

UNIVERZITA KARLOVA V PRAZE  
EVANGELICKÁ TEOLOGICKÁ FAKULTA

Bakalářská práce

**The Solution to the Mind-Body Problem in  
Searle's Philosophy of Mind**

Jan Popelář

Katedra Filozofie  
Vedoucí práce: Jan Kranát, PhD.  
Studijní program: Evangelická teologie (6141R007)  
Studijní obor: Evangelická teologie

**Praha 2016**





## **|Prohlášení**

Prohlašuji, že jsem tuto bakalářskou práci s názvem *The Solution to the Mind-Body Problem in Searle's Philosophy of Mind* napsal samostatně a výhradně s použitím uvedených pramenů.

V Amsterdamu dne 27.4 2016

\_\_\_\_\_

## **Acknowledgments**

For carefully reading and correcting the entire text I am indebted to Evan Jehl, whose fine touch has made this text a much better one. For a great help with technical issues and logistics, I want to thank to my brilliant brother. For all the rest, my thanks goes to Rosa.

## **Anotace**

Tato práce se zabývá problémem mysli a těla ve filozofii mysli amerického filozofa Johna Searla. Zkoumá jeho kritiku filozofické tradice tohoto problému, která nás podle jeho mínění vedla k přijetí zavádějících kategorií a kritiku vědecké metodologie, která nás nutí k redukci mysli a vědomí na jejich fyziologii. Dále se věnuje jeho navrhovanému řešení problému mysli a těla, jeho definici vědomí, intencionality, subjektivity a také jeho slavnému argumentu “Čínského pokoje“. Ačkoliv tato práce zkoumá Searlovi teze kriticky a poukazuje na jejich častou problematičnost, není to komentář či rozvedení jedné jeho specifické teze. Jejím primárním účelem je zhodnocení Searlovi koncepcí vědomí v kontextu filozofické tradice problému mysli a těla.

## **Klíčová slova**

mysl, vědomí, problém těla a mysli, subjektivní zkušenost, dualismus, materialismus, redukcionismus, kognitivismus, psychologie

## **Summary**

This paper studies the mind-body problem in Searle's philosophy of mind. It thoroughly evaluates his criticism of the philosophical tradition and the scientific methodology. As he argues, the former has been misleading us with its dualistic distinctions and the latter has been pushing us into various forms of reductionism. Most importantly, though, it examines his proposed solution to the mind-body problem, his definition of consciousness, intentionality, subjectivity, and in addition, his famous "Chinese room" argument disputing claims of strong AI proponents. Although this paper does deal with many of the problems his claims and theories inherently contain, it is primarily meant to outline Searle's view on consciousness and its implications in the context of the mind-body problem tradition, rather than to be an extensive elaboration of his one specific thesis.

## **Keywords**

mind, consciousness, subjective experience, reductionism, materialism dualism, mental states, cognitivism, computation, psychology





# Content

<b>Introduction</b>	1
<b>Chapter I</b>	
<i>The mind-body problem in philosophy</i>	5
<b>Chapter II</b>	
<i>The mind-body problem in psychology</i>	14
<b>Chapter III</b>	
<i>Philosophy of science: The matter of methodology</i>	21
<b>Chapter IV</b>	
<i>John Searle's biography</i>	26
<b>Chapter V</b>	
<i>Critique of a materialistic and a dualistic approach to the mind</i>	28
<b>Chapter VI</b>	
<i>The solution to the mind-body problem</i>	36
<b>Chapter VII</b>	
<i>The definition of consciousness and its features</i>	41
<b>Chapter VIII</b>	
<i>Functionalism, Cognitivism, AI and thinking machines</i>	48
<b>Chapter IX</b>	
<i>What is wrong with Searle's philosophy of mind?</i>	57
<b>Conclusion</b>	66
<b>Appendix</b>	
<i>Yet another naturalistic theory of mind</i>	68
<b>Bibliography</b>	73



## Introduction

*“The important thing in science is not so much to obtain new facts as to discover new ways of thinking about them.”*

William Lawrence Bragg

---

What does it mean to have a mind? Is it to be conscious of one’s own consciousness, to be a subject of qualitative mental states? Or is it just a pattern of sequences of very complex routines in the information-processing unit, our brain? How do we come to have consciousness? Does it arise from the sheer complexity that only the human brain and possibly some other animals' brains achieve, or is it something else, something that in some ways transcends brain tissue? What is the relation, in general, between mental and physical phenomena?

There is no settled agreement on the correct answer. The trouble is that the question presents us with a problem: each *possible* answer to it has consequences that appear unacceptable. This problem has traditionally gone under the heading “The Mind–Body Problem“ and is today as complex as ever. The seemingly obvious nature of who one is becomes very uncertain under the inspection of the mind-body relation and the nature of our consciousness. After all, to study consciousness is to study the fundamental nature of our personal, subjective existence.

Conscious mental states undoubtedly define our lives but their subjective nature seems to defy scientific methodology and pose quite a challenge to current science. It is probably the single most important gap in our understanding of the natural world. The methods of scientific investigation, when applied to consciousness, inevitably leave out some of its essential features. The problem is that in the world of physics and the other natural sciences where particles and force fields exist in an objective way and causally interact with each other, there is simply no room for the inner subjective life that we daily experience. How do we then fit consciousness together with the world-view of science that is in many ways directly in conflict with our subjective experience?

It often seems that the scientific methodology leaves us no choice but to embrace some form of materialism while our subjective experience pushes one toward some form of dualism. Ludwig Wittgenstein expressed this realization with precision in his *Philosophical Investigations* (part I, section 412, p.124), writing: "...the feeling of an unbridgeable gulf between consciousness and brain-process ...When does this feeling occur in the present case? It is when I (for example) turn my attention in a particular way on to my own consciousness and, astonished, say to myself : This is supposed to be produced by a process in the brain! As it were clutching my forehead." The problem of consciousness lies uneasily at the intersection of philosophy and science, and there is no doubt that the scientific part of the problem will require philosophical methods of understanding.

While the “easy“ questions such as the neurophysiology of consciousness might be answered relatively quickly, the „hard“ problem of consciousness – that is, the relation between neurophysiological states and qualitative mental states - may not be resolved so readily. However, alongside the philosophical discourse dealing with this hard problem, a new science of consciousness has taken shape that integrates experimental and theoretical work across many fields including neuroscience, psychology, cognitive science, artificial intelligence and computer science. Developing a scientific account of consciousness is recognized by many as a major objective for twenty-first century science. The key factor in the transition to scientific legitimacy is the realization that it may not be necessary to explain why consciousness exists in order to unravel the physical and biological mechanisms that underlie its various properties. After all, physicists solved many mysteries of the universe without accounting for the brute fact of its existence.

In the following three chapters, I shall explicate the traditional mind–body problem and the evolution of our understanding of the mind and consciousness over time, indicating in some detail the background to Searle's philosophy. We will see that although the historical context and attendant philosophical framework had indeed changed, older concepts and ideas were oftentimes simply taken up and reused at a more sophisticated level.

The next four chapters, 5 to 8, are solely dedicated to Searle's criticism of dualistic and materialistic tradition, his solution to the mind-

body problem and the definition of consciousness and its essential features. In chapters 8 and 9, I study Searle's arguments against cognitivism and strong AI and point out the most blatant problems his theories and concepts contain.

# Chapter I

## The mind-body problem in philosophy

---

Although people have wondered about consciousness since they wondered about anything, the study of consciousness and the problem of the mind-body relation had its historical beginnings. We shall see that the modern version of the problem began to unfold when the philosophers of the seventeenth century tried to capture the nature of knowledge and how we come to acquire it. Soon after, the questions about the nature of the mind, subjective mental states, experience and even reality naturally followed. In other words, the epistemological inquiry consequently resulted in forming the first theories and concepts of the mind. And as far as the history of the mind-body problem is concerned, there is one man standing at the beginning.

Rene Descartes (1596-1650) was undoubtedly the key figure in the early history of the mind-body problem. After all, it was him that introduced the distinction between the mind and the body as the difference of two substances. However, it was an epistemological question of how one comes to have knowledge that led him into inquiring about the nature of the mind. In his skeptical epistemological enterprise, Descartes concluded that the thinking subject, “I”, is ultimately beyond all doubt and therefore its existence absolutely

certain.<sup>1</sup> The famous, indubitable “Cogito, ergo sum” became the very foundation of his theory of innate ideas - ideas, directly “given“ that are simply too “clear” and “distinct” to be doubted (and what is more, whose certainty is guaranteed by God) such as axioms of mathematics. Descartes thus believed the primary source of knowledge lies in the activity of mind whose content is accessible exclusively through one's introspection and as such is closed to scientific observation, i.e. to an objective, third-person point of view. On his view, consciousness (Descartes' concept of thinking covered all conscious mental states with no distinctions) is some sort of 'inner space', a theatre in which the mind's eye is like a homunculus watching the screen. He considered the mind (*res cogitans*) that is not spatially extended and seemingly follows rules on its own to be fundamentally different from body, which, as *res extensa*, is subject to the laws of nature. Descartes' definition of the difference between the two substances in terms of spatial extension will be, as we see later

---

<sup>1</sup> Descartes came to his famous realization in the following passage of his *Discourse on the Method*: “...considering that the very thoughts we have while awake may also occur while we sleep without any of them being at that time true, I resolved to pretend that all the things that had ever entered my mind were no more true than the illusions of my dreams. But immediately I noticed that while I was trying thus to think everything false, it was necessary that I, who was thinking this, was something. And observing that this truth ‘I am thinking, therefore I exist’ was so firm and sure that all the most extravagant suppositions of the sceptics were incapable of shaking it, I decided that I could accept it without scruple as the first principle of the philosophy I was seeking.” (Descartes, *Discourse on the Method*; CSM I: 126-7, 1637)



on, a very important one. This “dualistic“ distinction, however, poses a serious problem. How can two fundamentally different substances interact with each other, much less act upon each other? Quoting Gilbert Ryle, Descartes' mind seems like “a ghost in a machine” mysteriously acting upon the body, affecting bodily action. While Descartes was aware of this problem and tried to address it, his suggestion that the connection somehow takes place in the pineal gland was not quite convincing. The question thus remained: How does the mind relate to the brain? It is evident that Descartes' conception of mind (as the *inner* theatre) that is essentially beyond the reach of any scientific investigation could not be taken into much consideration, much less accepted by his peers so deeply impressed by Bacon's empirical method. As Francis Bacon suggested, in order to purge the intellect of “idols“, prejudices and illusions that hinder an empirical investigation, an observation and an experiment were to be the *only* legitimate sources of knowledge.

Perhaps, that was one of the reasons why the English philosopher Thomas Hobbes (1588-1679), also for some time Bacon's secretary, came to an understanding of the mind as a mere result of physical processes. Since there is nothing in this world but matter, according to Hobbes, and all effects have mechanical causes, Descartes' *res cogitans* cannot hold up. Unlike Descartes, Hobbes did not believe the mind is free from the necessity of natural laws; hence, freedom of the mind is nowhere to be found. In his view, there are no independent mental phenomena and mind, as anything else for that matter, is

reducible to biological processes. This view, also held by many today, is generally known as *eliminative materialism*.

However, despite the fact that the mind-body problem had been somehow established, most of the philosophical discourse of that time still focused on epistemology which is most evident in the works of the British empiricists, namely John Locke, George Berkeley and David Hume. No doubt, in many aspects they all held similar views on mind to Descartes' but unlike him, especially Locke and Hume, believed the mind can be actually studied - that is, its content described and classified.

To that end, John Locke (1632-1704) formulated association laws under which consciousness is supposed to derive all its content from sensory data. In his theory, mental representations, which he refers to as "complex ideas", are produced by associating external sensations with their internal copies. For example, the complex idea of "coffee" is a result of association of external sensations such as hot, liquid, brown, bitter with an *internal reflection* on them - simple ideas. In other words, physical properties such as *hot, liquid, etc.*, of an object cause sensations which then give rise to ideas. Locke, as a true empiricist, emphasized *receptiveness* of the mind rather than its *productiveness*. The mind in Locke's theory only mirrors simple ideas, copies a sensory input. Despite that, however, he still insisted that the internal images of the external objects are identical with those objects by assuming material reality behind them. He also rejected Descartes' concept of innate ideas on the ground of common sense, simply because

mathematical concepts or language are presumably not present in children's minds. That presumption is based on the notion of the human mind as *tabula rasa*<sup>2</sup> - that is being gradually filled up with sensory impressions that later translate into experience. Although Locke rejects this pillar of Descartes's theory, i.e. innate ideas, his theory of mind bears surprisingly similar consequences. Firstly, Locke's internal images of external objects seem to be "projected" on some kind of "mental screen", which evokes the question of who is watching the "screening" (similar to Descartes' homunculus). Secondly, what really is this complex idea and how can anyone verify the correspondence between *the world* and its *internal image* that helps to create them? And lastly, the knowledge obtained under Locke's association laws, is the knowledge of *ideas* and not of *the world* no matter how strongly Locke would have insisted that they are identical.

While Locke's theory contained the homunculus fallacy and could not sufficiently argue for the identity of the external objects with its

2 Locke writes: "The senses at first let in particular ideas, and furnish the yet empty cabinet, and the mind by degrees growing familiar with some of them, they are lodged in the memory, and names got to them. Afterwards, the mind proceeding further, abstracts them, and by degrees learns the use of general names. In this manner the mind comes to be furnished with ideas and language, the materials about which to exercise its discursive faculty. And the use of reason becomes daily more visible, as these materials that give it employment increase. But though the having of general ideas and the use of general words and reason usually grow together, yet I see not how this any way proves them innate. The knowledge of some truths, I confess, is very early in the mind but in a way that shows them not to be innate." (John Locke, *An Essay Concerning Human Understanding*, paragraph 15, 1689)

internal representations, it was still theory grounded in the physical reality. In contrast, George Berkeley (1685-1753), Locke's fellow countryman, was not bothered by any of those fallacies and even took some of them to yet another level. According to Berkeley, the mind is imprisoned in some kind of inescapable Cartesian theatre, nailed to its seat and never able to see what the real world is actually like. Everything in this world is subject-dependent, a constant conjunction of sensations, and sensation is all there is. He claimed the world exists only in the mind since its representation cannot be verified and thus avoided Locke's problem of correspondence. As he claimed in a *Treatise Concerning The Principles of Human Knowledge*, "to exist is to be perceived". So as far as we know, *matter* does not exist, since the mind is the only substance of which we surely know. Unlike Locke's ideas that at least still originate in sensation that come from material reality, Berkeley's theory does not need any. However, it makes one wonder who fills up the empty corners of the mind with sensations if there is no ground for them in material reality? Berkeley as Bishop of Cloyne had seemingly a simple answer: it is God who provides all the entertainment. God is the source of ideas and the existence of the perceived world essentially depends on him. This extreme version of Berkeley's idealism is well expressed in limericks written by Monsignor Ronald Knox, *God in the Quad*.<sup>3</sup> Berkeley's subjective idealism

---

3 "...here once was a man who said: God, must think it exceedingly odd if he finds that this tree continues to be when there is no one about in the Quad. Dear Sir, Your astonishment's odd: I am always about in the Quad and that's why the tree will

represents a complete opposite to Hobbes' reductive materialism. Surprisingly, this binary opposition (*no* mind versus *only* mind) still, although in a different form (nobody really advocates for idealism), frames much of the contemporary discussion.

No introduction to British empiricism is complete, though, without mentioning its youngest member, David Hume (1711-1776). Hume's work had a lot in common with both Locke's and Berkley's - espousing the former's laws of association and the latter's skepticism concerning the correspondence between sensations and the outside reality. Hume famously disparaged certainty of natural laws (mental habits) and discarded Descartes' indubitable 'I' as merely a virtual centre of perception along with his mental-material substance distinction. The association laws, introduced by Locke, Hume adapted as follows: law of similarity or resemblance - objects in the world bear a *resemblance* to their internal images; law of spatial or temporal contiguity - being reminded of something from the past may trigger a thought of something else that happened in that *time-space* frame - and law of causal connection -the associations are in a mutual *causal* relationship, i.e. their effects cause other associations. The association laws became very important as the predecessor of the reflex theory (association of *stimuli* and *responses*) that was later on adopted by behaviorism. One may wonder, however, about the reasons behind the ambition of the British empiricists to formulate those "mental" laws.

---

continue to be since observed by Yours faithfully, GOD." (Knox, *The Complete Limerick Book*, 1924)

Perhaps the mechanistic worldview that had emerged during the Renaissance urged them, just as it urged Newton in physics, to give mechanical explanations to seemingly organic mental processes.<sup>4</sup>

The last philosopher from this pre-modern scientific period, but perhaps the most influential one, Immanuel Kant (1724-1804), combined British empiricism with Descartes' rationalism, and realism with idealism. According to him, sensory experience is indeed fundamental but yet "blind" without innate ability to form concepts and categories that give a structure and sense to what is perceived. The outside reality, the physical world, does exist independently of an observer, but it is the mind that imposes a meaningful order on reality (such as relation between a cause and an effect). Kant himself recognized three so-called absolutely irreducible faculties of the mind: knowledge, feeling and desire. He believed they are irreducible because we cannot get one from the other, i.e. that there is something qualitatively different about thinking something from wanting something or feeling something. Sadly, he did not really express this idea in much detail. Kant's irreducible faculties of mind, however, were later renamed and found its use in psychology as the trilogy of mind, namely cognition, emotion

---

4 One of the letters written by Johannes Kepler, though concerning astronomical observation, represents this paradigm shift very well: "My aim is to show that the heavenly machine is not a kind of divine, live being, but a kind of clockwork, insofar as nearly all the manifold motions are caused by a most simple, magnetic, and material force, just as all motions of the clock are caused by a simple weight. And I also show how these physical causes are to be given numerical and geometrical expression." (Kepler, *The Logic of Personal Knowledge*, 1605, p. 52)

and motivation. Despite the influence Kant's work had in psychology, Kant himself did not believe in anything like scientific psychology. As he understood it, consciousness and its subjective experience has no spatial extension and therefore cannot be measured or mathematically organized. Moreover, the only method at hand to observe this subjective experience is introspection that is, according to Kant, by definition biased.<sup>5</sup>

In summary, the modern version of the mind-body problem was laid out by Rene Descartes. His dualistic concept of mind and body and an emphasis on activity of the mind and its inner aptitudes helped to create its basic matrix. The British empiricists contributed to the study of the mind by formulating mental laws and first theories of learning while placing emphasis on content of the mind and its relation to external reality. Thomas Hobbes, as one of the first, argued that since there is nothing but matter, all human beings are entirely material and thus the mind - subjected to the laws of nature – is not as independent as it may seem. And finally, Immanuel Kant expanded all of those concepts into a more complex and complete theory of mind.

---

5 In the most important text on psychology in his philosophy, *The Metaphysical Foundations of Natural Science*, Kant writes: "...even the observation itself alters and distorts the state of the object observed." (*Cambridge Edition of the Work of Immanuel Kant*, 1786, Ak. IV:471)

## Chapter II

### The mind-body problem in psychology

---

While Kant may have been right about the limitation of the scientific study of consciousness, it did not stop early psychologists from trying to figure out the nature of phenomenal experience.

In the nineteenth century, the mind-body problem made its way out of philosophy and took the form of introspecting the phenomenal nature of consciousness in psychology. The questions of the relation between the mind and the body or the origin of knowledge ceased to dominate the philosophical discourse, and the study of consciousness became the main subject matter in the newly established discipline of psychology.

Early psychologists focused mainly on so-called experimental observations. That included various techniques of introspection and response-stimuli experiments. One of the first such individuals was German psychologist and structuralist Wilhelm Wundt (1832-1920). Like many others at the time, he also regarded the method of introspection as the only possible way, given the phenomenal nature of consciousness, to comprehend our conscious experience. He believed that by using the introspection, one can discover single elements of consciousness and analyze different modalities of sensations (later known



as *qualia*). To Wundt and to other structuralists, consciousness was some sort of a chemical compound whose elements could be taken apart and studied one by one. They wanted to break up complex experience into irreducible elements, i.e. to find the most basic elements of conscious experience that cannot be reduced any further. For that purpose, Wundt founded a laboratory in Leipzig where he carried out various experiments on the introspective method. However, having been aware of the unreliability of introspection in normal conditions (he called it self-observation/every-day introspection), Wundt used well-trained participants that reported on their experience under strictly controlled conditions (inner perception/experimental introspection). Unlike *self observation*, the method of *inner perception* was not supposed to be biased. Nevertheless, it turned out that this method could capture only the most basic sensation. To Wundt's disappointment, the higher mental processes and the complex experiences that go along with them remained apparently inaccessible.

Another approach to the study consciousness was represented by psychophysics. Gustav Fechner (1801–1887) and Hermann von Helmholtz (1821-1894) studied the relation between physical properties of the physical stimulus and psychological properties of experience to which the stimulus gives rise. And because they thought sensation to be the most elementary conscious experience, they began with that. The principle of psychophysics is that every psychological quality of sensory experience is related to a physical property of the corresponding stimulus. Because the stimulus and the sensual experience are tied

up together, Fechner and Helmholtz studied what physical conditions give rise to conscious experience and developed methods to trace them, such as assigning numbers to certain experience or thresholds for conscious awareness. Although their endeavor did not bring the results they were expecting, the legacy both psychophysics and structuralism left behind was not forgotten. It later on became the basis for modern sensory neuroscience.

However, the true shift in psychology and the study of consciousness came with William James (1842-1910). James defined psychology as a description and an explanation of all mental states of consciousness as such.<sup>6</sup> He argued that among all the methods of psychology, introspection is the one we have to rely on “first and foremost”. This means we must turn our attention inward in order to analyze our mental life. James was the first one to have described the intrinsic features of consciousness such as personal subjectivity, unity, intentionality and selective attention. He understood these features in a mutual connection. As James wrote, all our thoughts tend to take personal form, i.e. every thought is *my* thought and every feeling is what *I* feel. This thinking *I* is always a subject of unified experience. Although thoughts may be constantly changing, no state once gone can occur

---

<sup>6</sup> James regarded introspection highly: “Introspective observation is what we have to rely on first and foremost and always. The word introspection...means, of course, looking into our own minds and reporting what we discover. Everyone agrees that we there discover states of consciousness. So far as I know, the existence of such states has never been doubted by any critic, however skeptical in other respects he may have been.” (William James, *Principles*, 1890, Chapter VII, p.185)

again and be exactly the same as was before. Consciousness does not appear to itself chopped up in bits but rather flows as a stream, as James famously put it. It is also sensibly continuous. Even when we fall asleep or something else renders us unconscious, we simply pick up where we left off, i.e. we feel being the same person as we were before. In addition, thoughts always have a content. They are always about something. James argued that all mental life is characterized by intentionality since every conscious mental state is intentional in its nature. That is why we are always being aware of or paying attention to only selected parts of what we perceive or think. James' characteristic of the intrinsic features of consciousness was the first of its kind and had much influence in psychology.

However, this interest in subjective experience and qualitative states of consciousness came to an end with behaviorists, who decided to solely focus on study of behavior. The key figures of behaviorism, J.B. Watson (1878-1958) and later B.F Skinner (1904-1990) did not think that studying phenomenal nature of consciousness is useful in any way. Moreover, they found its principal method – introspection - to be absolutely unreliable, no matter how controlled the conditions would be. They wanted to establish psychology as a genuine scientific discipline that produces verifiable results and carry out laboratory experiments whose outcome can be predicted. Since behavior, unlike phenomenal states of mind, is objectively observable, behaviorism became the first school in psychology openly disparaging mind as unworthy of serious consideration and a proper study. Behaviorists were obsessed

with scientific methodology and could not accept the kind of psychology that eludes it. As a result, psychology abandoned mind (at least in America) and everything that pertained to qualitative mental states of our consciousness and shifted its focus to behavior and functional relations between stimuli and responses. It was an example of the most extreme type of reductionism that has ever occurred in psychology. Behaviorists went as far as to claim that all of our mental lives can be explained through various types of conditioning and reinforcement. What is interesting about the academic hegemony of behaviorism, however, is the fact that it ignored not only the phenomenal part of our conscious experience but also cognition as such. And what is more, it lasted nearly four decades until the onset of the cognitive revolution. Despite behaviorist hegemony however, the interest in consciousness and phenomenal experience still persisted in German Gestalt psychology since the continental tradition was not as much driven by usefulness and applicability as its American counterpart. Historically, the continental tradition was more interested in *understanding* whereas Anglo-Saxon tradition more in causal explanations, functionality and utility. Although behaviorism has not succeeded in its effort to turn psychology into a science of behavior, it most certainly contributed to our understanding of learning mechanisms. We can see that in the history of psychology, introspectionists, Gestalt psychologists and behaviourists not only differed in their empirical approach to consciousness, but also, and perhaps first and foremost, in their philosophical ideas

about what science is, what the mind is and which methods are scientific.

The so-called cognitive revolution eventually overthrew the hegemony of behaviorism in the 1960's and became the dominant research line of inquiry in most fields of psychology. Yet it did not reconsider the idea of consciousness. Early cognitive scientists talked mostly about selective attention, memory, perception and language acquisition. The intrinsic features of conscious experience such as subjectivity and intentionality were once again neglected - needless to say, for the very same reasons that led behaviorists to focus solely on behavior. Cognitivists were dazzled with the rapid development of digital computers and quickly succumbed to the temptation to treat mind *only* as an extraordinary information processing software. Despite the fact that cognitive revolution did find a way to bring the concept of mind back into scientific research - by studying memory, attention, decision making, judgment etc., it still avoided the question of consciousness in its phenomenal totality (although it is somehow arguable whether there is a study of cognition without implicit or explicit study of consciousness).

American philosopher Owen Flanagan (1949) elaborated some of the reasons for this “consciousness shyness” in his book *Consciousness Reconsidered*. The first and the most obvious one is, as he calls it, “positivistic reserve”. Whatever consciousness may be it cannot be objectively studied and therefore must be left out of scientific investigation. This is most likely the reason why cognitive scientists prefer to avoid the subject of consciousness and focus instead on perception

and memory (Flanagan calls this second aspect the “piecemeal approach”). The third aspect is “Conscious inessentialism” i.e. we may agree on having minds and having mental states but it yet does not mean they are essential to anything that is going on in our brain. In other words, those mental states and processes could be essentially unconscious same as they are in computers. The fourth and the last aspect Flanagan calls “epiphenomenalist suspicion”, the idea that consciousness actually does not matter because it does not play any causal role in the world. Our mental states just *happen* in consciousness but all the important processes are unconscious. We may well observe things around us as much as we like and have whatever thoughts, but ultimately, it does not make any difference in the real world.

Even though it is not difficult to understand some of the reasons that Flanagan describes, it goes against our intuition and common sense that consciousness nearly fell into total scientific oblivion so many times. However, in the last two decades, thanks to philosophers such as John Searle, David Chalmers and many others, consciousness has become an independent scientific project in which scientists from many research fields are actively engaging.

## Chapter III

### Philosophy of science: The matter of methodology

---

Given the definition of the mind-problem that I outlined in the previous chapters, it is important to ask what it is about science that makes the problem so difficult? What is it about scientific methodology that consciousness just cannot fit in?

As we could see in chapter one, philosophers such as Descartes, Bacon or Hume already had some fundamental ideas of scientific methodology. At that time, the views on the sources and justification of knowledge could be schematically divided into two categories – *Idealism* (rationalism) that viewed reality as primarily existing within the mind and therefore focused on one's own subjectivity and *Realism* (empiricism) that believed the world to be objective i.e. mind-independent, only pictured or mirrored in the mind through representations and theories. Knowledge in idealism was more about constructing reality and truth was defined by coherence inside a theory, while in realism knowledge was viewed as a matter of representation and truth was based on correspondence between the statements and the facts of the world. Although neither of those views offer any guidelines for scientific investigation, their respective theoretical assumptions about the nature of reality had an impact on defining and shaping the scientific methods later on. The standard

image of science and the scientific knowledge is generally characterized by at least two things. First, by *well defined methods*, i.e. methods regulated by norms and prescriptions that determine what counts as legitimate questions, valid data and acceptable explanations and second, by *reductionism*, i.e. reduction to a more basic level that explains macro-phenomena through their underlying causes. Science thus provides explanations for observations and forms theories about causal properties of phenomena, proceeding from common sense observations to micro-structural explanations. That is the standard image of science. However, the real demarcation criterion, i.e. what distinguishes genuine science from speculation or pseudoscience has been a matter of discussion for quite some time.

In logical positivism, which shared the same assumptions about the nature of the reality with rationalists, the main criterion was *verification*. This means that a theory must be first tested against empirical data and then successfully verified in order to be accepted. For positivists, the source of all scientific knowledge lay in an observation that they believed was (or rather expected to be) neutral, objective and empirical. The theory was understood as being constituted by logical statements that have a basis in observations. Moreover, any statements about reality that cannot be translated into direct sensory observations were rejected as unverifiable. The methods of positivists were the same as those of empiricists - induction, cumulation, verification and confirmation. On their view, science progresses by



cumulating objective facts. However, many of these assumptions are rather problematic.

First of all, the strict separation between a theory and an observation, uncritically assumed in positivism, is false. As Wilfrid Sellars convincingly argued in his paper *Myth of the Given* there is no such thing as directly-given sense data that can serve as the basis for observation. Theory is underdetermined by the facts - changing theoretical assumptions can change the impact of observations (cf. Tycho vs Kepler). Simply, no observation is theory-independent. Secondly, the statements can never objectively represent facts of the world due to the nature of language (as was proven by Wittgenstein). And lastly, verification through cumulation of confirmed observations is in principal impossible as the “black swan” fallacy clearly shows – confirmation can never be certain. A single observation may lead to rejection of the related hypothesis as was convincingly proven by Karl Popper, who thus proposed *falsification* as the demarcation criterion of science. According to Popper, the theory needs to be tested by being repeatedly exposed to attempts of falsification in order to be considered as scientific. He argued that falsification is a far better characteristic of genuine science than verification since it also shows us whether a theory is in principal falsifiable and therefore scientific. For instance, Freudian psychoanalysis could never be considered a genuine scientific theory not because it cannot be proven but because it cannot be falsified in the first place. Freud notoriously always came up with ad hoc explanations. As Adler wrote: “Freud is always right”. The consequences of fal-

sification in Popper's theory are radical - every hypothesis that is falsified should be radically rejected.<sup>7</sup>

Despite the logical appeal of Popper's theory, one of the most influential theorists of science Thomas Kuhn (1923-1996) rejected falsification as the demarcation criterion and argued instead that it is paradigm that defines what science is and what it is not. Kuhn understood paradigm as a framework (worldview) in which the scientific practice takes place and described scientific development through the phases those paradigms undergo.<sup>8</sup> There is no cumulative progress but revolution. It is a dogmatism within a paradigm, Kuhn says, that determines the framework. Thus, every scientific paradigm has its specific set of demarcation criteria that cannot be applied across the paradigms (c.f. worldview of Aristotle's physics, Newton physics, Einstein physics). In a way, a scientific paradigm is a result of social activity and collective decision-making rather than a result of some universal methods.

---

7 However, many experienced scientists would probably still hold on to *falsified* hypotheses, arguing that relying on ad-hoc hypotheses has many times proven a good strategy. Radical rejection of the whole hypothesis because of its one falsified element is simply impractical. An ad hoc hypothesis may lead to observations that can eventually prove the original hypothesis.

8 First is the *pre-paradigm phase* in which there is only data collection but no framework yet. Second is the *paradigm phase*, which is a framework in which normal scientific practice takes place. Third, the *crisis phase*, is the phase in which too many anomalies that cannot be ignored appear. The final phase is the *revolution phase* in which the paradigm loses its hold and new methods, data, criteria and institutions emerge. After the revolution phase, the same cycle repeats itself (when the new paradigm stops working it gets simply replaced by a different one).

In addition, Austrian philosopher Paul Feyerabend (1924-1994) refuted the idea of the single and universal demarcation criterion as well. He argued that any single rule, however plausible, is going to be violated at some point in time.

Although it is evident that the scientific worldview and the scientific methods have undergone significant changes over the centuries, our conceptual apparatus apparently has not changed sufficiently enough. However, yet it may seem inconceivable from our current perspective, in principle, the *next* scientific paradigm or the one after may be capable of accommodating consciousness within its boundaries with no troubles at all. But that would most likely require a complete revision of *the* science in which our research methods dictate the subject matter of our research.

## Chapter IV

### John Searle's biography

---

John Searle was born in 1932 in Denver, Colorado, to the family of a business executive and a physician. He studied at the University of Wisconsin and Oxford University where he received a doctorate in philosophy in 1959. Shortly afterwards, Searle left Oxford for the University of California, Berkeley, where he has taught philosophy ever since.

His contribution to philosophy has been significant. One of Searle's earliest and most important contributions was in the area of philosophy of language. While he was at Oxford, Searle studied under British philosopher J. L. Austin, who had developed speech act theory. Searle is credited with having elaborated and expanded speech act theory and his notion of the "performative utterance" by examining the importance of rules in communication and by introducing the role of intentionality in constituting the meaning of speech acts.

In the early 1990s, Searle turned his attention to the problem of consciousness and since then has written about a dozen books on this subject. Searle's work in philosophy of mind covers most of the domains of the problem of consciousness. It contains sharp criticism of dualistic philosophical tradition, materialistic theories, and the theories

in cognitive science and artificial intelligence research. He is probably best known for his famous thought experiment, the “Chinese room”. His work has undoubtedly helped to bring the problem of consciousness to a broader audience since his lectures were broadcasted on the BBC radio and his articles published in *The New York Review of Books*. Consequently, it contributed to establishing institutes of Consciousness studies and academic programs at universities around the world. Since he has entered philosophy of mind, he has arguably become its most recognizable figure. For his various contributions to philosophy, he received the *Jean Nicod Prize* in 2000; the *National Humanities Medal* in 2004; and the *Mind & Brain Prize* in 2006.

## Chapter V

### Critique of a materialistic and a dualistic approach to the mind

*“Consciousness thus differs from other biological phenomena in that it has a subjective or first-person ontology, but this subjective ontology does not prevent us from having an epistemically objective science of consciousness.”*  
(J. Searle, *The Rediscovery of the Mind*)

---

As the previous chapters have shown, there have been and are many approaches to the mind-body problem and the study of consciousness. They can be generally divided in scientific materialism, philosophical dualism and cognitivism. There are many variations of both approaches, yet Searle is not in favor of either, and in this chapter, we are about to see why.

Searle spent a great deal of time arguing against materialism, cognitivism and dualism and tried to prove them wrong. He challenged the claims that the mind works like a computer just as much as the materialistic claims that either reduce or straightaway eliminate consciousness in favor of neurophysiology and the preceding dualistic claims suggesting that the mind and the body are two entirely different entities. In his comprehensive work on the subject of consciousness, Searle thoroughly analyzes the reasoning behind those claims while

showing why he finds them ultimately false and misleading. In the following paragraphs, we shall see that Searle's own account for consciousness and its related features is to a large extent defined by refutation of the philosophical tradition and some of the modern theories of the mind. However, before we get into the specifics of his critique of the materialistic tradition and cognitive science, we must first examine what those traditions have in common.

When we for instance take James' definition of consciousness, as far as its intrinsic characteristics are concerned, it is obvious that they are all by definition related to the observer. Science, however, operates on the presumption that all things that exist are physical and must be in principal observer-independent so they can be accessible to any observer. Therefore, it follows that consciousness under James' definition (which is the kind of definition most of us would agree upon) cannot ever become scientific subject matter unless its features are reduced to something else – i.e. anything that is free of the observer's subjectivity (neurons, functionality etc.). This implies that whoever would insist that subjectivity or intentional mental states such as beliefs and desires are irreducible and yet as real as any other physical phenomena is committed to essentially unscientific views.

Hence, consciousness (most notably intentional mental states) was reduced to behavior, seen as mere dispositional stances in behavioristic tradition<sup>9</sup> and left out entirely in cognitive science where

---

<sup>9</sup> In a behavioristic account, any intentional states can be translated into behavioral dispositions or behavior itself. "To say that John believes that is going to rain is

only objectively observable brain activity is of concern (such as information processing, memory etc).

Searle finds those traditional distinctions and vocabulary, “inherited apparatus”, obsolete and full of poorly defined categories for handling the problem. According to him, it forces us to basically choose between materialism and dualism as if there was no other option and nothing in between those two. Equally, he blames the scientific community for being blinded by its own methodology that makes impossible to approach consciousness properly, i.e. not leaving out essential features of our mind. On his view, this methodological rigidity and objectifying tendency (*positivistic* in its nature) can lead only to the ultimate absurdity: treating consciousness independently of conscious subject from the third-person point of view. Searle argues that actual ontology of mental states is *first-person* and not third-person ontology (beliefs or desires are always *someone's* beliefs and desires) and although mental phenomena may be inherently “subjective,” this does *not* mean they are subjective in the sense of existing only relative to human interests – the reality of our experience of pain for example, is not a matter of convention, and the pain exists whether or not we want it to.

He suggests to see the mental as part of the physical make-up of reality, not as something that goes against it. As with concepts like eco-

---

simply to say that John will be disposed to close the windows, put garden tools away and carry an umbrella if he goes out.” (Searle, J.R. *The Rediscovery of the Mind*. Cambridge, MA: MIT Press, 1992, p.33)



nomics, governments or hockey scores no one would deny to exist in spite of not consisting of physical particles, the mental phenomena, too exist, in objective reality. It is a mistake to divide the world up into physical and mental. Why could mental not be physical at the same time? Searle finds this distinction, based on the Cartesian definition of the physical as something spatially extended, simply obsolete and inadequate to describe the facts that correspond to statements about physical reality. Typical dualistic definitions of the physical and the mental that Searle finds outdated would probably go like this: The physical things are made of matter and energy; they are both indestructible and can transform into one another; and they are unconscious. Matter has extension and mass, and energy at least could potentially have these attributes since it can transform into matter. The mental, on the other hand, is destructible and cannot change into either matter or energy; lacks and could never acquire extension or mass; and most distinctively, is conscious.

Searle argues that one does not have to embrace the Cartesian metaphysics of the thinking and corporal substances to acknowledge the existence of mental states. However, given the fact that materialism has become the synonym to “scientific”, one feels the pressure to adopt such a view, assuming that it is the only choice. Searle thus appeals to our common sense and intuition that we all (more or less) share about consciousness while blaming various forms of materialism (in their respective research fields) for deliberately ignoring it in order to preserve their scientific status. While there is no

doubt that we have to reconcile our intuition about the mental with the common scientific worldview (considering how often our intuition may actually go against the factual e.g the Copernican paradox), it does not mean that we are left with no choice but to abandon consciousness all at once. He argues that insisting on the irreducibility of our mental states does not make one a dualist nor violate our scientific worldview in any way. However, as long as science, because of its methods, excludes the subjective qualitative states of our consciousness, no form of materialism, in Searle's view, can present a plausible theory of mind.

### **|Reductionism and Eliminative materialism**

Among the methods that define science, the method of reduction, i.e. the ability to reduce (ontologically) macro-phenomena on the basis of causal reduction, is the most prominent one. Due to the method of reduction we have been able to advance from common sense observations and descriptions to causal explanations. For instance, what was once described as a tasteless and colorless liquid – *water*, can now be simply reduced to the group of H<sub>2</sub>O molecules in *nothing but* fashion. Heat to molecular motion, genes to DNA, etc. While Searle has nothing against the method of reduction as such, he insists that we cannot do this type of reduction (ontological) in case of consciousness. Because although consciousness is, Searle says, a “causally emergent property of the behavior of neurons”, *causally* reducible, it is by no means reducible *ontologically*. Even if we developed a complete

account for causal elements (causes and effects), we could not make the standard reduction without missing out on essential features of our consciousness (e.g. subjectivity, qualia, etc.). The problem of subjective experience would simply not go away. No matter how much the third-person epistemology proved to be useful, the ontology of mental states is a “first-person ontology”. Moreover, the ontological reduction is not possible for another reason, namely, there is no subjective appearance-reality distinction since consciousness consists of appearances itself, Searle adds.

Searle has taken on many materialistic theories over the years, but he has always paid extra attention to its most radical form, so-called “eliminative” materialism. Unlike the classical form of reductionism that identifies a higher-level phenomenon with a lower-level phenomenon, this form of materialism, often accompanied with a crusade against “folk psychology”, states that there are *really* no mental states at all. The notorious advocates of this view in the scientific community are, for instance, the neuroscientists Patricia and Paul Churchland. According to them, in the perfect science of neurobiology there will be no room for expressions so typical for folk psychology like fears or desires or anything else that refers to *mental phenomena*.<sup>10</sup>

---

10 The Churchlands believe folk psychology to be a stagnating program (in Kuhn's sense) isolated from scientific progress and as such having to be eliminated and ultimately replaced by “neuro-speak”. Just as the famous phlogiston was once replaced in the theory of combustion by oxygen (although as a theory consistent with all of the observations made at the time and widely accepted as valid) or geocentrism by heliocentrism, folk psychology must be displaced in order to make a room for a psy-

All mental states will be at some point possible to reduce to neurophysiological processes and replace with neural-processes terminology and if somebody has a problem conceiving consciousness as a property of a neuron, it is simply due to a failure of imagination. Light at first was also not believed to be electromagnetic radiation, as they point out.

However, Searle alongside philosophers such as Daniel Dennett and Andy Clark do not see why such an elimination would be necessary. Folk psychology is obviously not a theory on the same level as neuroscience. Intentional states such as beliefs or desires are only descriptive, they do not aim at explaining internal causes. Folk psychology is just a manner of speaking, such as the sentence “People sometimes eat because they are hungry” clearly shows. Physicist Arthur Eddington captured the essence of this problem quite precisely

---

chology grounded in the real-world findings of neuroscience. The Churchlands, being very serious about their conviction, once gave a little example of how such a task should be carried out: “Paul, don't speak to me, my serotonin levels have hit bottom, ma brain is awash in glucosteroids, my blood vessels are full of adrenaline and if it weren't for my endogenous opiates I'd have driven the car into a tree on the way home. My dopamine levels need lifting. Pour me a Chardonnay and I'll be down in a minute”. (New Yorker magazine, the march 2007 issue). The main problem of their proposition, however, is that they got it somehow backwards. To state the obvious, the relation between the particular neurotransmitter and its affect on mood for instance (an example in NYM), can be established only because of people's self-reports on their mood states. Nobody would probably know what serotonin has to do with mood without measuring its level while the person reports on their particular state of mood.

in the introduction to his book, *The Nature of the Physical World*. It is evident that there are several levels of description and explanation that coexist in parallel (e.g. nomological, functional explanations) and from their respective viewpoints *exist*.<sup>11</sup>

---

<sup>11</sup>“I have settled down to the task of writing these lectures and have drawn up my chairs to my two tables. Two tables! ...One of them has been familiar to me from earliest years. It is a commonplace object of that environment which I call the world. How shall I describe it? It has extension; it is comparatively permanent; it is colored; above all it is substantial. By substantial I do not merely mean that it does not collapse when I lean upon it; I mean that it is constituted of "substance" and by that word I am trying to convey to you some conception of its intrinsic nature....Table No. 2 is my scientific table...My scientific table is mostly emptiness. Sparsely scattered in that emptiness are numerous electric charges rushing about with great speed; ...I need not tell you that modern physics has by delicate test and remorseless logic assured me that ...my second scientific table is the only one which is really there-- wherever "there" may be. On the other hand I need not tell you that modern physics will never succeed in exorcising that first table.“ (Arthur Eddington, introduction to *The Nature of the Physical World* (Cambridge, Eng., The University Press, 1928) ix

## Chapter VI

### The solution to the mind-body problem

---

Searle has repeatedly claimed that the primary challenge of the philosophy of mind is to show how our common sense picture of the mind can be reconciled with modern science, which tells us that nothing except particles in fields of force *really* exists. Searle attempts to bypass the conflict of intuition and science by proposing a redefinition of the philosophical framework in which the mind-body problem is set. While he calls himself a materialist, he is not in favor of any materialistic reduction as far as consciousness is concerned (for the reasons explicated in the previous chapter). His insistence on irreducibility of consciousness, however, earned him the reputation of a hidden dualist, a claim he has resolutely refuted many times. In fact, although Searle and dualists have something crucial in common: namely, they agree that mental states as standardly conceived exist and are not *really* illusions, behavior, functions, or computer programs, his solution to the mind-body problem appears to be explicitly anti-dualistic. He has even insisted that, despite the points of agreement mentioned above, we have to reject dualism because it can never explain the mind-brain interaction. With evident skepticism, Searle asks, "Are we supposed to think that our thoughts and feelings can somehow produce chemical effects on our brains and the rest of our nervous system? How could

such a thing occur? Are we supposed to think that the thoughts can wrap themselves around the axons or shake the dendrites or sneak inside the cell wall and attack the cell nucleus?"<sup>12</sup>

Searle insist we have to situate consciousness within our scientific understanding of the world. That means that our notion of consciousness must not be inconsistent with the fundamental theories of modern worldview such as atomic and evolutionary theory. To meet these conditions, Searle trivially describes consciousness as the result of a very long process of evolution in which our brain has developed to the point where it is capable of causing and sustaining conscious states. Searle is confident that although there is not much known about the neurophysiology of consciousness at the moment, it gets sorted out in the future neuroscience since consciousness, as he believes, is entirely caused by “the behavior of lower-level biological phenomena“ and hence common biological process. At the same time however, he argues that a denial of the reality of the mental, rather than being the necessary implication of science, is in fact a profoundly unscientific attempt to say that reality can only contain what our theories adequately account for. What, then, is Searle's theory that is neither dualistic nor materialistic (in the sense of reductionism) and yet can somehow offer the solution that may just resolve the age-old problem? Searle's solution, which he calls “biological naturalism”, basically consists of two simple propositions: First, the mind is caused by the brain,

---

12 John Searle, *Minds, Brains, and Science* (Cambridge, MA: Harvard University Press, 1984), p.17

and the second, the mind is a feature of the brain. "Mental phenomena are caused by neurophysiological processes in the brain and are themselves features of the brain so mental events and processes are as much part of our biological natural history as digestion, mitosis, meiosis or enzyme secretion."<sup>13</sup> That is to say that the mind is both caused by the neurons and at the same time is a feature of the neurons. Searle's favorite example that illustrates this thesis is the liquidity-H<sub>2</sub>O molecules relationship. The feature of liquidity that we observe in water, says Searle, is caused by the underlying molecular features of water. At the same time however, liquidity is not some extra property of the H<sub>2</sub>O molecules. It is one feature of the molecules. Consciousness is thus "a higher level or emergent property of the brain in the utterly harmless sense of higher-level or 'emergent' in which solidity is a higher-level emergent property of H<sub>2</sub>O molecules."<sup>14</sup> As Searle explains, it is all matter of a level of description. Often, we can describe something either at the micro- or the macro-level. In all of these cases, the macro-level is caused by the micro-level, and is at the same time identical with the micro-level. It follows that on this model, the mind is a macro-level property of the brain, and the neurons are the micro-level of the brain. The brain causes the mind, and the mind is a higher-level feature of the brain.

The very essence of Searle's solution is to deny both that the mind is a different thing than the brain and that the mind cannot cause

---

13 John Searle, *The Rediscovery of the Mind* (Cambridge, MA: MIT Press, 1992), p.24

14 Searle, *The Rediscovery of the Mind*, p.25



anything. As Searle puts the apparent dilemma: "Either you have dualism and an unintelligible account of causation or you have an intelligible account of causation and abandon the idea of the causal efficacy of the mental in favor of some version of the identity thesis with an attendant epiphenomenalism of the mental aspects of psychophysical events."<sup>15</sup> In other words, the solution to this dilemma is hence saying that the mind is a higher-level features of the brain and that since the mind itself is physical, causation is conceivable and indeed real. Just as the liquid state of water can be causally dependent on the behavior of the molecules and can also be a feature of the system made up of the molecules, so too the mind can be caused by the brain and be a feature of the brain at the same time. On his view, consciousness is a mental property of the physical system and therefore physical as well just like liquidity is a property of the physical system of molecules. Searle attempts to escape the obvious hopelessness of Cartesian dualism by simply stating that the mental is also physical, so it is obvious that mental things could causally interact with non-mental things, because they are both physical. And even though it may be difficult to understand how they relate to each other (on the neurophysiological level), it is easy to imagine that they could (unlike dualism, which posits "two realms", making it hard to see the possibility of the interaction, much less its mechanism).

---

15 John Searle, *Intentionality: An Essay in the Philosophy of Mind* (Cambridge: Cambridge University Press, 1983), pp.264-265

Searle's way of bringing consciousness back into the subject matter of science is defining consciousness as an emergent property or higher feature of neuron interactions and *as such* as quite an ordinary biological process similar to digestion or enzyme secretion. "Consciousness in short 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."<sup>16</sup>

---

<sup>16</sup> Searle, *The Rediscovery of the Mind*, p.90

## Chapter VII

### The definition of consciousness and its features

---

After all that has been said in the previous chapters, it may be the right time to ask what Searle actually means when he refers to *consciousness* and *mental states*. In this chapter we take a closer look at Searle's definition of consciousness and its intrinsic features. Searle describes consciousness as follows: "When I wake up from a dreamless sleep, I enter a state of consciousness, a state that continues as long as I am awake. When I go to sleep or am put under general anesthetic or die, my conscious states cease."<sup>17</sup> It is evident that his description of consciousness in terms of *conscious states* is meant to fit his proposition for an understanding of consciousness as a *state* of our neural network. Like William James, Searle distinguishes several structural features of consciousness that constitute its phenomenal nature. Although he recognizes about a dozen of them, he considers subjectivity, qualia and intentionality to be the essential ones. "The essential features of consciousness, in all its forms, are its inner, qualitative, and subjective nature."<sup>18</sup>

---

17 Searle, *The Rediscovery of the Mind*, p.83

18 John Searle, *Mind, Language and Society: Philosophy in the Real World* (W&N, London, 1999) p.40

## |Qualia

Searle describes qualia as follows: “Conscious states are *qualitative* in the sense that for each conscious state there is a certain way that it feels, there is a certain qualitative character to it.”<sup>19</sup> Similar to Nagel, Searle argues that every experience has its specific quality<sup>20</sup>. That is to say, it feels like something to *me* to swim in a sea or climb a rock as opposed to how it feels like (or rather must feel) to *you*. There is always something like to be in certain mental state. This qualitative character of our conscious experience is also referred to as *sensory modalities*. These sensory modalities include vision, sound, taste and smell and

---

19 Searle, *Mind, Language and Society*, p.42

20 British philosopher Peter Hacker, however, questioned this notion of to-be-like-something. He argues that the notion makes sense only if there is also a notion of to-be-not-like-something something or if one is capable of actually becoming something *else*. Thus Nagel's 'there is something like' to be human being or 'something like to be me for me' make no sense since there is nothing *else* opposed to it. While there may be one possible case – ceasing to exist – it most likely has no qualitative feeling to it, since death is defined as the termination of all biological functions. It follows that the statement 'there is something like to be me for me' is devoid of any meaning, since I cannot be something or someone else. Although in terms of the pure logic, the statement is right - I cannot be someone else thus there *must* be something like to be me for me - it still lacks the content without the opposition. The same goes for statements like 'there is something like to have experience for the subject of that experience' since there is no such case of having no experience. Moreover, “something“ by definition has no specific meaning and therefore depends on the presumption that everybody *knows* what this something is. (P.M.S.Hacker, Is there something like to be a bat? *Philosophy* 77 (2002) p.157-174)

qualities of sensation within those modalities - red vs blue, sour vs sweet - and it is this *quality* that makes them so difficult to deal with.<sup>21</sup> The problem is, as Daniel Dennett puts it, that qualia are *ineffable, unmediated, unanalyzable* and have *no interpersonal comparisons*. That is to say, we cannot really describe or analyze what it is like to see red or compare someone's experience of seeing red with someone's else experience of seeing red. Moreover, the basic qualities of sensory experience are directly apprehended that is, they are experienced as what they are, without any judgment. In addition, the existence of qualia is the best example of the subject-dependent feature of our consciousness.<sup>22</sup>

For many, it is qualia that pose one of the biggest challenges in philosophy of mind. To echo the question that defines the modern mind-body problem, how can something that is exclusively accessible to an experiencing subject be investigated by means of standard scientific methodology?

---

21 Knut Nordby, the famous color-blind scientist, once said in an interview that although he has acquired thorough theoretical knowledge of the physics of colors and the physiology of the color receptor mechanisms, nothing of that could help him to understand the phenomenal nature of colors.

22 Already Galileo Galilei described the dependence of qualia on the experiencing subject: "I think that tastes, odors, colors, and so on are no more than mere names so far as the object in which we locate them are concerned, and that they reside in consciousness. Hence if the living creature were removed, all these qualities would be wiped away and annihilated". (Galileo Galilei, *The Assayer*, 1623, p.274)

## |Intentionality

Like Brentano and James, Searle understands intentionality as the property of mental states to either represent or to refer to states of affairs. "Intentionality, to repeat, is the general term for all the various forms by which the mind can be directed at, or be about, or of, objects and states of affairs in the world".<sup>23</sup> All conscious states are thus about being conscious of something as such-and-such, they always have content. Similar to Kant, Searle argues that in order to make sense of the *of* of something, consciousness must involve organization and categories that have to exist prior to any conscious experience. These categories enable us to perceive things as something. Consciousness is therefore the fundamental condition for having genuine intentionality.

However, Searle points out that there has been a systematic effort in cognitive science and linguistics to separate intentionality and consciousness and treat intentionality as independent of consciousness. In functionalist theory for instance, intentionality has no qualitative content that matters to the functions of any intentional states such as desires or beliefs. Searle distinguishes between intrinsic intentional states (I am thirsty) and ascribed *as-if* intentional states (the flower must *be* thirsty or 'the program *wants* to launch a sequence). The former, intrinsic intentionality is what we and maybe higher animals have as a part of our biological make-up and the latter, as-if intentionality is only a matter of figurative speaking. By intrinsic, Searle means something that is genuine and essential as opposed to something that is derived from

---

<sup>23</sup> Searle, *Mind, Language and Society*, p.85

something or appears to be something. The way we often refer to computers or machines, for instance, would suggest “they” really *think* this or really *want* that, as if they had some intrinsic intentional states such as beliefs or desires. In other words, unlike intrinsic intentional states and their mental content such as my belief that Dallas is a great city, the attribution of intentional states to some other unconscious system (as-if) does not imply the presence of any mental states. Even if the computer program, based on certain parameters, appeared to believe that Dallas is a great city, the belief would inevitably lack a mental content. As-if intentionality is solely derived from the intentionality of the conscious human agents and is merely attributed to the unconscious systems. Without this distinction, Searle says, anything in the universe could be *mental*, i.e. having intentional states.

### **|Subjectivity**

All of our conscious mental states, without an exception, have one defining feature: *subjectivity*. Subjectivity is a mark of the mental in the same way as intentionality. However, unlike intentionality that can be in principle attributed to unconscious systems, such as computers in *as-if* fashion, subjectivity is the feature of our consciousness that as far as we know, no other natural phenomena possess. However, it is not the kind of subjectivity we normally use in connection to the particular judgement we talk about here. Searle talks about subjectivity in the sense of an “ontological category” or “mode of existence“. This mode of existence is always “first person“. That means that every conscious

state belongs to somebody, it is *somebody's* state. In that sense, neither *you* nor the others have conscious mental states observable from the third-person point of view. We cannot observe others' "subjectivities", nor can our subjectivity become an object of introspection. He argues that if we attempt to observe the consciousness of others we inevitably end up observing nothing else than their behavior (which we can, at best, only presume is causally related to their conscious mental states). And so far as introspection goes, Searle finds no distinction between an observer and the thing observed. On his view, whatever is being observed is always a part of *self-observation*. On that ground, Searle rejects the notion of "privileged access" since it relies on the possibility to make the distinction between the observer and the observed.

[The main points of his position on the problem of consciousness, Searle summarizes as follows:

"1| Consciousness consists of inner, qualitative, subjective states and processes. It has therefore a first-person ontology.

2| Because it has a first-person ontology, consciousness cannot be reduced to third-person phenomena in the way that is typical of other natural phenomena such as heat, liquidity, or solidity.

3| Consciousness is, above all, a biological phenomenon. Conscious processes are biological processes.

4| Consciousness processes are caused by lower-level neuronal processes in the brain.



5| Consciousness consists of higher-level processes realised in the structure of the brain.

6| There is, as far as we know, no reason in principle why we could not build an artificial brain that also causes and realises consciousness.“<sup>24</sup>

---

24 Searle, *Mind, Language and Society*, p.53

## Chapter VIII

### Functionalism, Cognitivism, AI and thinking machines

---

If there is something Searle is particularly known for besides his proposed solution to the mind-body problem, it is his criticism of functionalist models of mind, cognitivism and artificial intelligence. Let us start with functionalism.

In certain research fields, functionalism is today widely considered orthodox. In short, the theory states that there is nothing special about the mind because all the mental phenomena are simply a result of right inputs and outputs of the system. Mental processes are treated as mere functional states of a system and characterized by their causal role (that is, what they do as opposed to what they are made of). The characteristic feature of functionalist theory is thus the concept of multiple realizability – that the same mental states can be realized in different physical systems. As we know, many functional states, such as hunger or thirst, are being realized in different nervous systems with the same causal roles across many different species. The specific physical realization is thus causally irrelevant. The mind just happened to be “implemented” in the brain, but in theory, it could be any other system, made of whatever so long as it has the right causal properties. For example, a digital computer could in principle have a “mind” or

consciousness if running the “right” program with the right causal relations. There is no need for reduction nor identity because there is no distinction between mental and physical - the function is all that matters. The brain is basically a deterministic formal device that carries out symbol manipulations in the same way a computer does.

Searle refutes functionalist theory for many reasons, like its disregard for the brain as an accidental hardware (the physical realization) that in principle does not matter or its treatment of mental states as merely functional states of the system.

As far as cognitivism is concerned, Searle rigorously disputes its embedded presumption that at some level of description the brain processes are syntactical, (Fodor's language of thoughts),<sup>25</sup> and hence potentially computational.<sup>26</sup> He argues that syntactical ascriptions are

---

25 Fodor's idea of LOT (language of thoughts) describes mental representations as symbol structures. Fodor suggested that mental processes are simply symbol manipulation according to formal syntactical rules (algorithms). Thoughts are strings of symbols and thus have syntax similar to the rules of grammar in language. Intentionality, semantics and rationality are realized in a formal syntactic system. “Thinking is computing“, says Fodor. This view is closely related to Chomsky's theory of *universal grammar* (the idea that the knowledge of the general rules of syntax determining the construction of sentences is hard-wired into the brain).

26 In addition, all computational states are in cognitive science inherently unconscious processes (in the sense of *information-processing*) and as such are inaccessible to our consciousness - even all forms of intentional mental states. However, Searle argues that our idea of unconsciousness is solely derived from our notion of consciousness and “that we understand the notion of an unconscious mental state only as a possible content of consciousness“. (*The Rediscovery of the*

observer-relative and as such depend on an interpretation from outside. That is why any physical object or event can be in principle described in terms of syntax. Given that syntax cannot be intrinsic to a physical system by any definition (as opposed to density, for instance, that is intrinsic to any physical system), “computational states are not discovered within the physics, they are *assigned* to the physics.”<sup>27</sup> Therefore it follows, Searle points out, that computation is an observer-relative feature of reality and as such cannot have causal power over any physical system (only the implementing medium). He also believes that cognitivist computational metaphor inherently contains the homunculus fallacy because if computational states are observer dependent and the brain is a digital computer, somebody or something has to operate on those binary symbols or any other symbols inside of our mind. In the case of digital computers, there is no problem because we, the users, are homunculi that assign computational states to the system. But the brain is obviously a different case since there is

---

*Mind*, p.155) That means that if we call something an intentional state it must be in principle accessible to our consciousness. For instance, while we are asleep we still have a number of beliefs that, although unconscious at that moment, are in principle accessible to our consciousness. When we are unconscious, says Searle, there is neither information-processing nor unconscious intentionality going on in the brain - there are only neurophysiological processes. Period. Those processes are either mental or not and those that are, are potentially conscious. Searle thus rejects the idea of unconscious syntactic rules that are in principle inaccessible to our consciousness.

<sup>27</sup> Searle, *The Rediscovery of the Mind*, p.210

no external agent that takes care of attaching the computational interpretation to the patterns of neuron interactions, to the physical events that take places in our neural network.

According to Searle, somebody has to do the job, if reasoning is seen as symbol manipulation according to a certain algorithm that needs an operator. It does not take care of itself. Without this homunculus, the neurophysiological operations of the brain remain brute, non conscious biological activity where no “implemented algorithms” are to be found. In Searle's opinion, we should focus on figuring out how the human mind *really* works rather than trying to find out how the human mind resembles computer software.

The theoretical assumption of cognitivism, that at some level of description the human mind is a result of running complex syntactical operations, is widely adopted in the research field of artificial intelligence. Despite the refusal of this assumption, Searle does not reject the research program of artificial intelligence as such but instead argues that it is important to recognize the implications such research may have in our study of the mind. He distinguishes between so-called “weak” AI and “strong” AI. Weak AI, as Searle understands it, claims that the brain processes and mental processes can be simulated computationally, whereas “strong AI” claims that machines that act intelligently literally have a mind. While Searle readily agrees that simulation is indeed possible (although reminding of the fact that a computer simulation of rain does not make one wet), he strongly rejects the idea of attributing mind to a computer solely on a basis of

displaying intelligent behavior. At best, the system can behave as *if* it has conscious states but can never really have them without the causal duplication of the neurophysiological structure of the brain. In his view, exhibiting intelligent behavior is not a sufficient constituent for “having the mind” because only the existence of “intrinsic mental content” can be.<sup>28</sup>

In strong AI theory, the mind is seen as just a state that necessarily emerges when systems reach the kind of functional complexity needed to sustain complex behavior. The question is whether meeting certain computational specifications is enough to have mental states such as feelings, thoughts and understanding. Searle finds that possible only if

---

28 Searle also extends the objections against the strong AI and applies it to the problem of “other minds”. As mentioned above, he insists that behavior itself is not sufficient proof of having mental states because if we took strong AI claims seriously any machine that exhibits a certain level of intelligent behavior would literally be conscious. No machine rises to that level yet but in theory it can. Searle challenges this conception of the mental as having some essential connection to behavior. In one of his thought experiments, Searle invites the reader to imagine the situation in which our brain is entirely replaced by silicon chips. Although the silicon chips preserve all the mental states and consciousness, our external behavior is reduced “to total paralysis” and we end up being imprisoned in the paralyzed body. The moral of his thought experiment is to show that behavior is not ultimately relevant to the existence of mental phenomena, much less a proof. Searle suggests that the only indication that others have a mind as well is physiology because as far as anybody can tell, the brains are much alike, so there is no reason to assume that one brain would produce conscious mental states and the other, though all its regions are functioning normally, would not.

we can duplicate causal powers of the brain. If his theory of biological naturalism is right, consciousness then naturally emerges.

Thus, in principle, Searle acknowledges the possibility of creating consciousness artificially (although in very different way than most AI researchers). However, to demonstrate that mere symbol manipulation is not sufficient for genuine understanding, Searle formulated the famous “Chinese room“ thought experiment, that is intended to prove that even if the program is able to simulate intelligent behavior, in this case understanding Chinese, it still does not *literally* understand Chinese. Therefore, running the appropriate program cannot be, by definition, a sufficient condition for having a mind.

### **|The Chinese room argument**

The thought experiment now generally known as the Chinese Room Argument was first published in the paper “Minds, Brains and Programs” in the journal *The Behavioral and Brain Sciences* in 1980 as a response to claims of strong AI proponents that an appropriately programmed computer could have a mind and be *literally* said to understand semantics. In the paper, Searle imagines himself alone in a locked room full of boxes of Chinese symbols and with a book of instructions for manipulating the symbols (the program) written in English. He speaks no Chinese and has no understanding of that language whatsoever. People outside the room send in other Chinese symbols which, unknown to him are questions in Chinese (the input) and he, simply following a book of instructions (computer program) for

responding to Chinese characters, slips Chinese characters back under the door. Searle understands nothing of the symbols, and yet, by following the program for manipulating symbols and numerals just as a computer does, produces appropriate strings of Chinese characters that fool those standing outside into thinking that there is a genuine Chinese speaker in the room. Searle's conclusion of the argument is that implementing a program in a digital computer can make the computer look like it understands the language but in reality does not produce genuine understanding.

Searle argues that the thought experiment underscores the fact that computers merely use syntactic rules to manipulate symbol strings, but have no understanding of meaning or semantics. Therefore, Searle concludes that the "Turing Test" is inadequate because although the program is giving the right answers to the questions it does not have the real understanding of the questions it receives nor the answers it gives. The output that seems to be based on genuine understanding is nothing but simulation of such understanding. In other words, if Searle or any other man in the room does not understand Chinese on the basis of implementing the appropriate program for understanding Chinese, then neither does any other digital computer solely on that basis because no computer, because it is a computer, has anything the man does not have. Computation, says Searle, is defined purely formally and syntactically and therefore has no actual semantic content



and therefore “syntax is not sufficient for semantics”.<sup>29</sup>

The Chinese room argument has received dozens of critical replies over the years (the Robot Reply, the Brain Simulator Reply, etc.) and among the ones that have been discussed the most is “the Systems Reply”. The Systems Reply argument basically points out that the hu-

---

29 The Chinese room argument has its important antecedents. The first is known as “Leibnitz's Mill”. Like in Searle's argument, Leibnitz (1646–1716) asks the reader to imagine a machine that behaves in such a way that would trick him into believing that it actually thinks and has experience. The argument goes as follows: “Moreover, it must be confessed that perception and that which depends upon it are inexplicable on mechanical grounds, that is to say, by means of figures and motions. And supposing there were a machine, so constructed as to think, feel, and have perception, it might be conceived as increased in size, while keeping the same proportions, so that one might go into it as into a mill. That being so, we should, on examining its interior, find only parts which work one upon another, and never anything by which to explain a perception. Thus it is in a simple substance, and not in a compound or in a machine, that perception must be sought for.” (section 17, *Monadology*) Leibnitz points out that the internal mechanical operations in that machine are just parts moving from point to point, and thus there is nothing that is conscious or that could explain thinking, feeling or perceiving. For Leibniz, physical states of whatever complexity are not sufficient for constituting mental states.

A second antecedent to the Chinese room argument is the idea of a paper machine implemented by a human. The author of this idea is Alan Turing. Turing, in the paper “Intelligent Machinery” (1948) imagines a program for a “paper machine” to play chess. The paper machine has a program, a series of simple steps like a computer program, written in natural language and followed by a human. The human operator of the paper chess-playing machine does not need to know how to play

man in the room (CPU) may not have the understanding, but the *whole* system may have it. Thus, the capacity to understand should not be constrained by the limits of the processing unit because the larger system would indeed contain a huge database, memory, etc. Yet, Searle argues that even if the human in the room memorizes the whole book and the stacks of paper (the database, memory), it still does not make any difference in terms of him *understanding* Chinese. If neither the human nor any other single element in the room has any understanding of Chinese, then neither does the whole room, which is made of those elements. However, the proponents of the SP argue that Searle's failure to understand Chinese is irrelevant: he is only the implementer. The larger system implemented would indeed understand.

---

chess. All the operator does is follow the instructions for generating moves on the chess board. In fact, the operator does not even have to know that he or she is playing chess—the input and output strings, such as “N-QB7” could mean nothing to the operator of the paper machine. Turing was quite optimistic that computers would soon be able to exhibit intelligent behavior like answering questions and carrying on conversations. In 1950, Turing proposed what is now known as the Turing Test. The test is designed to show the extent to which a computer can pass for *human* in an “on-line” chat and thus be considered intelligent.

## Chapter IX

### What is wrong with Searle's philosophy of mind?

---

#### **|The fallacy of *the solution***

It is obvious that theories as bold as Searle's have its weak spots, and his proposed solution to the mind-body problem, *biological naturalism*, is no different.

Searle's proposition that the mind can be caused by the brain and be a feature of the brain at the same time may appear rather incoherent since it seems that something can be identical with something else, or be caused by something else, but not both. To return to Searle's metaphor, how can "The liquid state of this water...be causally dependent on the behavior of the molecules, and...also be a feature of the system made up of the molecules'?"<sup>30</sup> If Searle's H<sub>2</sub>O-liquidity example was a genuine instance of causation and identity at the same time, his case would be probably strong. However, he is simply wrong to think that there are any examples of simultaneous causation and identity. The liquidity is not caused by H<sub>2</sub>O molecules but is simply one of their features.

The proof of this can be found in Searle's own argument, in the so-called "different levels of description". As Searle has pointed out

---

<sup>30</sup> Searle, *The Rediscovery of the Mind*, p.252

himself, we must be careful to distinguish those attributes that are intrinsic physics from those that are observer-relative on any level of description. As he explains, the expressions “mass“, “gravitational attraction“ and “molecule“ name features of the world that are intrinsic. If all observers ceased to exist, the world would still contain mass, gravitational attraction, and molecules. “Expressions such as 'bathtub,' and 'chair' do not name intrinsic features of reality. Rather, they name objects by specifying some feature that has been assigned to them, some feature relative to observers and users.”<sup>31</sup> But what kind of attribute are these levels of description? It is obvious that there are not any *levels* in a thing apart from observers since the level of description is not intrinsic to any physical system. Strictly speaking, there are as many levels as there are ways to observe something, and if there were no observers, there would be no “levels“. Therefore, the levels of description are observer-relative. Now, it makes no sense to say that observer-relative attributes cause anything; these attributes are not in the things observed at all, but are rather assigned to the world by observers. So how could the levels be causally interacting when they are not intrinsic to the world at all? We cannot create new causal relations just by looking at one and the same object from different perspectives.

Ironically, Searle made a similar case against computation when he argued that computational states cannot have causal power over any physical system since they are not intrinsic to physics. It almost seems

---

31 Searle, *The Rediscovery of the Mind*, p.211

like Searle does not pay enough attention to his own notion of intrinsic vs. observer-relative features. If he did, he would most likely realize that the different levels of description must be observer-relative, not intrinsic; hence, different-level features of one thing cannot interact or be in causal relation. It follows that something cannot be both caused by X and identical with X. Thus, liquidity as a feature of H<sub>2</sub>O molecules cannot really cause anything. The point is that although all of the features that we perceive, at higher and lower levels alike, are intrinsic and have causal powers, things do not acquire or lose intrinsic attributes when we look at them on different levels (although it may be sometimes useful to discern certain properties on a particular level). This is because there is no intrinsic feature of a thing that demands that it be split up into any particular levels. Instead, the levels get attributed by observers, i.e. they are observer-relative. And these observer-relative features, by definition, cannot have any causal powers since in a fundamental sense they are not in the things at all.

Let us take the example of water. Water objectively exists, no matter what level of description we choose. We can look at it on the micro-level and on the macro-level. But that does not mean that the levels are causing anything. It only means that we have two different perspectives on one and the same thing. Now, we may loosely talk about H<sub>2</sub>O molecules “causing” the behavior of water, but they *really* do nothing of the kind. Learning about the molecules may help us understand why the behavior of the macro-level is the way that it is, but the relationship is identity and not causation. To deny this is to accept the absurd view

that we could literally create new causal relations in the world simply by observing them from additional perspectives. And the same goes for the mind and the brain. If the mind is merely a feature of the brain, the two cannot causally relate, and if they do causally relate, they cannot be identical. Searle's dilemma thus remains unscathed because we either end up in Epiphenomenalism where the mind has no causal power over the brain or dualism where we have no account for the causal relation at all. Searle's solution to the mind-body problem simply does not work the way it might at first look. However, his dilemma may just be a part of a bigger conceptual problem, and no one should be blamed for trying to find a solution.

According to Thomas Nagel and Colin McGinn, the problem with any solution to the mind-body problem is that we lack the requisite conceptual apparatus to even begin to resolve the problem.<sup>32</sup> Nagel,

---

32 According to British philosopher Colin McGinn (called by Owen Flanagan the proponent of “new mysterianism“ or “anti-constructive naturalism“), we must accept the fact that our cognitive limitations prevent us from ever come to an understanding of the relation between the brain and our consciousness. McGinn claims that neither perception nor introspection alone will ever enable us to witness the dependence between qualitative mental states of our consciousness and the brain's neural activity. Although he indeed presumes the problem must have some sort of naturalistic explanation, he is very skeptical about ever finding one. He introduces his idea of so-called *cognitive closure* as follows: “Our concepts of consciousness are inherently constrained by our own form of consciousness, so that any theory the understanding of which required us to transcend these constraints would *ipso facto* be inaccessible to us.” (p. 356) McGinn further argues that it does not even matter how advanced our descendants may become because “so long as they (will) form

for instance, argues that since causal explanations in science are necessary, given a certain theory, the observed effects must follow. Thus given the molecular composition of H<sub>2</sub>O, its features like solidity, liquidity, etc. has to be strictly deducible. But the problem with consciousness is that no necessary connection exists between the physical and the mental. No matter how much we know about the brain, we can never deduce a single mental predicate. Even if we knew everything about the physical world of molecules or forces, we would

---

their empirical concepts on the basis of perception and introspection they will face the same problem.” (p. 361) Until all concept formation is tied to perception and introspection, however loosely, no mind will be capable of understanding how it relates to its own body. States of our consciousness are simply not potential objects of perception, just as neural activity cannot be an object of introspection. The problem is that every scientific method depends on methods of measurements, but what is measurable depends on scientific method. Perhaps only a higher form of life that has some sort of independence of perception and introspection may find the answer. McGinn believes the mind-body problem requires obtaining data in some other way than through perception and introspection. Moreover, we can hardly think of conscious processes in the spatial terms that apply to the brain. Here, McGinn strongly opposes Searle's spatial metaphors of liquidity or solidity by saying: “the case is exactly unlike (say) the dependence of liquidity on the properties of molecules, since here we do think of both terms of the relation as spatial in character; so we can simply employ the idea of spatial composition.” (p.363) Like Thomas Nagel, McGinn concludes that as of now, the problem has no solution, no matter how advanced neuroscience may become. There is no philosophical problem distinct from a scientific one. As it is now, the solution to the mind-body problem is cognitively closed. (Colin McGinn, “Can we solve the Mind- Body Problem?” *Mind*, New Series, Volume 98, Issue 391, 1989)

not be able to predict the most obvious mental fact unless we slip a mental premise into the argument. To put it simply, we cannot get out of the subjectivity of our consciousness to see its necessary relation to its material basis, and Searle's attempt to get around the problem by redefining mental phenomena as *property* of their underlying neurophysiology does not remove this conceptual obstacle. It also seems that when Searle says that the mental is also physical, it stretches the meanings of mental and physical completely out of shape, depriving them both any of content.

### **|Inconsistent criticism and the accusation of property dualism**

As we can see in the previous paragraphs, Searle's theories and criticism are oftentimes inconsistent, and his disparaging view on dualism is a fine example of it. He rejects the argument of dualists who insist that just because we cannot account for causal interaction between the mental phenomena and the physical states, does not mean there is none. But at the same time, he concedes that there are many brute facts in the world, such as gravity, which we accept as real even though we have no explanation of how they work. It may very well be that the causal interaction of the mind and the brain is one of those brute facts of our reality for which we have yet no explanation. Since no one has presented any plausible *scientific* theory about their interaction, and since it is logically impossible to deduce a mental



statement from a non-mental statement, the dualist mind-brain interaction is not out of the question.

Ironically enough, though, Searle has been accused of being a dualist himself many times. To many of his critics such as Nagel and Chalmers, biological naturalism seems to be little more than property dualism in disguise. The reason why it sounds like property dualism is that there is a significant difference, by Searle's own admission, between liquidity and other *objective* phenomena on the one hand and consciousness and other *subjective* mental phenomena on the other. The problem is that consciousness is, unlike liquidity, not identical nor ontologically reducible to the microphysical structures which cause it. If the physical processes which cause consciousness are objective third-person phenomena, and consciousness and other mental phenomena are subjective first-person phenomena, it is reasonable to describe the latter as being of a *fundamentally different* kind than the former. That is to say that they exist in the universe of dual properties. Moreover, what is the point of appealing to liquidity when Searle himself concedes that such examples fail to be analogous to consciousness whose ontological irreducibility distinguishes it from features like liquidity?<sup>33</sup> The main problem with this move is that it ignores the fact that the distinction between mental and physical is obviously not merely a matter of arbitrary

---

33 Searle writes: "When we come to consciousness, we cannot perform the ontological reduction...a perfect science of the brain would still not lead to an ontological reduction of consciousness in the way that our present science can reduce heat, solidity, color, or sound." (Searle, *The Rediscovery of the Mind*, p.116)

definition. As we have seen, mental phenomena are in Searle's view uniquely subjective and therefore uniquely ontologically irreducible. That is the reason they are contrasted with physical phenomena. It just does not sound right that Descartes or anyone else decided one day to arbitrarily define "mental" to mean "non-physical", and then concluded, trivially, that some form of dualism must be true. Moreover, consciousness does not *emerge* from the brain nor the brain *gives rise* to consciousness.<sup>34</sup> And if it does, it is difficult to imagine how does that not imply that consciousness is "over and above" the brain - claim of property dualism - in one way or another.

## **|Free will**

Another example of Searle's inconsistency is the problem of free will. There is no need to show that his view leads inevitably to the denial of free will. He freely admits it. Since Searle has said that the mind is merely a higher-level feature of the brain and the brain is made up of neurons, and neurons certainly do not have any free will, the mind has no free will either. "All of our mental states are caused by neurobiological processes in the brain, and they are themselves realized in the brain as its

---

34 Searle uses the term *give rise* in relation to consciousness many times. Here is one example: "The relation is this: brain states that are nonconscious can be understood as *mental* states only to the extent that we understand them as capable, in principle, of *giving rise* to conscious states." (Searle, *Mind, Language and Society*, p. 86)

higher level or system features."<sup>35</sup> Searle's bottom-up model of explanation simply inevitably leads to determinism. For if the brain is deterministic, and the mind is a feature of the brain, then the mind must be deterministic as well because we have to assume that there are causal antecedents to our actions. If there is a cause, the event *has to* happen. To him, it just seems incoherent to think that the higher-level could in turn cause the lower-level. However, unlike other philosophers who intellectually accept determinism, Searle admits that free will seems to be an obvious fact. After all he has said, he continues to act on the assumption of free will. Although he cannot see how free will can be consistent with everything we know about the world he continues to believe that he has free will all the same.

---

35 John Searle, *Freedom and neurobiology: Reflections on free will, language, and Political Power*, (New York: Columbia University Press, 2007) p.40

## Conclusion

---

It is a relief to read that Searle, though he takes his biological naturalism to have solved the mind-body problem, also acknowledges that his position “raises a thousand questions of its own... [such as] how exactly...the elements of the neuroanatomy produce mental phenomena?”<sup>36</sup> and admits that “we don’t have anything like a clear idea of how brain processes, which are publicly observable, objective phenomena, could cause anything as peculiar as inner, qualitative states of awareness or sentience, states which are in some sense ‘private’ to the possessor of the state”<sup>37</sup> It is hard to see exactly, however, what problem Searle thinks he has solved since explaining that we do not have the slightest idea how the brain causes consciousness is what most philosophers *mean* by “the mind-body problem”. Certainly, it is no revelation to be told that brain processes cause mental processes. In reality, Searle’s solution appears to simply redefine the mind-body problem in such a way that the difficulty of fitting what is *irreducibly* subjective into the objective physical world is re-classified as a problem for neurobiology. However, there is no reason to believe that the methods of neurobiolo-

---

36 Searle, *The Rediscovery of the mind* p. 1

37 John Searle, *The Mystery of Consciousness*. New York: The New York Review of Books, 1997, p.8

gy are any more likely to be able to deal with the objective-subjective divide than those of philosophy. In fact, given the inherently *conceptual* nature of the problem, they are certainly less likely to be able to do so. For all of his skillful critiques of materialism, dualism and AI, Searle's philosophy of mind is troubled by many obvious misconceptions and paradoxes. His argumentation is oftentimes vague and not well-founded, and when faced with criticism, leans on intuition or appeal to common sense. And although Searle's biological naturalism frames the brain-consciousness problem in somehow useful manner, it does not solve it for good. As the history and the present has shown us, there is no single solution to the mind-body problem. But all the same, John Searle has already checked off the mind-body problem on his "to-be solved" list as *solved*.

## Appendix

### Yet another *naturalistic* theory of mind

---

Contemporary philosophical discourse is very complex and includes as many theories of consciousness as there are scientific disciplines. Cognitive scientists, neuroscientists, AI researchers etc. each postulate theories or have a set of presumptions that answer a certain set of questions in their particular framework. This framework, however, is usually confined to a rather specific area of research, and the theories thus often contain, age-old conceptual problems and fallacies. Despite this fact, many of these theories introduce some new useful distinctions and are relevant to the discussion. However, I will limit myself, for the purpose of this paper, to only one other theory that I find to offer an interesting, different way of looking at consciousness.

#### **|David Chalmers**

Australian philosopher David Chalmers is a well-known figure in philosophy of mind for coining the term the “Hard Problem” of consciousness. He argues that although science could probably solve many of the “easy” problems of consciousness, such as the neural correlates of conscious phenomena, it can never solve the “hard” problem, i.e. the problem of explaining how any physical system could have or produce any qualitative states. Chalmers believes that the *hard* problem cannot

be solved by reductive explanation because subjective experience or qualia cannot be explained by referring to their underlying physical mechanisms. Thus, instead of reduction, they must be taken as *fundamental* features of the universe. Those features are fundamental in a sense that they cannot be explained by further reduction because they do not seem to be dependent on anything *more* fundamental (like mass, gravity, etc.).

According to Chalmers, a theory of consciousness thus should be more like a theory in fundamental physics than like a theory in the biological sciences. However, since no theory of consciousness can tell us what consciousness is or why it exists, it should at least present to us how consciousness is connected to other phenomena in the world - for example, our brains. Chalmers himself has described his theory as *naturalistic dualism* (a combination of functionalism and property dualism). It is a type of dualism that he believes may be much easier to incorporate into the modern scientific view of the world than any other dualistic theory. As a variety of dualism, it postulates that consciousness is a fundamental but nonphysical feature of the universe, coupled to the physical through so-called "fundamental psychophysical laws". Those laws (part of fundamental laws of nature) are what determine how phenomenal experience correlates with physical features (such as information-processing in our brain).

In Chalmers' view, phenomenal experience itself is not a necessary part of the physical world. He defends this idea through a thought experiment that presents a world in which, although it is physically

identical to ours, no one is conscious; everyone is a neurobiological zombie. Yet, our counterparts in that world have physically and cognitively identical brains with us – it is just that the psychophysical laws of nature do not control that world as they do in ours. Chalmers claims that this zombie thought experiment is not logically incoherent or inconceivable and thus consciousness cannot be a logically necessary part of the physical. We can imagine all of the objective physical, biological or cognitive components of the world without any of the phenomenal experience. It follows, however, that in Chalmers' theory, phenomenal consciousness, since detached from the physical reality in this fashion, has no causal powers over physical phenomena. In other words, phenomenal consciousness is rendered causally ineffective and thus cannot make any difference in the physical world (epiphenomenalism).

The theory also implies the idea of panpsychism. Chalmers argues that phenomenality may exist, at least in some extremely simple form, in all information-bearing events in the universe, since phenomenal properties are the internal aspect of information.

The only difference is that in the human brain, information is organized in more complex ways than in perhaps any other system. And from this information processing and functional organization emerges human consciousness. However, the consequence of relating phenomenality and information is that every simple physical systems, such as thermostats or pocket calculators, must also have some sort of simple consciousness. As with all panpsychist theories, Chalmers' theory sees



consciousness everywhere around us. Subjective experience is distributed widely in the universe, anywhere you look around, there are physical systems carrying information and thereby having conscious experiences.



## Bibliography

- Searle, J. *Freedom and neurobiology: Reflections on free will, language, and Political Power*. New York: Columbia University Press, 2007
- Searle, J. *The Mystery of Consciousness* (review collection). The New York Review of Books, 1997
- Searle, J. *The Rediscovery of the Mind*. Cambridge, MA: MIT Press, 1992
- Searle, J. *Minds, Brains and Science*. Cambridge, MA: Harvard University Press, 1984
- Searle, J. *Intentionality: An Essay in the Philosophy of Mind*. Cambridge University Press, 1983.
- Searle, J. Minds, brains, and programs. *Behavioral and Brain Sciences* 3 (1980), 417-57.
- Searle, J. Why I am not a property dualist, *Journal of Consciousness Studies* 9 (2002), 57-64
- Dennett, D. *Consciousness explained*. Back Bay Books, 1992
- Flanagan, O. *Consciousness Reconsidered*. MIT Press, 1992
- Chalmers, D. *The Conscious mind: In Search of a Fundamental Theory*. Oxford University Press, 1996
- Farthing, G. *The Psychology of Consciousness*. Prentice Hall, 1992
- Revonsuo, A. *Consciousness: The science of subjectivity*. Psychology Press, 2009
- Nagel, T. *Other Minds*. New York: Oxford University Press, 1995

- Nagel, T. "What Is it Like to Be a Bat?", *Philosophical Review* (1974), 435–50
- Hacker, P.M.S. Is there something like to be a bat? *Philosophy* 77 (2002) 157-174
- McGinn, C. "Can we solve the Mind- Body Problem?" *Mind*, New Series, Volume 98, Issue 391 (1989), 349-366
- Seth, K. A. 2010. The grand challenge of consciousness, *Frontiers in Psychology* (2010)
- Block, N. Consciousness, Philosophical Issues about, *Encyclopedia of Cognitive Science*. Nature Publishing Group, 2003
- Kuhn, T. *The Structure of Scientific Revolutions*. University of Chicago Press, 1960
- Russell J. & Norvig P. *Artificial intelligence: A modern approach*. Pearson, 2013
- K. Popper, *The Logic of Scientific Discovery*, Psychology Press, 2002
- Munger, P.M. *The History of Psychology: Fundamental Questions*, Oxford University Press 2003