, he notes that:

Real color has undergone a property ontological reduction to light reflectances. Similar remarks could be made about the reduction of heat to molecular motion, the reduction of solidity to molecular movements in lattice structures, and the reduction of sound to airwaves. In each case, the causal reduction leads naturally to an ontological reduction by way of a redefinition of the expression that names the reduced phenomenon.³⁰

Each mentioned case not only undergoes a causal reduction, but an ontological reduction follows. I believe Searle wants to avoid this kind of reductive pattern for the mind because it would carve off the ‘ontologically subjective’ character of conscious states. It seems to give everything in the world at least two properties: their ontologically objective properties, such as the light reflectance of colors, and their ontologically subjective properties, like the particular redness of a red in a particular subject’s experience.

²⁸ Searle, *Rediscovery*, 114

²⁹ Searle, *Rediscovery*, 114

³⁰ Searle, *Rediscovery*, 115

The ontological distinction that Searle consistently raises between the first-person and third-person ontological modes of existence makes it difficult to see him escaping the worries of property dualism. For instance, what property does the first-person mode have that the third-person mode does not? What is it about consciousness that is not identical to whatever causes it? Does Searle manage to escape these worries? I would hope so since he has even written a paper titled, "Why I am Not a Property Dualist," but it is still worth starting an analysis of his philosophy of mind by introducing the theses of biological naturalism. Picking apart his theses serves as a useful springboard to address his views on differences in ontological modes and, if he succeeds in his segregation of the two, what it means for something like consciousness to be a feature of the brain without being identical to it, reduced to it, or completely eliminated.

§3. Thesis One: Consciousness is Irreducible

Thesis [1]: Conscious states, with their subjective, first-person ontology, are real phenomena in the real world. We cannot do an eliminative reduction of consciousness, showing that it is just an illusion. Nor can we reduce consciousness to its neurobiological basis, because such a third-person reduction would leave out the first-person ontology of consciousness.³¹

Searle begins the first thesis of biological naturalism by outright stating his belief in conscious states. And Searle is insistent on this point: Conscious states, or better yet, consciousness, is a real phenomenon in the real world. There is no denying that our experiences are felt and experienced by us in a first-person point of view. I think most of us are inclined to agree here – I do. Whatever it is that I am experiencing as I live and breathe is done through my perspective. Although there are many interpretations of what such a ‘subjective, first-person ontology’ may imply, I believe Searle is attempting to address Thomas Nagel’s line of thought in “What Is It Like to Be a Bat?” Nagel emphasizes the subjective, phenomenal character of our experiences, stating that “the fact that an organism has conscious experience *at all* means, basically, that there is something it is like to *be* that organism.”³² Nagel further clears up his use of ‘what it is like’ in a footnote, stating that “it does not mean ‘what (in our experience) it *resembles*,’ but rather ‘how it is for the subject himself.’”³³ In other words, no matter how good our materialist theories currently are or how much the field of physics has advanced, it is not possible to reduce our conscious experiences down to an objective, third-person ontology without leaving out the subjective, first-person ontology of our phenomenological experience. This subjective ‘what it feels like’ character of our experience is what Searle wants to make room for.

³¹ Searle, *Mind*, 113

³² Thomas Nagel, “What Is It Like to Be a Bat?” *The Philosophical Review* 83, no. 4 (October 1974): 436.

³³ Nagel, “What Is It Like to Be a Bat?” 440

Here are some examples. When I eat a dish of my mom's red rice and refried beans, it is possible to explain the process on a purely empirical basis – we can talk about the digestive processes that my body will undergo and the kinds of chemical reactions that eating rice and beans will incite. But that would seem to leave out the subjective experience of what it is like *for me* to eat my mom's rice and beans. The sensation of eating this plate of rice and beans is even likely to be a different sensation for my mother. I may describe feelings of nostalgia and comfort from eating the dish, whereas my mom may describe mixed feelings of turmoil from cooking or having a long work day yet knowing her work is for her son. The qualitative character of the experience is, according to Searle, not captured by a third-person, objective reduction. Such a third-person reduction would result in an eliminative reduction of the experiential qualities.

If you have seen Pixar's *Ratatouille*, such subjective – what it is like – sensations can be seen when Anton Ego, a harsh food critic, is eating a dish of ratatouille. Ego had previously released a negative review of the film's central restaurant's (Gusteau's) food, but upon tasting this dish of ratatouille, he is immediately hit with the sensation and memory of eating his mother's freshly cooked ratatouille after a long day of being a tormented and angsty child. Eating the food then brings a calm smile of relaxation and comfort to his face. In such a case, I believe it is especially clear that there is a certain sensation of what it is like to be Ego eating the ratatouille, and such a subjective experience is not going to be the same for everyone. The subjective experience of *what it is like* for Ego to eat the ratatouille is unique to him. Other people's experience with the ratatouille may *resemble* Ego's, but ultimately Ego's experience is something that belongs to him, the subject of the experience.

When Searle mentions reductions in his first thesis, I think he is most fearful of following the camps of the Churchlands and Daniel Dennett. As explained earlier, Searle believes

reductions can come in various forms, but they tend to be *ontological* in the philosophy of mind. An ontological reduction is “the form in which objects of certain types can be shown to consist in nothing but objects of other types”³⁴ in the way that chairs are objectively nothing but collections of molecules in terms of their ontology. Ontological reductions, as conceived by Searle, often lead to ontological eliminations in this way – one ontological level (chair-level) is eliminatively reduced to a different ontological level (collections-of-molecules-level). But in the case of consciousness, we can only *causally* reduce. A causal reduction is “a relation between any two types of things that can have causal powers, where the existence and a fortiori the causal powers of the reduced entity are shown to be entirely explainable in terms of the causal powers of the reducing phenomena.”³⁵ Searle uses the example of solidity here: solid objects are causally explained in terms of the vibratory movements of molecules in lattice structures, but they also have causal powers of their own (i.e., resistance to pressure and impenetrability to other objects). In formal terms, Searle states:

We can say that phenomena of type A are *casually reducible* to phenomena of type B, if and only if the behavior of the A’s is entirely causally explained by the behavior of the B’s, and A’s have no causal powers in addition to the powers of the B’s... [In contrast,] phenomena of type A are *ontologically* reducible to phenomena of type B if and only if A’s are nothing but B’s.³⁶

The ‘nothing but’ component in Searle’s definition of ontological reduction is what holds most of the weight in this situation. Searle does not think that making reductions is intrinsically negative, but in the case of consciousness, a ‘nothing but’ reduction (ontological reduction) is not possible. Such a reduction would effectively *eliminate* consciousness because it leaves out the first-person, subjective ontology that we experience. Despite Searle’s distaste for reductions, he

³⁴ Searle, *Rediscovery*, 113

³⁵ Searle, *Rediscovery*, 114

³⁶ Searle, *Mind*, 119

does see a causal reduction being apt for consciousness. Even with Searle's allowance of causal reduction for conscious states, I find it difficult to accommodate the first-person, subjective character he is so concerned about; how is it that an objective, third-person ontology gives rise to a subjective, first-person ontology? For Searle, the causal reduction of conscious states still *prima facie* leaves out the first-person, subjective ontology. I will delve into these specific issues after going over each of the theses of biological naturalism because they can sometimes overlap, as thesis two will show. It is important to remember that although we may have issues with individual theses, it is really their conjunction that raises more issues.

§4. Thesis Two: Consciousness is Causally Emergent

Thesis [2]: Conscious states are entirely caused by lower-level neurobiological processes in the brain. Conscious states are thus *causally reducible* to neurobiological processes. They have absolutely no life of their own, independent of the neurobiology. Causally speaking, they are not something “over and above” neurobiological processes.³⁷

Searle’s second thesis distinguishes lower-level processes from higher-level ones. Lower-level processes for conscious states include the level of neurons or particles – the lowest level of explanation we can reach for biological processes. But Searle claims consciousness is not identical with these lower-level processes; consciousness is *caused* by the processes, but the processes are not themselves identical to consciousness. This is another area where Searle’s distinction between causal reduction and ontological reduction comes into play in the same way as it did in thesis one. Most of the time, when we derive a causal reduction, it correlates with or leads us to make an ontological reduction as well. But consciousness is unique in this sense because we can identify a causal reduction, yet the ontological reduction does not follow. Here, confusion arises on reading and describing Searle’s view because most philosophers would think that if consciousness is ontologically irreducible, then it must be something ‘over and above’ the neurobiology of the brain. Searle responds, saying:

No, causally speaking, there is nothing there, except the neurobiology, which has a higher level of consciousness. In a similar way, there is nothing in the car engine except molecules, which have such higher level features as the solidity of the cylinder block, the shape of the piston, the firing of the spark plug, etc. ‘Consciousness’ does not name a distinct, separate phenomenon, something over and above its neurobiological base, rather it names a state that the neurobiological system can be in.³⁸

³⁷ Searle, *Mind*, 113

³⁸ John R. Searle, “Why I Am Not a Property Dualist.” *Journal of Consciousness Studies* 9, no. 12 (2002): 60

What Searle is introducing in his use of causal reductions is a position he calls ‘levelism.’

Levelism is a model of “different levels or layers of description of a system.”³⁹ A car engine can be described at the level of its molecules or at the level of solidity, but both descriptions ultimately describe the same system.

If levelism is a correct way of describing some of the world’s processes, then Searle’s position is that consciousness can be explained at the level of third-person, neurobiological processes and at the level of first-person, subjective experiences, but both descriptions are of the same system. Neurobiology is a low-level description of the brain system and consciousness is a higher-level description of the brain system. Yet I do not think this is exactly what Searle wants to say because the lower-level description so plainly leaves out the subjective, first-person ontology he wants to keep in biological naturalism. The disconnect present in Searle’s appeal to levelism is a big hole he desperately needs to fill. Levelism seems to operate as a kind or variation of emergentism, where different levels of description describe different properties or levels of complexity. But such a position would also leave Searle facing issues of causal overdetermination or epiphenomenalism because a higher-level of description, such as consciousness, may lack causal powers, or two events, one physical and one mental, would be describing one event. Searle’s continuous grasping of the subjective ontology of consciousness, while appeasing the objective ontology of the underlying neurobiological processes, constantly puts him in a promiscuous position for a philosopher of mind.

In the process of distinguishing the causal reduction he advocates for from the ontological reduction he denies, Searle is also trying to distinguish himself from identity theorists in thesis two. Because neurobiological processes “have absolutely no life of their own, independent of the

³⁹ John R. Searle. “Consciousness, the Brain and the Connection Principle: A Reply.” *Philosophy and Phenomenological Research* 55, no. 1 (March 1995): 217.

neurobiology” they cannot be identified as conscious states or consciousness themselves. The identity theorist often uses the example that lightning is an electric discharge. Searle would agree with such an identification, but it is not analogous to the case of consciousness. Where the identity theorist would say that a conscious state is simply a neurobiological process, Searle would say that a conscious state is *caused* by a neurobiological process. Identifying the conscious state with a neurobiological process would be a type of eliminative reduction, says Searle, where consciousness is ontologically ‘nothing but’ a neurobiological process. And an ontological reduction, such as the one an identity theorist would make, is unacceptable and unrealistic for Searle. Again, we are left to teeter-totter between first-person subjective and third-person objective ontologies (and the smells of Cartesian dualism that the distinction brings up).

Searle’s emphasis in thesis two is on the neurobiological account of consciousness being a distinct thing from the experiential or phenomenological account of consciousness. Consciousness *can* be causally explained by neurobiology, yet your neurobiology is not what is experiencing what it is like to be you. No individual neuron is operating as a homunculus, thinking to itself, “It is my time to fire off! I am producing the experience of the color red along with my pals Neuroino and Neuronina the neurons.” Such a description would make no sense. Although the posited homunculus-neurons Neuroino and Neuronina are all *causing* your conscious experience, they themselves are not having the experience or any experience for that matter. At best, they are simply a lower-level description of a conscious state, yet not identical to the conscious state.

§5. Thesis Three: Consciousness is a Feature of the Brain

Thesis [3]: Conscious states are realized in the brain as features of the brain system, and thus exist at a higher level than that of neurons and synapses. Individual neurons are not conscious, but portions of the brain system composed of neurons are conscious.⁴⁰

Thesis three follows and expands on the line of thought mentioned in thesis two concerning car parts, levelism, and system features/descriptions. Searle's analogy to car parts asserts that "there is nothing in the car engine except molecules, which have such higher level features as the solidity of the cylinder block, the shape of the piston, the firing of the spark plug, etc."⁴¹ Features such as the shape, weight, or velocity of an object are 'system features' for Searle. The piston's shape would be a system feature. On the other hand, features such as solidity, liquidity, and transparency are 'causally emergent system features.'⁴² The cylinder block's solidity would be a causally emergent system feature. So,

On these definitions, consciousness is a causally emergent property of systems. It is an emergent feature of certain systems of neurons in the same way that solidity and liquidity are emergent features of systems of molecules. The existence of consciousness can be explained by the causal interactions between elements of the brain at the micro level, but consciousness cannot itself be deduced or calculated from the sheer physical structure of the neurons without some additional account of the causal relations between them.⁴³

Searle's stance on the role of features can be confusing, so I will try to take it step by step, if not for the reader's sake, then for my own. Solidity is a causally emergent feature of a system of molecules. But solidity can also be ontologically reduced to the vibratory movements of molecules. Consciousness is a causally emergent feature of, presumably, a system of neurons. But consciousness cannot be ontologically reduced to the neurons. Solidity is a different level of description of a system of molecules' movements. Yet consciousness both is and is not a

⁴⁰ Searle, *Mind*, 113-114

⁴¹ Searle, "Why I am Not a Property Dualist," 60

⁴² Searle, *Rediscovery*, 111

⁴³ Searle, *Rediscovery*, 112

different level of description of a system of neurons. It *is* a different level of description because of its causal relations; consciousness causally emerges from the activity of neurons. But it is *not* a different level of description because, despite being technically correct, the description will still leave out the first-person ontology we experience in consciousness. This same problem seems to continue arising for Searle! I understand his need to distinguish the ontological from the causal, but he provides no further clarity on how it is that a first-person ontology can exist within a third-person ontology. Searle seems to be saying that the subjective ontology of consciousness is caused by neural activity but no ontological reduction can be made from the brain's objective ontology without mentioning the causal reductions. No matter what, causality must be mentioned in the relationship between the brain and consciousness. The analogy between solidity's emergence and consciousness's emergence does not really seem to operate "in the same way" where it matters most.

Perhaps explicitly defining the implied notions of emergentism can help. Searle's notion of emergentism – which he labels 'emergent1' – in the case above is separated from a stronger kind of emergentism that Searle labels 'emergent2.' When something is 'emergent1' it is "a causally emergent property of systems... in the same way that solidity and liquidity are emergent features of systems of molecules."⁴⁴ In comparison,

A feature *F* is emergent2 iff *F* is emergent1 and *F* has causal powers that cannot be explained by the casual interactions of *a*, *b*, *c*... If consciousness were emergent2, then consciousness could cause things that could not be explained by the behavior of the neurons. The naïve idea here is that consciousness gets squirted out by the behavior of the neurons in the brain, but once it has been squirted out, it then has a life of its own.⁴⁵

Searle believes that emergent2 features are not even possible, as they would violate the laws of transitivity of causation that we operate on. If they do *not* violate natural laws, then emergent2

⁴⁴ Searle, *Rediscovery*, 112

⁴⁵ Searle, *Rediscovery*, 112

features would likely imply epiphenomenal properties, and Searle does not question the causal potency of his conscious states. When he wants his left arm to go up, his left arm goes up. Searle emphasizes that consciousness is not something separate from the neurons in the sense that it is not something extra or over and above them. But his ontological distinction between the third-person character of neurons and the first-person character of experience paints a different picture.

Overall, the main idea present in thesis three, that conscious states are realized as brain features, is meant to highlight that consciousness exists as part of the real, natural world and is not something that exists in an ‘over and above’ fashion. Searle says that consciousness is causally emergent from neurobiological processes. Although there are plenty of Neuroinos and Neuroinas around in our brains, they themselves are not having conscious experiences. But their operations taken together do give rise to our conscious states. In sum:

All forms of consciousness are caused by the behavior of neurons and are realized in the brain system, which is itself composed of neurons. What goes for thirst goes for all forms of conscious life whatever, from wanting to throw up to wondering how to translate the poems of Stéphane Mallarmé into colloquial English. All conscious states are caused by lower-level neuronal processes in the brain. We have conscious thoughts and feelings; they are caused by neurobiological processes in the brain; and they exist as biological features of the brain system.⁴⁶

If Searle’s conception of consciousness is still unclear, I would put my money on the confusion coming from the role he assigns causal reduction in our conscious states or how he links the first-person, subjective ontology of consciousness with the third-person objective ontology of our biology. I aim to address these issues in the section on subjectivity, especially pressing into how deep the analogy between causally emergent features of systems runs in comparing features such as solidity and liquidity with consciousness.

⁴⁶ Searle, *Mind*, 112-113

§6. Thesis Four: Consciousness Functions Causally

Thesis [4]: Because conscious states are real features of the real world, they function causally. My conscious thirst causes me to drink water for example.⁴⁷

Thesis four posits consciousness as causally efficacious and addresses the potential problems of causal overdetermination and epiphenomenalism. Because consciousness is experienced as and exists in a first-person ontology for Searle but is still part of a third-person ontology of the world, it may be difficult to assign or recognize causal powers of conscious states. If we grant the existence of conscious states under Searle's view, they seem epiphenomenal. Additionally, if not epiphenomenal, the threat of causal overdetermination lingers – what would we need an additional explanation of actions for if they can all be explained in third-person ontological terms?

From my experience, I know that when I want my arm to go up, it goes up. When I am hungry, I eat or make plans to eat. When I have a desire, I try my best to fulfill that desire. No matter the decision I make, it is true that I had (or at least feel like I had) a decision – the conscious states I have and experience cause *something* in my actions and beliefs. Despite these experiences of what it feels like to have a desire to raise my arm or satiate my hunger, some reductionists would try to reduce the desires, more broadly, the experiences, to their third-person ontology. Searle states how the eliminativist may speak about mental activity without 'mental vocabulary,' but the manner of speaking does not eliminate the occurrence of the activity in practice:

The point that we can talk about mental phenomena without using a mental vocabulary does not change the fact that the mental phenomena continue to have mental properties. My yellow-orange after-image remains qualitative and subjective whether or not we choose to mention those features... To put the point succinctly, the fact that one can mention a

⁴⁷ Searle, *Mind*, 114

phenomenon that is intrinsically qualitative and subjective in a vocabulary that does not reveal those features does not remove the features.⁴⁸

No matter how hard you try to remove it, the level of description of consciousness, the mental phenomena, will always exist. Talking about conscious states without mentioning them does not remove them from existing since they are ontologically irreducible. We may be tempted to characterize the descriptions in terms of their ontological properties (i.e., first-person, subjective vs. third-person, objective), but what matters most in the relationship is the causal properties of the entire system since they are what give rise to the system features. We can try to redefine our notions of what it means to have conscious states, but that does not take away the actual experience of them. Similarly, if we redefine and speak of solidity solely in terms of molecule movements, it does not mean that solidity does not exist – we just described it at a lower level of physics. A solid object is still solid no matter how you describe it. Analogously, a conscious state is a conscious state no matter how hard you try to leave it out from vocabulary.

Searle's stance on consciousness's causal efficacy may also give off hints of Jaegwon Kim's strong mind-body supervenience. Kim's supervenience states that "the mental supervenes on the physical in that if anything x has a mental property M , there is a physical property P such that x has P , and any object that has P has M ."⁴⁹ Any mental property is dependent on some other physical property. Furthermore, Kim could specifically address Searle's biological naturalism with a derivative of strong supervenience: "If an organism is in some mental state M at t , there must be a neural-physical state P such that the organism is in P at t , and any organism that is in P at any time is necessarily in mental state M at the same time."⁵⁰ If Searle agrees with Kim's

⁴⁸ Searle, *Mind*, 58-59

⁴⁹ Jaegwon Kim, *Philosophy of Mind*. 3rd ed. (Boulder, CO: Westview Press, 2011), 9

⁵⁰ Kim, *Philosophy of Mind*, 302

position, he may also face epiphenomenalism of the mind, which Kim accepts as possible within his materialist framework.

Searle addresses supervenience, saying that “it is certainly true that consciousness is supervenient on the brain;”⁵¹ and “mental states are supervenient on neurophysiological states in the following respect: Type-identical neurophysiological causes would have type-identical mentalistic effects.”⁵² But, unlike Kim, Searle does not think talk on supervenience adds much to philosophical discussion. Supervenience does no more work for Searle because it is only *causally* supervenient and not *constitutively* supervenient. Neurobiological processes *cause* our conscious states, but they themselves do not *constitute* our conscious states. This distinction is of the same vein as Searle’s notions of ontological and causal reduction. After conceding that consciousness is causally supervenient on brain processes, Searle states that

if [it] is right, and everything we know about the brain suggests that it is right, then the concept of supervenience adds nothing to the concepts that we already have, such concepts as causation, including bottom-up causation, higher and lower levels of description, and higher-order features being realized in the system composed of the lower-level elements.⁵³

Even if it is true that supervenience does not add to this philosophical discussion, Searle does an upsetting job of addressing the issue. He blows it off as a technicality, a ‘gotcha!’ moment in philosophy:

The fact that the mental features are supervenient on neuronal features in no way diminishes their causal efficacy. The solidity of the piston is causally supervenient on its molecular structure, but this does not make solidity epiphenomenal; and similarly, the causal supervenience of my present back pain on micro events in my brain does not make the pain epiphenomenal.⁵⁴

⁵¹ Searle, *Mind*, 148

⁵² Searle, *Rediscovery*, 124

⁵³ Searle, *Mind*, 149

⁵⁴ Searle, *Rediscovery*, 126

Mind-body supervenience is more than a technicality though. For instance, besides epiphenomenalism, it addresses some major philosophical dilemmas such as the explanatory gap – “the problem of closing the apparent ‘gap’ between pain and [neural state] N or, more generally, between phenomenal consciousness and the brain.”⁵⁵ On Kim’s account of supervenience, since the mind is epiphenomenal, the gap is not as pressing of an issue. For Searle, supervenience may pose a graver issue because he believes in a causally potent mind. I do not believe Searle directly addresses this issue, but I think he would point to the puppet master that is conceptual dualism as the source of our troubles. Because we are so entrenched in using the language of physical vs. mental, material vs. immaterial in our active philosophical framework (of conceptual dualism), we are left to make up implicit dualities instead of accepting the mental as simply coming from something physical. We are drunkenly searching for something more, for some great depth. But maybe Searle is drunk on the brain, an *ivresse du cerveau*. Searle likely believes he has no explanatory gap he even needs to address. There is no ‘hard problem of consciousness’ for Searle; we just need to get the scientists into the labs to figure out what system of neurons exactly is causally responsible for our conscious states. This just seems to leave him with a blatant reductionism.

It is of no help in Searle’s writings that, despite his denial of the traditional language and vocabulary of the philosophy of mind, he himself still uses the language and makes it difficult to parse out the truth. His attempt to “rescue the truth from the overwhelming urge of falsehood”⁵⁶ still leaves us using and presupposing what he claims is a false dichotomy to begin with. This issue is brought up by Searle when he explains his account of mental causation in his *Mind: A*

⁵⁵ Kim, *Philosophy of Mind*, 302

⁵⁶ Searle, *Rediscovery*, 2

Brief Introduction. There are typically four primary propositions regarding mental causation that, when taken together, are inconsistent. They are:

1. The mind-body distinction: the mental and physical form distinct realms.
2. The causal closure of the physical [CCP]: the physical realm is causally closed in the sense that nothing nonphysical can enter into it and act as a cause.
3. The causal exclusion principle: where the physical causes are sufficient for an event, there cannot be any other types of causes of that event.
4. Causal efficacy of the mental: mental states really do function causally.⁵⁷

One of these must be wrong if Searle wants to propose a coherent theory of the mind. Searle, as a firm believer and supporter of science, thinks proposition [2] must be true. The same thing can be said for proposition [3]. And because we know and experience our mental states as causally potent, proposition [4] must also hold true. That leaves proposition [1], which Searle gladly denies because of its implicit, dualistic assumptions. On the propositions, Searle says:

In general, as we have seen over and over, when you have one of these impossible philosophical problems it usually turns out that you were making a false assumption. I believe that is the case in the present instance. The mistake is expressed in proposition 1, the traditional mind-body distinction... The way out of this dilemma is to remind ourselves of a result we achieved in [chapter 4]: the reality and irreducibility of consciousness do not imply that it is some separate type of entity or property “over and above” the brain system in which it is physically realized. The consciousness in the brain is not a separate entity or property; it is just *the state that the brain is in*.⁵⁸

This is where Searle most clearly asserts he has no explanatory gap to address (making him seem like something of a reductionist). In his words, “there is no ‘link’ between consciousness and the brain, any more than there is a link between the liquidity of water and H₂O molecules. If consciousness is a higher-level feature of the brain, then there cannot be any question of there being a link between the feature and the system of which it is a feature.”⁵⁹ The rest of us are the ones that need to clarify our confused notions of the mental.

⁵⁷ Searle, *Mind*, 207

⁵⁸ Searle, *Mind*, 207-208

⁵⁹ Searle, *Rediscovery*, 105

Let's assume that Searle is right in some respects on mental causation. Importantly, let's assume that the physical is causally closed, and mental states have causal powers. How does Searle confront the issue of causal overdetermination? Again, he gives himself an out by arguing that it is traditional philosophy that is confused, not him. Our attachment to conceptual dualism leaves us with unnecessary puzzles:

There are supposed to be two problems about mental causation: First, how can the mental, which is weightless and ethereal, ever affect the physical world? And second, if the mental did function causally would it not produce causal overdetermination? The way to answer these questions is to abandon the assumptions that gave rise to them in the first place. The basic assumption was that the irreducibility of the mental implied that it was something over and above the physical and not a part of the physical world. Once we abandon this assumption, the answer to the two puzzles is first that the mental is simply a feature (at the level of the system) of the physical structure of the brain, and second, causally speaking, there are not two independent phenomena, the conscious effort and the unconscious neuron firings. There is just the brain system, which has one level of description where neuron firings are occurring and another level of description, the level of the system, where the system is conscious and indeed consciously trying to raise its arm... There are, of course, very difficult problems about how it actually works in the neurobiology, and for the most part we do not yet know the solutions to these problems.⁶⁰

The first unnecessary puzzle we assume is separating the mental from the physical. It is not that the mental does not exist, but for Searle, the mental *is* physical. As Searle might say, "I am a thinking being therefore I am a physical being."⁶¹ Anything that thinks, has conscious states, must be a part of the physical world not apart from it. The second unnecessary puzzle is how the mental could operate on the physical. But Searle believes this is not an issue for him because the 'mental' is not separate from the physical. The mental must be physical, and since it is physical, it is not causally overdetermining anything. We can give causal explanations of physical events, like the raising of my arm, in terms of intentionality and my desire to raise my arm *or* in terms of the biological processes that underpin the raising of my arm. If these two causal explanations can

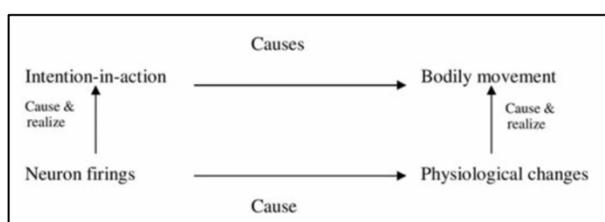
⁶⁰ Searle, *Mind*, 209-210

⁶¹ Searle, *Rediscovery*, 15

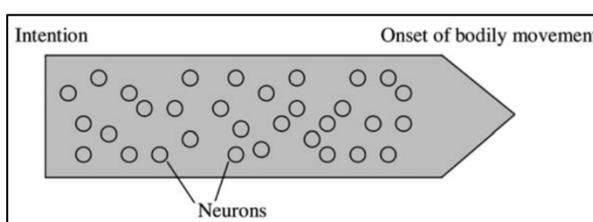
be *identified* with each other, then there is no problem of causal overdetermination. But Searle seems to deny this position because he believes that describing your conscious states in solely third-person terms will be leaving out the first-person character of the states. This issue is a continuous concern for Searle, but any issues we have with his conception are not philosophical, but neurobiological. For Searle, they are bound to be solved by the sciences in their due time.

Searle's explanation of the relation between the mental as a system feature of the physical just makes him seem like a reductionist, which he wants to steer clear of. He puts forth a few diagrams to further assist in his division between parts of a system and features of the whole system. A whole system, in the case of humans, the brain, is responsible for what we typically consider the mental. But make no mistake, this is a physical phenomenon. Searle insists that along with our implicit dichotomy of conceptual dualism, we also have confused notions of causality that contribute to our confusion. Figure 1 illustrates how such an explanation shapes up in the brain system.

Figure 1: Causation



*Figure 2: Causation in a System*⁶²



© Oxford University Press. Source: John R. Searle, *Mind: A Brief Introduction*, Fundamentals of Philosophy Series (Oxford; New York: Oxford University Press, 2004), 210-211.

In Figure 1, the physical, neuron firings, cause physiological changes and cause and realize an intention-in-action. The physiological changes cause and realize a bodily movement

⁶² Searle, *Mind*, 210-211

and the intention-in-action causes the bodily movement. But this just looks like causal overdetermination. Searle says that this model of thinking is just meant to clarify things, but it can still be misleading because of our already confused notions of causation. Figure 2 is meant to be a better depiction of how our conscious states, such as our intention-in-action work causally; “the conscious intention is shown as existing throughout the system and not just on top. In this one the circles represent neurons and the shading represents the conscious state as spread through the system of neurons.”⁶³

When Searle says that neurobiological processes are *causing* our conscious states, he does not mean that event A, such as neuron firings, happens, and then event B, such as an intention-in-action, follows. Causation does not always occur in discrete events. For instance:

Look at the objects around you and notice that they are exerting pressure on the floor of the room you are in. What is the causal explanation of this pressure? It is caused by the force of gravity. But the force of gravity is not a separate event. It is a continuous force operating in nature. Furthermore, there are lots of cases of simultaneous causation that are, so to speak, bottom up, in the sense that lower-level microphenomena cause higher-level macrofeatures.⁶⁴

On Searle’s account, our neurobiology gives rise to our conscious states in a similar fashion in which gravity operates on everything around us. It is not that gravity happens (event A) and then objects exert pressure on the floor (event B), but instead, there is a continuous gravitational pull that causes objects to exert pressure on the floor.

The comparison to gravity seemed to put my immediate qualms to rest. But after more time with Searle’s philosophy lingering in my mind, it seemed to me that Searle is left with a disjointed belief system. There is a disconnect in the comparison of the events of gravity and of the brain (and nervous system). The brain causes physiological changes and causes and realizes

⁶³ Searle, *Mind*, 211

⁶⁴ Searle, *Mind*, 123

intentions-in-action with neuron firings, which then leads to bodily movement, but this process is also happening in a first-person, subjective ontology. I presume gravity does not cause and realize a first-person, subjective intention-in-action while causing objects to exert downward pressure. Who knows, I could be wrong, but I doubt that is the belief Searle would like to ascribe to with his biological naturalism. His view consistently posits the first-person, subjective character of consciousness because we know and experience it, but it becomes difficult to see where it comes from.

Yet, still, he does believe that we can come up with an epistemically objective science of an ontologically subjective phenomenon. He must accommodate for the disconnect in events like gravity lacking an intention-in-action while neuron firings cause physiological changes *and* cause and realize intentions, conscious states, and the like. The biological properties of the world seem to have something special to them, most obviously that they can cause a subjective ontology in an ontologically objective world. Because this notion of subjectivity keeps coming up, I will spend more time discussing what it is that Searle is trying to preserve by positing such an ontological realm.

Chapter III

§1. Subjectivity and Other Loose Ends

As seen throughout the theses of biological naturalism, Searle's idea of consciousness relies heavily on defending the first-person, qualitative, and subjective character of our conscious states. He believes defending this conception of consciousness is necessary for justifying what we know of as our conscious, human experiences. Although defending the idea can be a noble project in the philosophy of mind, there are many ways to defend it, and I do not think Searle's has been the most efficient in the field as it leaves many readers puzzled about which label best fits him. His view does bring a much-needed push for what I think is the correct direction the field needs to go to, but this value can often be lost in the confusion. This concern has constantly been brought up in analyzing the theses of biological naturalism – we are usually left with the confusion of trying to figure out how it is that a first-person ontology emerges out of a third-person ontology. Part of this confusion arises from his use of subjectivity and how it ties into the various features of his definition of consciousness.

Searle's notion of subjectivity for consciousness just means that consciousness is experienced by a subject. Consciousness or a conscious state is necessarily *somebody's* consciousness or conscious state. Consciousness is experienced in the first-person perspective and includes features such as qualitateness (philosophically know as qualia) and intentionality. His notion of subjectivity most often refers to this first-person ontology. The ontological notion of subjectivity needs to be pulled apart from the common use and conception of subjectivity that we use in our daily lives. This fallacy of ambiguity, as Searle refers to it, is where one of our first confusions in understanding biological naturalism arises. When we think of subjective claims,

they are usually referring to epistemology, whereas Searle is referring to an ontological claim of subjectivity.

An example on music should drive the distinction between epistemic and ontological notions of subjectivity. When I am hanging out and talking with my friends, I might say, “Otis Redding’s ‘(Sittin’ On) the Dock of the Bay’ is the greatest song of all time.” My friends may then pose their own views: “That can’t be the greatest song of all time because ‘Girl’ by The Internet and Kaytranada is the greatest,” or “I 100% agree – Otis Redding’s performance is phenomenal.” The conversation can keep going on and on, with no one willing to change their own view on the matter. But this commonsensical conception of subjectivity is not what Searle frequently refers to in his use. The subjectivity being used in the conversation on the greatest song of all time is *epistemic* subjectivity. “We often speak of judgments being [epistemically] “subjective” when we mean that their truth or falsity cannot be settled “objectively” because the truth or falsity is not a simple matter of fact, but depends on certain attitudes, feelings, and points of view of the makers and hearers of the judgment.”⁶⁵ This notion of epistemic categories contrast epistemically subjective statements with epistemically objective statements. Although “Otis Redding’s ‘(Sittin’ On) the Dock of the Bay’ is the greatest song of all time” is an epistemically subjective statement, “Otis Redding was born on September 9, 1941” is an epistemically objective statement. An epistemically subjective statement depends on someone’s beliefs or opinions, while an epistemically objective statement is true no matter what anyone’s opinion on it is. But the epistemic mode of subjectivity is not what Searle refers to for consciousness; he refers to the ontological mode of subjectivity.

⁶⁵ Searle, *Rediscovery*, 94

The ontological form of subjectivity brings up some other peculiarities for consciousness.

The first is that

In consequence of its subjectivity, pain is not equally accessible to any observer. Its existence, we might say, is a first-person existence... Every conscious state is always *someone's* conscious state. And just as I have a special relation to my conscious states, which is not like my relation to other people's conscious states, so they in turn have a relation to their conscious states, which is not like my relation to their conscious states.⁶⁶

This notion of consciousness is murky and starts to mold itself into something like the privileged access that Searle vehemently denies. It will be worth addressing the issue of how Searle replies to privileged access and the problem of other minds in this section since it is so entrenched with ontological subjectivity.

Another peculiarity of a similar vein is that “all of my conscious forms of intentionality that give me information of the world independent of myself are always from a special point of view. The world itself has no point of view, but my access to the world through my conscious states is always perspectival, always from my point of view.”⁶⁷ Searle again emphasizes the ontologically subjective mode of consciousness's existence, this time adding some content on intentionality – the mental content we have that represents information in the world and in our minds. All the content, information, or, in all-encompassing terms, the intentionality, we have of the world comes to us from our point of view: “Consciousness and intentionality are essentially connected in that we understand the notion of an unconscious intentional state only in terms of its accessibility to consciousness.”⁶⁸ And Searle takes this to just be a baseline fact of our experiences. Intentionality, or “aboutness,” is also coming from our ontologically subjective perspective.

⁶⁶ Searle, *Rediscovery*, 94-95

⁶⁷ Searle, *Rediscovery*, 95

⁶⁸ Searle, *Rediscovery*, xii

How does Searle arrive at his view while abiding to the framework of the sciences and without eliminating consciousness from the equations? He provides a somewhat detailed explanation of how we get to subjectivity in his *Rediscovery of the Mind*.⁶⁹ My immediate thought on Searle's subjectivity was that he should just start with the ontologically subjective, first-person character of consciousness as a baseline fact, and *then* move on to explain the ontologically objective sciences we abide by. He does posit the ontologically subjective character of consciousness as a baseline fact of experience to an extent, but he still wants to arrive at the fact from a scientific position. Being the scientific erudite he is, Searle falls into some of the same traps he says that 17th Century empiricists like Descartes and Galileo fell into by being so loyal to the sciences. Despite recognizing the conceptual dualism we may operate under in philosophical discussion, Searle does no better to fully escape working under its scheme.

Searle starts reconciling the sciences and consciousness by clarifying how the sciences work, recognizing that his claims will “sound almost self-contradictory: On the one hand I will claim that consciousness is just an ordinary biological feature of the world, but I will also try to show why we find it almost literally inconceivable that it should be so.”⁷⁰ He begins the contemporary scientific worldview in the 17th Century with Descartes and Galileo, who provided useful heuristic tools to advance the field. Heuristics like the separation of an ethereal mind and material body were useful then, in the early days of our scientific worldview, but now they tarnish our perspective of seeing the mind as something material. Our attachment to such heuristics is one early reason we fail to advance a science of consciousness. We cannot accept

⁶⁹ Searle, *Rediscovery*, 83-93

⁷⁰ Searle, *Rediscovery*, 85

obvious truths because of our attachment to popular cultural heuristics. It is as if we are still operating on a geocentric as opposed to a heliocentric view of the world.

Following the advances in science that come in the 17th to 20th Centuries, we can better formulate various theories that we now take for granted. The most foundational of these theories for our everyday existence is the atomic theory of matter. “According to the atomic theory of matter, the universe consists entirely of extremely small physical phenomena that we find it convenient, though not entirely accurate, to call ‘particles.’”⁷¹ Everything physical in the world exists because of the system of particles that make it up. The laptop I am now writing on is made of a system of particles and so is the copy of *The Rediscovery of the Mind* that I am reading from. This notion of systems was mentioned in the third thesis of biological naturalism, where consciousness is said to be a feature of a brain *system*. This is the line of thought of Searle’s levelism, which is a kind of weak emergentism since higher-level system properties or features can be said to ‘emerge’ from the lower-level description of a system. For instance, the lower level of particles that make up the higher level of my laptop are not laptop-shaped, but the laptop’s shape emerges as a higher-level feature of the laptop’s particles. Searle qualifies this description as a causal explanation, where the lower level of particles causally explains the higher-level features of my laptop such as its shape. But the form of emergence employed by Searle is not meant to posit some over and above entity besides the system itself at a different level of description.

Building on the atomic theory of matter, Searle then introduces the principles of evolutionary biology into our common contemporary scientific worldview. Searle summarizes the principles of the theory in philosophers’ terms:

⁷¹ Searle, *Rediscovery*, 86

Over long periods of time, certain *types* of living systems evolve in certain very special ways... The ways in which they evolve are complicated, but the basic procedure is that token instances of the types cause similar tokens to come into existence. Thus, after the original tokens are destroyed, the type or pattern that they exemplify continues in other tokens and continues to be replicated as subsequent generations of tokens produce yet other tokens... Tokens that have a greater probability of survival relative to their environment will therefore have a greater probability of producing further tokens like themselves, tokens with the same genotype. And thus does the type evolve.⁷²

This description of our theory of evolution fits with other biological theories and/or models we have adopted such as Mendelian and DNA genetics. To build on this theory, we also *describe* evolution on various levels, in this case, the ‘functional’ and ‘causal’ (or ‘hardware’) levels of the underlying biology. These notions are meant to work with Searle’s levelism.

The functional level of description in this case is more general-audience-friendly in a sense since it can explain the evolutionary phenomena at a higher-level description than the chemical description can. For Searle, the functional level describes a causal phenomenon in terms of our interests. Strictly speaking, the functional level *is* a causal level. The causal level of description is meant to be more technical and focuses on the internal happenings of a system. It can be better described as the hardware level of description because of its focus on the relation of some phenomena to the environment. Searle uses the example of plants moving toward the sun in the process of photosynthesis to distinguish the levels. The hardware explanation of such movement is that “variable secretions of auxin cause plants to turn their leaves toward the sun” while the functional explanation is that “plants that turn their leaves toward the sun *are more likely to survive than plants that do not.*”⁷³ The functional and causal descriptions can be molded together to provide a better overall picture of how a system operates, but they both distinctly exist and are identifiable. Carrying the results of our two baseline theories of science – the

⁷² Searle, *Rediscovery*, 88

⁷³ Searle, *Rediscovery*, 230

atomic theory of matter and the theory of evolution – over to consciousness, we can come up with a description of how consciousness can exist and how it is possible that we are experiencing conscious states:

The products of the evolutionary process, organisms, are made of subsystems called ‘cells,’ and some of these organisms develop subsystems of nerve cells, which we think of as ‘nervous systems.’ Furthermore, and this is the crucial point, some extremely complex nervous systems are capable of causing and sustaining conscious states and processes. Specifically, certain big collections of nerve cells, that is, brains, cause and sustain conscious states and processes.⁷⁴

Searle does not see his theory of consciousness as something far-fetched and out of reach, but as the state of the world as we know it and as the next logical leap in our progression of theories on consciousness. Of course, theories could be wrong. The atomic theory of matter and the theory of evolution could be wrong, and Searle’s theory is not excluded from the possibility. But on this matter, Searle no longer thinks we are left with a philosophical question, but an empirical one. The hard problem of consciousness is a problem of neurobiology, not philosophy.

Searle places a lot of weight on having a theory that aligns with the empirical facts of the world. Most of the time, the empirical facts of the matter can tell us a lot about the world or our theories of it. But some thought experiments test out the limits of these empirical possibilities. For instance, philosophical zombies provide a challenge to Searle’s view of the mind: it is possible to imagine a world operating exactly as it currently does but without any consciousness or conscious states. In other words, it is logically conceivable to have your neurobiology function exactly as it currently does, but somehow there are no conscious states happening in your mind. Really, there is no mind. All there is to zombie-you are a bunch of neuron firings in the brain, blood pumping through the body, and food being digested in the gut – but there is no mental activity as we conceive of it. Searle likes this form of argument as it seems to give rise to an

⁷⁴ Searle, *Rediscovery*, 89

absurd question for mental activity: How is it that an ontologically objective chunk of meat is having ontologically subjective, qualitative experiences?

Searle proposes a rhetorical structure for the zombie argument.⁷⁵ Step one starts with Searle using the zombie argument against reductionism, most applicably, behaviorism. It is possible to imagine a creature that looks and behaves like me (Saul) but has no conscious activity going on. We can call this creature Zaul. Zaul shaves in the same way as me, Zaul performs the same weightlifting program as Saul, and eats the same foods that Saul does. But Zaul has no conscious states behind the hunk of human meat that he is – he just follows all the same behaviors and outward appearances that Saul would.

Step two addresses the conceivability of such Zaul-like creatures. If a creature like Zaul is logically conceivable, then we can generalize the possibility to the rest of the universe. Every creature with consciousness can be conceived of as having no conscious states at all; instead of Searle, there is Zearle and instead of Sehon there is Zehon. If such a world is possible, it seems that property dualism must follow. In a world with no conscious states, there seems to be only one property – the material. Whereas in a world with conscious states, there are seemingly two – the material *and* the mental. In order to accommodate for our mental states we must posit another set of properties or substances that could explain our conscious states.

This is where Searle introduces his third step and clarifies his position on the logical conceivability of zombies. When Searle says that it is possible to conceive of a beast like him – with all the same behavior, activity, and general demeanor – Searle meant it at a more superficial level than we may have expected. Sure, it is possible to imagine a beast with the same *behavior*

⁷⁵ Searle: *3 Philosophy Courses (UC Berkeley)*. Searle: Philosophy of Mind, 2014. <https://youtube.com/playlist?list=PL553DCA4DB88B0408>. Philosophy of Mind, Lecture 11, Minutes 34:30-37:00

and physical *appearance*, but if the beast has the same, underlying biological *microstructure* and the same laws that constitute the behavior for that microstructure, then it follows that the beast *must* be conscious. The exact same microstructure in the biology *necessitates* the existence of a conscious being. This is not a dualism for Searle, either, since it is clearly the result of biological processes – the material. It may seem odd that the world works this way, but it is just a matter of fact. There is no need to get drunk on the great depths.

Searle puts forth an analogy for the situation with electrons: “the charge of the electron is part of what it is to *be* an electron.”⁷⁶ Every electron *necessarily* has a charge as part of its conditions of existence. It would be wrong to say that an electron has no charge – it is not even possible for an electron *not* to have a charge. Similarly, if some beast has all the same biological microstructures with the same causal laws for their behavior as we currently have, coming from the same type of system (i.e., the brain), then that beast must *necessarily* be conscious and have conscious states. Searle admits he could be wrong here, but, again, it is no longer a philosophical claim, it is an empirical one. As Searle would summarize it, when your brain goes, your consciousness goes with it.

Searle’s response to the zombie argument also contributes to the confusion in understanding his philosophy of mind because of his use of “necessary” conditions. In a typical construal, necessary truths are true in every possible world – they are not just dependent on what labs discover. If his argument does work though, Searle’s position seems suspiciously like a reduction. But let’s say that Searle’s view is right: consciousness is the next logical leap in our scientific world view, building on top of the atomic theory of matter and the theory of evolution. Any claim denying consciousness would have to be empirical rather than philosophical because

⁷⁶ Searle: 3 Philosophy Courses (UC Berkeley). Philosophy of Mind, Lecture 11, Minute 34:00

consciousness “is a biological feature of human and certain animal brains. It is caused by neurobiological process and is as much a part of the natural biological order as any other biological features such as photosynthesis, digestion, or mitosis.”⁷⁷ The atomic theory of matter works at the level of physics, the theory of evolution at the level of biology, and the theory of consciousness is just an expansion on the knowledge we have from atoms and evolutions.

At this point Kevin Corcoran raises an eyebrow. Corcoran points out that the analogy Searle makes between photosynthesis, digestion, or mitosis and consciousness is relevantly *dissimilar* because consciousness operates under a first-person, subjective ontology, while photosynthesis, digestion, or mitosis can all be explained in third-person terms with no additional features or properties like a subjective ontology coming out of them. If consciousness is only experienced in a first-person manner, then it starts to look suspiciously like an endorsement of privileged access – “a model that suggests that consciousness is like a private room into which only we are allowed to enter.”⁷⁸ And Searle does not want to advocate for such a position as it seems utterly ridiculous to him.

Searle starts his reproach of privileged access by deconstructing it as a metaphor. The notion of privileged access is getting at a spatial metaphor, where “only I can go inside the space of my own consciousness.”⁷⁹ For Searle, this metaphor breaks as soon as it is posed, though, because the suggestion that there is even a ‘room’ that one is entering does not work. There is no separation between the room and the person with the privileged access inside it:

The metaphor of a private inner space breaks down when we understand that there isn’t anything like a space into which I can enter, because I cannot make the necessary distinctions between the three elements of myself, the act of entering, and the space in which I am supposed to enter.⁸⁰

⁷⁷ Searle, *Rediscovery*, 90

⁷⁸ Searle, *Rediscovery*, 98

⁷⁹ Searle, *Rediscovery*, 98

⁸⁰ Searle, *Rediscovery*, 98

I find Searle's position on privileged access intriguing because it breaks down the metaphor of privileged access, but it still seems to be leaving out some important components of explanation for Searle's biological naturalism. Even if the notion of privileged access is a confused one, how is it that we arrive at ontological subjectivity?

Because of his focus on the ontological subjectivity of experiences, Searle breaks down the metaphorical notion of privileged access quite well but glosses over the straightforward notion that the term is really trying to get at. A more straightforward version of privileged access may go as follows: I have privileged access to p if I can directly sense or know p , but it is impossible for anyone else to sense or know p (apart from my report of p). I believe Searle would concede that such a version of privileged access is mostly correct. We, as subjects of our experiences, are directly aware of our own conscious states but cannot observe the subject in question's consciousness. There may be some good guesses on the content of consciousness due to our underlying biology presumably being of the same type as other people and animals and because of what we know about how the environment affects what it is that we are conscious of, but ultimately, the ontologically subjective consciousness cannot be accessed by anyone other than the person whom the conscious state belongs to. On top of that, Searle may even assert that, in some senses, even we do not have access to our conscious states and their subjectivity. This latter claim can be confusing because our experience tells us that there is definitely a subject having experiences, but that the subject is also something fleeting and volatile. He says:

If I try to observe the consciousness of another, what I observe is not his subjectivity but simply his conscious behavior, his structure, and the causal relations between structure and behavior. Furthermore, I observe the causal relations between both structure and behavior, on the one hand, and the environment that impinges on him and on which he in turn impinges, on the other. So there is no way I can observe someone else's consciousness as such; rather what I observe is him and his behavior and the relations between him, the behavior, the structure, and the environment. Well, what about my own inner going-ons? Can I not observe those? The very fact of subjectivity, which we were

trying to observe, makes such an observation impossible. Why? Because where conscious subjectivity is concerned, there is no distinction between the observation and the thing observed, between the perception and the object perceived.⁸¹

For Searle, the ontological subjectivity of our experiences is irreducible and entangled in the ontologically objective world. The ontologically subjective (i.e., consciousness) allows us to observe the ontologically objective (i.e., the world) and make epistemically objective and/or subjective claims about them. He states our conception and relationship to ontological subjectivity as follows:

Our modern model of reality and of the relation between reality and observation cannot accommodate the phenomenon of subjectivity. The model is one of objective (in the epistemic sense) observers observing an objectively (in the ontological sense) existing reality. But there is no way on that model to observe the act of observing itself. For the act of observing is the subjective (ontological sense) access to objective reality. Though I can easily observe another person, I cannot observe his or her *subjectivity*. And worse yet, I cannot *observe* my own subjectivity, for any observation that I might care to make is itself that which was supposed to be observed.⁸²

There is a paradoxical atmosphere to Searle's interpretation and use of ontological subjectivity. But he takes it as a baseline *fact* of the world. It is not that subjectivity is something that needs to be arrived at analytically but is simply something that really exists in the world. It is our conception of the mind under conceptual dualism that limits our thinking.

Imagine walking through a rose garden, but in a third-person, objective view. No, really, take a second to imagine such a world without evoking a conscious state.

In attempting to imagine the walk through the rose garden, the walk was likely coming from a subjective ontology. That imaginative act comes from a perspective that focuses on certain elements of the garden or the experience. Most people will likely start by imagining what the garden and roses look like. That's what I did. But even this act of imagination comes from a

⁸¹ Searle, *Rediscovery*, 93

⁸² Searle, *Rediscovery*, 98-99

visual perspective. It strengthens certain points of the conscious experience of walking through the rose garden. Most people may not imagine the sensation of their clothes on their body in this act. This perspective is left out in the conscious state of imagining walking through a rose garden. You may also include certain smells of the rose garden, certain shades of red on a rose. And all of these conscious states are experienced *by* someone, a subject. Your act of imagining had a human perspective. It is not as if imagining a walk through a rose garden immediately had you pretending to see as a bee does in a garden full of pollen.

In this scenario, I pointed to various aspects of the imaginative act that were subjective because they were coming from a perspective. They belonged to a subject, namely you, and included elements such as the size and color of the roses, various smells, certain bodily sensations, etc. Now, instead of imagining the situation, try imagining what the subjectivity from the situation is or looks like. Try to imagine what my subjectivity is. Then try to imagine what your subjectivity is. When we look for subjectivity, we simply cannot point to anything concrete and say, “Ah, there is the subjectivity!” We are left longing for something that is not to be seen:

Indeed, it is the very subjectivity of consciousness that makes it invisible in the crucial way. *If we try to draw a picture of someone else’s consciousness, we just end up drawing the other person* (perhaps with a balloon growing out of his or her head). *If we try to draw our own consciousness, we end up drawing whatever it is that we are conscious of.* If consciousness is the rock-bottom epistemic basis for getting at reality, we cannot get at the reality of consciousness in that way. (Alternative formulation: We cannot get at the reality of consciousness in the way that, using consciousness, we can get at the reality of other phenomena.)⁸³

As seen, in some sense, Searle *is* promoting a variation of privileged access. Although the metaphor may break down because there is no room or space into which one enters, Searle admits that everyone stands in some unique relation to their own consciousness that no one else can experience or fully explain; namely their subjective perspective of the world. In some

⁸³ Searle, *Rediscovery*, 96-97

intuitive sense, it does not seem so absurd that we each have *some* kind of unique relation to our own conscious experiences. Searle poses such a relation to experiencing pain: “For [a pain] to be a pain, it must be *somebody’s* pain; and this is in a much stronger sense than the sense in which a leg must be somebody’s leg, for example. Leg transplants are possible; in that sense, pain transplants are not.”⁸⁴ The unique relation we have to our own conscious states does not necessarily mean that conscious states are exclusively closed off from third-party observation, but some relation to the conscious state must be defined so we can get at exactly what makes our conscious states belong to us.

Ultimately, what I believe Searle must propose for his philosophy of mind to properly include subjectivity and other features of consciousness into his view is a grander paradigm shift within the philosophy of mind. I believe he recognizes parts of this by stating his problems with the conceptual dualism prevalent in the philosophy of mind, but as it stands, current models of study just simply do not work for ontological subjectivity: “there is no way on that model to observe the act of observing itself. For the act of observing is the subjective (ontological sense) access to objective reality.”⁸⁵ As seen by trying to think of a rose garden in third-person terms, it is difficult to separate an ontological subject from their object of observation. “There is, in short, no way for us to picture subjectivity as part of our world view because, so to speak, the subjectivity in question is the picturing.”⁸⁶

I have belabored the point on Searle’s subjectivity because I believe it is a feature of consciousness that is unique to his view and that he spends more than enough time on in his writings. Presumably, all the writing Searle has done on subjectivity is meant to stress its

⁸⁴ Searle, *Rediscovery*, 94

⁸⁵ Searle, *Rediscovery*, 98-99

⁸⁶ Searle, *Rediscovery*, 98

importance in conceiving what the mind is. But his lunge into the topic still leaves some unsettling truths stemming from the conjunction of various of his beliefs. For instance, there are still explicitly two ontological realms for Searle: the subjective and the objective. Searle's subjective ontology leaves us with at least some variation of privileged access that is meant to be solved by supposing that the underlying biological causes of the other minds is similar to our own. It is simply the way the world is for Searle. In his view, "if solidity were conscious, it would seem to it mysterious that it was caused by vibratory movements of molecules in lattice structures, but all the same those movements explain solidity."⁸⁷ In the same way, it could seem mysterious to us that the brain's biology is what is causally responsible for our whole mental life. But that is simply how the world is.

Even if that is simply a matter of fact, it does not clear Searle from the Cartesian path of dualism he desperately longs to avoid. If I were more religiously inclined, I could posit the same set of biological facts that Searle employs, but expand on how Searle's notion means there may be a spiritual plane that allows for the subjective ontology to continue far after our material bodies leave the material world. Kevin Corcoran's analysis of Searle's biological naturalism may be motivated in such a fashion. Even if it is not, it is worth directly going into the charge of dualism against Searle and Corcoran serves as an excellent start by labeling Searle as a biological-property dualist.

⁸⁷ Searle, *Rediscovery*, 103

§2. Searle as a Biological-Property Dualist

Searle's philosophy of mind is often charged with a residual smell of dualism. When reading him, the charge strangely seems to make sense despite Searle's materialistic insistence.

Sometimes it seems so obvious that he is a dualist, while at other times it seems like he can be nothing but a monist. Even if we have Cartesian intuitions about biological naturalism, a close inspection of Searle's philosophy of mind shows him attempting to be a sort of substance monist. Throughout his theses, Searle constantly reiterates the importance of the physical and puts it at center stage with consciousness. I am sure Searle would shiver at even the label of substance monism, as it works *with* the labels of conceptual dualism, but it is apt for a general understanding. Kevin Corcoran's analysis of Searle succinctly summarizes why Searle is *not* a generic property dualist, despite our immediate intuitions:

Property dualism consists minimally in the claim that there are some properties, i.e., mental properties, that are distinct from physical properties. Searle's biological naturalism, on the other hand, seems only to commit him to the claim that *some* physical properties are at some level of description mental and that *all* mental properties are at some level of description physical... What exonerates Searle is the fact that for him there appear to be *no* non-physical properties, i.e., the mental, although genuine and irreducible, is for all of that, a natural, *physical* phenomenon.⁸⁸

Corcoran wants to say that Searle is no property dualist, but a property monist – everything is physical or has a physical basis. I mostly agree with the idea behind Corcoran's labeling of Searle, but I think he really means to say Searle is a *substance* monist. Sure, everything has a physical basis on Searle's view, but not everything is conscious. Only some things, namely biological entities, are conscious. Thus, Corcoran continues his own analysis of Searle and proposes him to be a 'biological-property dualist.' Is *this* form of dualism fitting for Searle? It is worth exploring what a biological-property dualist may be and if Searle fits the label.

⁸⁸ Kevin Corcoran, "The Trouble With Searle's Biological Naturalism." *Erkenntnis* 55, no. 3 (2001): 311

Corcoran calls Searle a biological-property dualist to address and accommodate the mental *qua* physical properties that Searle gives the brain. For instance, Searle professes that the mental is causally supervenient, subjective (and leads into some variation of privileged access as seen above), and causally efficacious. Furthermore, Searle's stance on identity theories, combined with his wanting to reconcile claims that (1) consciousness is real, i.e., mental, and irreducible, and that (2) consciousness is biological, i.e., physical, already raises some worries:

Given a rejection of identity theory, it seems only a commitment to some form of biological dualism can possibly render claims (1) and (2) consistent. However, Searle's own version of dualism seems to fall short of its aim of rendering (1) and (2) compatible insofar as it fails to explain how a physical, spatial property of the brain can be in-principle inaccessible to third person observation. It would seem, in the words of Russell, that Searle's biological-property dualism has all the advantages of theft over honest toil.⁸⁹

Corcoran sees Searle's biological naturalism as fitting under a materialist branch, but it gives the biological properties of the world a dualist spin: it "is a cross pollinated view of the two theories Searle so vociferously rejects – monism and property dualism."⁹⁰ The baseline belief of this biological-property dualism relies on everything being material. But from the biological, which is material, we get another set of properties. These properties, although material, have some causal powers that non-biological properties lack. The way Searle describes them is as if he wants them to be conceived of as something very ordinary, yet it is like chasing a shadow – you'll never catch it.

Biological properties can produce effects that non-biological properties cannot. For instance, biological properties can produce new cells, regulate digestion, and manage our chemical systems. Living (biological) beings are likely to have these and other biological properties that nonliving (and, thus, nonbiological) beings do not have. (However far these

⁸⁹ Corcoran, "The Trouble with Searle's Biological Naturalism," 314

⁹⁰ Corcoran, "The Trouble with Searle's Biological Naturalism," 312

properties go down the phylogenetic tree is likely to be another empirical question.) But the bigger concerns for Corcoran involve those traits of consciousness that we typically associate as mental in Searle's view, such as subjectivity, qualitateness, and intentionality. While nonliving beings have nonbiological properties that can be explained in third-person objective terms, living beings have both biological and nonbiological properties, and the biological properties cause our consciousness with all its typical 'mental' traits. A rock for instance can be explained in fully physical terms without seriously wondering what the first-person, subjective experience of the rock is. Attempting to explain a conscious state, on the other hand, always comes in some subject's perspective that is not wholly explained by third-person, objective descriptions.

Searle's stress on consciousness needing to necessarily be *somebody's* consciousness (i.e., Searle's notion of subjective) is the focus of Corcoran's analysis. He believes that despite all of Searle's repudiations of privileged access of the mind, Searle is proposing something like it, and I would agree. Under Searle's notion of a subjective, first-person ontology, there is a *subject* that is having conscious experiences in their first-person view – a subject could be conscious of their thirst, their desire to visit another country, or of the shade of yellow on a car. The consciousness is subjective because there is a *subject* to experience it. In interpreting Searle, Corcoran raises an immediate concern with the view on subjectivity and how it does not fit with Searle's go-to analogy to other causally emergent system features he refers to such as liquidity and solidity:

Question: How can some physical and spatial properties of the brain be accessible only to consciousness itself and what is the natural explanation for this bifurcation of biological properties within the brain? Searle does not say... [On Searle's view], consciousness and the mental are physical features of the *whole* brain, not features of any of its parts or their states. The problem is this. The emergent system-features of liquidity and solidity, for example, with which Searle wants to compare consciousness, are relevantly *dissimilar* to consciousness. Indeed, that a physical, spatial property of a thing should be *in-principle* inaccessible to third-person observation is utterly mysterious. What other emergent

features of physical systems have that quality? Indeed, this third-person inaccessibility seems sufficient to make one wonder whether consciousness really is a physical, spatial property of brains after all.⁹¹

Before the above quote, Corcoran expresses his issue with reconciling the first two theses of biological naturalism (that (1) consciousness is mental and irreducible and (2) consciousness is physical) and proposes variations of the second thesis to attempt to reconcile the *prima facie* issue or concern of how something can be both mental and physical. One variation Corcoran proposes is that mental states are identical with physical states of the brain. But Corcoran also recognizes that Searle does not agree with such a position since that would be a kind of eliminative reduction, leaving out the first-person subjective experiences we know and have in our daily lives. For Searle, an identity reduction ignores, and effectively eliminates, the subject of an experience or conscious state, leaving out an essential element of consciousness.

By denying that a mental state is identical with a brain's physical properties Searle is left with a sort of informal, descriptive emergentism. I think this is where he wants to be, especially with his constant comparisons to liquidity and solidity and mentions of causally emergent system features. On Searle's view, solidity is a causally emergent property of a whole system of molecules, not something found in an individual molecule. A single molecule or atom is not solid, but the system of molecules *is* solid. In the case of liquidity, you cannot point to an individual atom of water and say it is wet, but the whole system that comprises the body of water *is* wet. Searle's conception of emergence is quite straightforward in this way. Consciousness for Searle "is a biological process and interacts with other biological processes. Consciousness is a biological process like digestion, photosynthesis, or the secretion of bile."⁹² It is a property of a whole system (i.e., the brain), not something found in a particular part of a system like the brain.

⁹¹ Corcoran, "The Trouble with Searle's Biological Naturalism," 313-14

⁹² Searle, "Why I am Not a Property Dualist," 60

Following the analogy with solidity and liquidity, a single neuron is not conscious, but the system of neurons (or the whole brain) is conscious. Searle compares consciousness, liquidity, and solidity as causally emergent system features, but the analogy does not seem to correspond here.

Corcoran takes issue with the analogy between solidity or liquidity and conscious states because it is disanalogous when speaking of the ‘mental’ properties we associate with our consciousness like subjectivity, qualitateness and intentionality. Consciousness cannot be fully observed in third-person, objective terms, while solidity, liquidity, and any other such emergent, system feature can be observed, and both causally and ontologically reduced to third-person explanations. “The view that seems to emerge from Searle’s biological naturalism, then, is something like the following. Some physical, biological properties of the brain are segregated from others such that they are inaccessible to all but the one whose brain they are properties of.”⁹³ This accessibility issue arises because of Searle’s desire to accommodate for his use of ontological subjectivity by causal means. The unique property of consciousness makes Corcoran question how much consciousness can really be physical.

Subjectivity, in the Searlean ontological sense, arises as one of the main biological-properties in biological-property dualism because it is difficult to conceive of how a first-person ontology can be explained in a third-person ontology. It seems like each ontology can distinctly exist in its own ontological realm without being a major concern for the other. Just focusing on the third-person ontology will leave us with a reductive materialist position, while including the first-person ontology seems to point to dualism. Any third-person description of a first-person experience seems to leave out the essential first-person elements we consciously have as was

⁹³ Corcoran, “The Trouble with Searle’s Biological Naturalism,” 313

seen in the rose garden scenario used before (Imagine walking through a rose garden, but in a third-person, objective view). Again, take a second to try to imagine the world in a way that it is not coming from a perspective, and it will come from a subjective ontology. Searle recognizes this and attempts to make room for it, but I think he just needs to accept some variation of dualism in order to do so. At one point, he even states, “if solidity were conscious, it would seem to it mysterious that it was caused by vibratory movements of molecules in lattice structures, but all the same those movements explain solidity.”⁹⁴

Another property relating to subjectivity that needs more explanation in Searle’s account of the mind is the relation between the ontological and epistemic: Are there certain things we can’t know about, like if consciousness or subjectivity are observable, for instance? On Searle’s account, the third-person, ontologically objective world is experienced in the first-person, by a subject, hence it is experienced as ontologically subjective. We can come to epistemic truths of both the ontologically objective and ontologically subjective, but these truths do not always come in the same way. For instance, we can be sure that our ontologically subjective conscious states exist because we are experiencing them and we have made detailed sciences of physics, chemistry, and biology, which all work with the ontologically objective. Ontologically objective truths of the world are ontologically subjective in experience because they are always experienced by some subject. But they are true independently of the subject. For instance, if you perceive a table in your room, that table is there, independent of your perception of it. In contrast, ontologically subjective truths only seem to be true insofar as the subject experiences them as such. If a subject has the experience of perceiving a table in their room, they are seeing something real, but it is under their scope of experience. In perceiving the table, the subject may

⁹⁴ Searle, *Rediscovery*, 103

not be conscious of the shelf next to it, as an example. Thus, both the table and shelf exist in an ontologically objective sense, but only the table was ontologically subjectively experienced. Beyond the relationship of something being experienced, there still seems to be a disconnect in the ontology and epistemology of objective and subjective ontologies.

Perhaps observing the role of causation could clear things up in these modal relations, since Searle believes consciousness is just another biological process. Again, Searle's overarching position for biological naturalism is that "mental phenomena are caused by neurophysiological processes in the brain and are themselves features of the brain."⁹⁵ Let's rephrase the thesis with some Searlean terms: mental phenomena, which are ontologically subjective, are caused by, and hence causally supervenient on, ontologically objective neurophysiological processes in the brain and are themselves causally emergent system features of the brain. And to reiterate why this is *not* a property dualism, Searle says,

The property dualist means that *in addition* to all the neurobiological features of the brain, there is an extra, distinct, nonphysical features of the brain; whereas I mean that consciousness is a state the brain can be in, in the way that liquidity and solidity are states that water can be in.⁹⁶

The causal relationship between ontologically objective processes (neurophysiological processes in the brain) gives rise to ontologically subjective states and/or features (mental phenomena, such as our conscious states). This relationship is causally supervenient, not constitutively supervenient. Physical phenomena cause mental phenomena, but the physical phenomena do not constitute the mental phenomena. There is a causal reduction, but not an ontological one.

It still seems that the causal relations cannot wholly explain the ontologically subjective component of our consciousness. The first-person component of mental phenomena is meant to

⁹⁵ Searle, *Rediscovery*, 1

⁹⁶ Searle, "Why I am Not a Property Dualist," 61

be taken as something real and causally potent, yet it does not have any direct ontological linking beyond the causal processes that give rise to it. This is again where Searle emphasizes the empirical nature of the question. The philosophizing is done. Let us leave it to science. Maybe I am drunk on the great depths, but I do not find Searle's explanation to be entirely satisfying. I find it more appropriate to simply say biological-property dualism is an apt label for Searle because it addresses that some of the 'powers' that biology has are not present anywhere else in the physical world. Considering his emphasis on the causal role of the neurobiology, why does Searle not just join a functionalist camp with the likes of Hilary Putnam or David Lewis? It looks as if Searle is reaching for some deeper explanation too since the causal role of the neurobiology does not explain everything away for us.

In Searle's view, the subjective ontology cannot be explained away through purely ontologically objective phenomena like neuron firings. We can come to epistemically objective facts about this subjective ontology, but all these facts come from and presuppose the subjective ontology itself. Searle is clearly operating between two ontological realms with his positing of the subjective ontology. This is clearly a kind of Cartesian dualism. Even Descartes worked on and admitted the religious backing behind his own view. Searle just needs to accept a similar position. Perhaps I would be more charitable to Searle if he is indeed correct that we have to wait for the scientific advancements of neurobiology, but he is not so charitable to the AI Optimists' similar claim that in due time AI will inevitably have mental states of its own.

In "Minds, Brains, and Programs," he replies to this line of thought, which he calls the Many Mansions Reply. The reply basically kicks the philosophical can down the road, waiting for the science to catch up to what is in principle possible to achieve. Searle states the Many Mansions Reply as follows:

Your whole argument presupposes that AI is only about analogue and digital computers. But that just happens to be the present state of technology. Whatever these causal processes are that you say are essential for intentionality (assuming you are right), eventually we will be able to build devices that have these causal processes, and that will be artificial intelligence. So your arguments are in no way directed at the ability of artificial intelligence to produce and explain cognition.⁹⁷

The comparison between Searle's position on the brain and the AI Optimists' position on AI may not seem to correspond exactly, but the form of argument stands: We have a current science of some topic such as neurobiology or AI, and it has led us to make certain reasonable assumptions for the time. Although we cannot prove everything about our position *now*, the sciences will soon catch up. Descartes had a similar thought: eventually we'll explain the mind with the science of the pineal gland. Now, Searle believes we'll explain the mind with the science of the brain causality.

⁹⁷ Searle, "Minds, Brains, and Programs," 422

§3. What Comes Out of the Room

Now equipped with Searle's biological naturalism and his conception of systems/system features, Searle's rejection of the Systems Reply, and thus the Chinese Room Argument, loses some of its power. Searle seems to be left with some contradictions in his use of causally emergent system features. As mentioned before, the Systems Reply attributes understanding, not to the person in the room, but to the entire Chinese Room System. On this conception, the person in the room is analogous to something like a computer's CPU. Searle formulates the Systems Reply to the Chinese Room Argument as follows:

While it is true that the individual person who is locked in the room does not understand the story, the fact is he is merely part of a whole system, and the system does understand the story. The person has a large ledger in front of him in which are written the rules, he has a lot of scratch paper and pencils for doing calculations, he has 'data banks' of sets of Chinese symbols. Now, understanding is not being ascribed to the mere individual; rather it is being ascribed to this whole system of which he is a part.⁹⁸

Although Searle thinks this is an embarrassing position to support, Searle's own use of system features gives it more argumentative power. Searle refers to system features throughout his writing and poses consciousness as a system feature of the neurobiological processes going on in the brain. So why is the Chinese Room as a system not capable of producing some causally emergent conscious state such as understanding?

It is worth reintroducing Searle's idea of system features and how they tie into the mind: Thesis [3] of biological naturalism is Searle's primary assertion of consciousness as a feature of the brain system. It rids itself of any homunculi, namely the brain's neurons, and asserts the brain as a conscious system. Searle compares this idea to other features of the world he deems similar. For instance, on Searle's account, solidity is a higher-level system feature of collections of

⁹⁸ Searle, "Minds, Brains, and Programs," 419

molecules. No one individual molecule is solid, but the molecule-system is. Similarly, no one neuron is conscious, but the neuronal system (the brain) is conscious. In both scenarios, the feature (solidity and consciousness) emerges from the causal processes of the system, making them causally emergent system features.

This view on causally emergent system features from Searle's early work is principled, even if we disagree with its position. Yet when revisiting the problem of the Chinese Room Argument in a book chapter,⁹⁹ Searle seems to lose some of the power behind his principled stance on systems. While retrospectively summarizing the main points of the Chinese Room Argument, Searle states that:

The notion "same implemented program" defines an equivalence class that is specified independently of any specific physical realization. But such a specification necessarily leaves out the biologically specific powers of the brain to cause cognitive processes. A system, me, for example, would not acquire an understanding of Chinese just by going through the steps of a computer program that simulated the behavior of the Chinese speaker.¹⁰⁰

He explicitly says that the biology of the brain is what gives it the ability to produce cognitive processes. Searle really starts to shape up as something of a biological-property dualist as aforementioned. It seems that there are physical properties and mental *qua* biological properties in this world. Now we just need to consolidate them with our empirical knowledge. But it currently seems impossible because of the distinction left between subjective and objective ontologies. The biology seems to have some *extra* power to it, most notably the power to cause a first-person ontology of experience.

To this, I believe Searle would say that the distinction is simply true and, although we do not have a current detailed scientific explanation, the science will one day catch up. For now, we

⁹⁹ John R. Searle. "Why Dualism (and Materialism) Fail to Account for Consciousness." In *Questioning Nineteenth Century Assumptions About Knowledge, Iii: Dualism*, edited by Richard E. Lee, 5–48. Suny Press, 2010.

¹⁰⁰ Searle, "Why Dualism (and Materialism) Fail to Account for Consciousness," 17

must accept that the subjective ontology of the world is what allows us to observe the objective ontology and make epistemic claims about it. Some may say that the ontological subjectivity in question could be an illusion, in the same way that a rainbow is not really an arch out in the sky. And Searle says that in the case of consciousness, if it consciously seems to you that you are conscious, then you must be conscious. But for the Chinese Room System this ‘possibility from illusion’ is not present at all.

In the case of the Chinese Room, there is no way it is like *to be* the Chinese Room. While there is a way that it is like to be Nagel’s bat, there is nothing it is *like* to be the Chinese Room. It may be possible for the Chinese Room to simulate correct behaviors of Chinese understanding, translation, and/or conversation, but there is no intrinsic phenomenon to the Chinese Room that makes it have a certain way to be the Chinese Room. It just performs a task, namely a simulation of cognitive processes, but it does not duplicate them. The intrinsic intentionality is left out of the room in this simulative process.

Searle’s conception of intentionality is also closely tied to consciousness and is something produced by the brain as a causally emergent system feature. For Searle,

Many conscious states are intrinsically intentional. My present visual perception, for example, could not be the visual experience it is if it did not seem to me that I was seeing chairs and tables in my immediate vicinity... Not all consciousness is intentional and not all intentionality is conscious, but... there are, in fact, logical connections between the two: mental states that are in fact unconscious have to be the kind of thing that could in principle become conscious... An example of a conscious state that is not intentional is the sense of anxiety that one sometimes gets when one is not anxious about anything in particular but just has a feeling of anxiousness. Examples of intentional states that are not conscious are too numerous to mention, but obvious cases are those that exist even when one is sound asleep.¹⁰¹

Searle’s close ties between intentionality and consciousness imply that for the Chinese Room to be able to produce intentionality, it will likely need or will have consciousness as well; the

¹⁰¹ Searle, *Mind*, 138-139

intentional states would need to be in principle accessible to a conscious state. But since there is nothing that it is like to be a Chinese Room, as it lacks the causal powers of the brain, there is also no intentionality. All roads to the mind lead to the brain.

Additionally, Searle earlier left open the possibility of other objects having mental states of their own by stating that “if solidity were conscious, it would seem to it mysterious that it was caused by vibratory movements of molecules in lattice structures, but all the same those movements explain solidity.”¹⁰² The fact that we have mental features is simply a brute fact of the world, but he does not exclude the possibility of other beings having mentality. On top of that, it seems that there is no way that Searle is able to confirm if the Chinese Room System can have some emergent feature such as understanding because you cannot fully confirm or deny if it has a first-person ontology of its own. Obviously, it seems a little mysterious to think about a whole room as exhibiting understanding, but if we carry the analogy over to machines as systems, then this idea at least becomes no more mysterious than consciousness arising out of brain states.

Let me clarify by continuing with the example of ChatGPT. Currently, it seems pretty clear that ChatGPT is not performing any cognitive functions such as understanding. We may attribute as-if intentionality to it, but it is not actually understanding what it is producing. It is just regurgitating some set of information that it was given. ChatGPT is still in the symbol manipulation stage of the Chinese Room. Now, if ChatGPT was able to better improve its representations of the information it receives and better simulate an intelligent conversation with a human being, then it will slowly become more difficult to distinguish the program’s responses from those of another human. In a sense, ChatGPT’s ‘rule book’ becomes much better at

¹⁰² Searle, *Rediscovery*, 103

providing accurate instructions. Slowly, it becomes easier and easier for the program to pass Turing's Imitation Game and it becomes less clear if it really lacks intentionality.

Daniel Dennett brings up this concern in his initial reply to Searle in the 1980 "Minds, Brains, and Programs" article. He says:

How, though, would we know that [the Chinese Room] lacked these properties [i.e., understanding], if all we knew was that it was (an implementation of) a formal program? Since Searle concedes that the operation of anything - and hence a human brain - can be described in terms of the execution of a formal program, the mere existence of such a level of description of a system would not preclude its having intentionality. It seems that it is only when we can see that the system in question is only the implementation of a formal program that we can conclude that it doesn't make a little intentionality on the side. But nothing could be only the implementation of a formal program; computers exude heat and noise in the course of their operations - why not intentionality too?¹⁰³

I agree with Searle that the simulation of an event is not the same thing as duplicating it, especially for some cognitive faculty such as 'understanding.' But his biological naturalism seems to imply that the only thing that can exhibit and/or duplicate consciousness and/or intentionality is another brain. Yet, if we are to say that consciousness is a system-level property that emerges out of lower-level properties in a way that we still find rather mysterious, then why can't we say the same thing about intentionality arising as a system-level property of an appropriately complicated Chinese room situation, even when we still find it rather mysterious how that would happen? The very move that Searle makes concerning consciousness – that it is a causally emergent feature of a system as a whole despite not being recognizable as such in any of the parts of the system—would seemingly be available as a response to the Chinese Room Argument.

Searle frequently uses his move from systems to justify how it is that the brain produces consciousness. In one instance I mentioned earlier, Searle said:

¹⁰³ Searle, "Minds, Brains, and Programs," 430

The products of the evolutionary process, organisms, are made of subsystems called 'cells,' and some of these organisms develop subsystems of nerve cells, which we think of as 'nervous systems.' Furthermore, and this is the crucial point, some extremely complex nervous systems are capable of causing and sustaining conscious states and processes. Specifically, certain big collections of nerve cells, that is, brains, cause and sustain conscious states and processes.¹⁰⁴

Searle's reasoning, if truly principled, should be open Chinese Room as a system as well. It seems that it would be possible to have a sufficiently complex program or Chinese Room System that could at least potentially be conscious or have intentionality if we truly believe in the potential of causally emergent system features in the way Searle describes for the brain.

But that doesn't seem right. Searle does not want to say that the Chinese Room System has any kind of understanding, or generally, any intentionality behind its function. Searle would claim that the Chinese Room is simply not the kind of thing that can exhibit mentality. If Searle is not using the same underlying principles for his biological naturalism in his rationale for the Chinese Room System, then Searle is really left having to claim something like biological-property dualism or admit that his school of mind is not as coherent as he has supposed.

Despite the problem Searle faces with the disconnect in his use of system features of the brain and of the Chinese Room, Searle's Chinese Room Argument can still be a helpful thinking exercise in the fields of philosophy, cognitive science, and computer science. The Chinese Room's ability to baffle us has forced the field of artificial intelligence to look more deeply into how it is that the bland-looking, wrinkly, oatmeal-colored blob in our head – the brain – can cause and sustain our conscious life. The increase in interdisciplinary work in these fields shows as much.

¹⁰⁴ Searle, *Rediscovery*, 89

Conclusion

John Searle's influence in the philosophy of mind does not go unnoticed – the philosopher's Chinese Room Argument alone has left a major mark in the field. His own teachings in the philosophy of mind reflect increasingly materialistic norms in philosophy. Although he wants to break free from classic Cartesian worries, Searle's methods of inquiry closely resemble Descartes's own methods and follow a similar path. Both philosophers attempt to have their philosophy of mind fit in with their era's norms. Descartes worked with both religious values and scientific advancements, while Searle's philosophy reflects an increasingly secular age that depends on rigorous empirical means of study.

Searle's biological naturalism is both an intuitive and entangled view in the philosophy of mind. In many senses, it is intuitive because it works from Searle's intuitions about what is obviously true in the world. More importantly, biological naturalism is an intuitive view because its starting position comes from a place of attempting to work with any 'obvious truths' from *any* school of mind. Instead of immediately trying to defend a certain philosophical school or another, Searle's biological naturalism attempts to look beyond the traditional delimiters of what makes a mind. It accepts consciousness as real and the world we experience as real, while mostly making use of Occam's razor and attempting to simplify our world view.

Despite trying to work with what seems so obvious, Searle's biological naturalism is also an entangled view. Even with his charge of conceptual dualism, Searle has trouble escaping the scheme. He says everyone in the field has a terror of falling into Cartesian dualism. But so does he. Positioning him as a biological-property dualist emphasizes this. If Searle simply accepted the biological-property dualist position, he would at least be admitting to some of the unique biological powers that stem from the brain, such as giving rise to our ontological subjectivity and

how we have sort of privileged access to our conscious states. Only some system with the ‘same causal powers as the brain’ would be capable of causing and sustaining mental states with all their qualitative, first-person richness. And, as far as we know, the only system that is capable of causing and realizing such causal powers for our subjective, qualitative, first-person conscious states is the brain. It is simply the way the world turned out and we need not look any deeper, Searle would say – no need to indulge in an *ivresse des grand profondeurs*.

Still, taking Searle’s philosophy at face value, he seems to run into many issues of his own that make his philosophy of mind difficult to hold together and make sense of. Both his use of causally emergent system features and subjectivity bring new challenges to the field. Particularly, they seem to set him up as a variation of a dualist where biology is able to make consciousness and other mental features causally emerge. Whether we are looking for something deeper or not, I can appreciate Searle’s philosophy for challenging us to think deeply about the role of ontological subjectivity of experiences. Through all his writing on subjectivity, Searle addresses Nagel’s concerns on what it is like *to be* a certain kind of being with this notion of subjectivity. By identifying it for what it is, Searle is taking a step towards accomplishing a more complete philosophy of mind. His notion of subjectivity provides a challenge for traditional reductions since he believes the first-person ontology of consciousness cannot be reduced to the third-person ontology of the world without leaving out the what-it-is-like-to-be locution. This seems like an obvious truth for our experiences, yet Searle’s biological naturalism and its attempts to show how this first-person ontology fits into our philosophy of mind are slippery at best.

Searle’s emphasis on brains, subjects, and systems presents us with a complex and challenging framework of the mind to comprehend. Within the intricacies of his philosophy there

lie certain points that can be pointed against his renowned Chinese Room Argument, even if they may not align with Searle's intended response. Namely, the Systems Reply starts to look like a more cohesive response than Searle gave it credit for because it seems to use the same notion of causally emergent system features that Searle is so adamant about for the brain. The Chinese Room Argument's challenge against the notion that a computer program executing a set of rules can genuinely understand or possess consciousness loses much strength if Searle's idea behind causally emergent system features holds up. By considering consciousness as a casually emergent property of a system, one could argue that a computer system that has a sufficient complexity and organization could indeed possess consciousness, despite seemingly lacking subjective experience. Given Searle's acceptance of some variation of privileged access, the idea should not be completely closed off to Searle.

Overall, Searle's discussions on consciousness have yielded much insight into its essence. However, the complexity of biological naturalism, and the philosophy of mind in general, often hampers Searle's ability to hold his arguments together seamlessly. Consciousness remains an enigmatic phenomenon, defying purely materialistic efforts to be defined in third-person ontological terms. Searle's exploration of subjectivity and its relationship to consciousness adds further layers of complexity to the already difficult problem of consciousness. These additional layers should not be taken for granted or ignored but incorporated into any philosophy of mind for it to be functional. Our subjectivity of experience seems so obviously true that we tend to take it for granted. In spite of these additional complexities, it is crucial to approach Searle's work on the mind with open ears and a critical lens. It is within this intellectual struggle that progress can be made in the philosophy of mind. At worst, Searle's philosophy of mind is an incoherent school with a useful list of considerations for what makes the mind. At best, Searle's

philosophy introduces and reintroduces many important mental features back into the classic Cartesian mind-body problem. Continued engagement, exploration, and refinement of Searle's work on the mind can contribute to a deeper understanding of ourselves and the potential of inorganic life. Maybe one day, a Chinese Room System understanding will seem as obvious as much as we understand. Until that day arrives, we have much philosophical work to do.

Bibliography

- A Discussion of Artificial Intelligence with John Searle and Luciano Floridi*, 2016.
https://youtu.be/b6o_7HeowY8.
- Consciousness & the Brain: John Searle at TEDxCERN*. TEDx Talks, 2013.
https://youtu.be/j_OPOgPIdKg.
- Consciousness in Artificial Intelligence | John Searle | Talks at Google*. Talks at Google, 2015.
<https://youtu.be/rHKwIYsPXLg>.
- Corcoran, K. “The Trouble With Searle’s Biological Naturalism.” *Erkenntnis* 55, no. 3 (2001): 307–24. <https://doi.org/10.1023/A:1013386105239>.
- Debate on AI & Mind - Searle & Boden (1984)*. Voices, 2022. <https://youtu.be/b5a9NJ10iGQ>.
- Debate on Mind-Brain Relation: Searle vs Eccles (1984)*. Voices, 2022.
<https://youtu.be/1KemhfmAsg8>.
- Descartes, René, Donald A. Cress, and René Descartes. *Discourse on Method ; and: Meditations on First Philosophy*. 4th ed. Indianapolis: Hackett Pub, 1998.
- Dreyfus, Hubert L. *What Computers Can’t Do: The Limits of Artificial Intelligence*. Rev. ed. 7th print. Perennial Library. New York: Harper & Row, 1986.
- John Searle - Can a Person Be a Soul?* Closer to Truth, 2016. <https://youtu.be/Bq2bfSzkTfU>.
- John Searle - Can Brain Explain Mind?* Closer to Truth, 2016. <https://youtu.be/ehdZAY0Zr6A>.
- John Searle - Does Consciousness Defeat Materialism?* Closer to Truth, 2015.
<https://youtu.be/6oYk7fMmflw>.
- John Searle - Solutions to the Mind-Body Problem?* Closer to Truth, 2016.
<https://youtu.be/IgWbExnceHE>.
- John Searle Interview on Perception & Philosophy of Mind*, 2021.
<https://youtu.be/f2qZdGmq8vw>.
- John Searle: Our Shared Condition -- Consciousness*. TED, 2013.
<https://youtu.be/eqDgt12m26c>.
- Kim, Jaegwon. “Mental Causation in Searle’s ‘Biological Naturalism.’” *Philosophy and Phenomenological Research* 55, no. 1 (March 1995): 189. <https://doi.org/10.2307/2108318>.
- . *Philosophy of Mind*. 3rd ed. Boulder, CO: Westview Press, 2011.
- Lewis, David. “Mad Pain and Martian Pain.” In *Philosophical Papers Volume I*, by David Lewis, 122–32, 1st ed. Oxford University Press New York, 1983.
<https://doi.org/10.1093/0195032047.003.0009>.
- . “Psychophysical and Theoretical Identifications.” *Australasian Journal of Philosophy* 50, no. 3 (December 1972): 249–58. <https://doi.org/10.1080/00048407212341301>.
- Nagel, Thomas. “What Is It Like to Be a Bat?” *The Philosophical Review* 83, no. 4 (October 1974): 435. <https://doi.org/10.2307/2183914>.
- Preston, John, and Mark Bishop, eds. *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence*. Oxford ; New York: Clarendon Press, 2002.
- Professor John Searle : Consciousness as a Problem in Philosophy and Neurobiology*, 2014.
<https://youtu.be/6nTQnvGxEXw>.
- Schneider, Susan, and Max Velmans, eds. *The Blackwell Companion to Consciousness*. 1st ed. Wiley, 2017. <https://doi.org/10.1002/9781119132363>.
- Searle: 3 Philosophy Courses (UC Berkeley)*. Searle: Philosophy of Mind, 2014.
<https://youtube.com/playlist?list=PL553DCA4DB88B0408>).

- Searle, John R. "Consciousness, the Brain and the Connection Principle: A Reply." *Philosophy and Phenomenological Research* 55, no. 1 (March 1995): 217.
<https://doi.org/10.2307/2108322>.
- . *Intentionality: An Essay in the Philosophy of Mind*. 1st ed. Cambridge University Press, 1983. <https://doi.org/10.1017/CBO9781139173452>.
- . *Mind: A Brief Introduction*. Fundamentals of Philosophy Series. Oxford ; New York: Oxford University Press, 2004.
- . *Mind, Language & Society: Philosophy in the Real World*. 1. paperback ed., [2. pr.]. New York: Basic Books, 2000.
- . "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 3, no. 3 (September 1980): 417–57. <https://doi.org/10.1017/S0140525X00005756>.
- . *Minds, Brains, and Science*. The 1984 Reith Lectures. Cambridge, Mass: Harvard University Press, 1984.
- . *The Rediscovery of the Mind*. Representation and Mind. Cambridge, Mass: MIT Press, 1992.
- . "Why Dualism (and Materialism) Fail to Account for Consciousness." In *Questioning Nineteenth Century Assumptions About Knowledge, Iii: Dualism*, edited by Richard E. Lee, 5–48. Suny Press, 2010.
- . "Why I Am Not a Property Dualist." *Journal of Consciousness Studies* 9, no. 12 (2002): 57–64.
- Searle, John R., D. C. Dennett, and David John Chalmers. *The Mystery of Consciousness*. 1st ed. New York: New York Review of Books, 1997.
- The Famous Chinese Room Thought Experiment - John Searle (1980)*, 2020.
<https://youtu.be/tBE06SdgzWM>.
- Turing, A. M. "Computing Machinery and Intelligence." *Mind* 59, no. 236 (1950): 433–60