Anglia 2018; 136(1): 100–120 DE GRUYTER

Dan Rebellato*

Is the Theatre a Zombie? On the Successful Failures of Émile Zola

https://doi.org/10.1515/ang-2018-0011

Abstract: Naturalist theatre, in its late-nineteenth-century incarnation, and particularly in the work of Émile Zola, is often seen as advancing a physicalist view of the mind, where all mind states can be reduced to brain states. The novels and the plays do not uniformly or unambiguously support this analysis, so is the theory or the practice wrong? Physicalism is an idea that has had a recent renaissance, helped by the discoveries of neuroscience. Nevertheless I express some caution about the claims made for the eradication of free will. A range of thought experiments in the philosophy of mind have cast doubt on physicalism, culminating in David Chalmers's much-debated zombie argument. I argue that zombies and their analogues represented deep social anxieties in the late nineteenth century, and make repeated appearances in Naturalism. The essay goes on to suggest that Naturalism should be considered to have conducted thought experiments, rather than just to have attempted to embody the theory on stage. Turning to John Searle's 'Chinese Room' thought experiment, I suggest that theatre-making itself may be a kind of thought experiment model of the mind.

"The mind is a kind of theatre", wrote David Hume, "where several perceptions successively make their appearance; pass, re-pass, glide away, and mingle in an infinite variety of postures and situations" (2001: I, 165). That the mind might be understood as a sort of theatre is an idea that occurs repeatedly in philosophical and psychological literature; but what of its opposite? That the theatre might be a kind of mind is less well rehearsed. How consciousness might be manifested onstage has tended to be a concern primarily for playwrights and actors. It is there in Aristotle's *Poetics*, in the descriptions of tragedy as the imitation of an action, which in turn is defined by the kind of "character and reasoning" involved (1996: 11). But it has returned to philosophical and theatrical attention lately, in important contributions by William W. Demastes (2002), R. Darren Gobert (2013), Bruce McConachie (2013), among others. In all disciplines, the place of consciousness

***Corresponding author: Dan Rebellato,** Royal Holloway, University of London
E-Mail: D.Rebellato@rhul.ac.uk

has had paricular force over the last twenty years, in part because of work in neuroscience and the bold, even extravagant claims made by cognitive scientists in their quest for the neurological basis of what it is to be human; so it is unsurprising to find this flurry of interest in theatre.

A good, challenging starting point for anyone thinking about theatre and the representation of consciousness is the Naturalist movement of the last third of the nineteenth century, and, in particular, the work of Émile Zola. In his role as chief agitator for the Naturalist project, Zola was a prolific commentator on its theory and practice in the novel and in the theatre, and key to his conception of the movement was that it would be a quasi- (or perhaps not even all that quasi-) scientific cultural movement. Naturalist authors would take their cue from the scientific advances of the nineteenth century and seek to understand human beings in purely physical terms. In his famous preface to the second edition of *Thérèse Raquin*, he writes:

> Thérèse and Laurent are human animals nothing more. In these animals I set out to trace, step by step, the hidden workings of the passions, the urges of instinct, and the derangements of the brain which follow on from a nervous crisis. Love, for my two heroes, is the satisfaction of a physical need; the murder they commit is a consequence of their adultery, a consequence they accept as wolves accept the slaughter of sheep; and last, what I have been obliged to call their remorse consists simply in an organic disorder, the revolt of a nervous system stretched to breaking point. There is a total absence of soul. (Zola 1992: 1–2)

It is a deliberately challenging and disorientating presentation of the argument. In philosophical terms, we might describe it as a physicalist reduction: love is nothing more than physical need; murder is just the result of a causal sequence; remorse is but a nervous disorder. Human beings are merely animals (indeed, the French word he uses is 'brutes' not 'bêtes', more 'beasts' than 'animals'). The theological notion of the soul has been abolished, but with it too perhaps even the mind, certainly in the sense of a centre of conscious rational control. Of his characters he says, "I had only one aim [...] to seek within them the animal, and even to see in them only the animal" (Zola 1992: 2).

His determination to show that he is eradicating mind and soul is perhaps at work in his repeated comparison of his characters to corpses: "I simply carried out on two living bodies the same analytical examination that surgeons perform on corpses" (Zola 1992: 2) and in an essay submitted in 1866 to the 33rd Congrès Scientifique de France, held in Aix, he wrote of the modern (implicitly Naturalist) author:

> I like to think of him as an anatomist of the soul and the flesh. He dissects man, studies the play of his passions, explores each fibre, analyses the entire organism. Like a surgeon, he feels

> neither shame nor revulsion as he probes these human wounds. He cares only for the truth and lays before us the cadaver of our own heart. (Zola, qtd. in Becker 2002: 256, my translation)

Zola is not consistent in his use of these images. Within a few lines he describes his characters as both "corpses" and as "living flesh" (1992: 2). This strange and uncanny image vaguely suggests that his characters are not merely dead but *zombies*, a thought to which I shall return.

Nonetheless, the challenge that Zola seems to be offering is a literary and theatrical vision of a world without consciousness. In the novel, the characters are described in mentalist terms, having thoughts, hopes, wishes, but if we are to take Zola at his prefatory word, we are to see those as epiphenomenal, causally inefficient. The real work is biological and chemical. This is easy to say and less easy to apply. Before Thérèse and Laurent drown Camille, Laurent sees an opportunity to kill him by stamping on his head. Is it possible to think of mental experience as causally inefficient here?

> For a few seconds, Laurent's foot hovered above the sleeping Camille's face. Then, slowly, he brought his leg down and moved a few paces further away. He had realized that to murder him like that would be utterly stupid; if he stove in Camille's head, he would soon have the whole police force after him. He only wanted to get rid of him in order to marry Thérèse, and he intended to live it up after the crime. (Zola 1992: 60)

To me, this passage suggests that Laurent has a complex series of interlocking *intentions*:

1. To marry Thérèse and live it up
2. To murder Camille
3. To crush Camille's head

He *believes* that (1) is dependent on (2). He then *realizes* that (3) will serve (2) but will prevent (1) (because he will be arrested and unable to marry Thérèse and live it up). Consequently he engages in a bit of preference-ordering and *decides* to abandon (3). His actions are driven by intending, believing, realizing, and deciding and the passage therefore seems to suggest that mental contents and processes are causally efficient. Is there a contradiction between the novel and the preface?

When I first encountered Zola's work, as an undergraduate almost thirty years ago, the consensus seemed to be that his theoretical writings perhaps naively overrated the significance of the new discoveries in biology and experimental method he had absorbed from his reading of Claude Bernard, Bénédicte-August Morel, Prosper Lucas and others (Brown 1996: 187–188; Mitterand 1999: 719–727), but that his advocacy of a scientific theatre had generated a much more important

theatrical movement, which, in its theatrical complexity and richness, superseded and thankfully contradicted the limitations of the theory. But in the last twenty years, we have entered a new era of hyperscientism. Barely a month goes by without some report of a neuroscientist having discovered the pattern of neurons that cause love, or the gene for friendship, or that having put a poet in an fMRI scanner, we now know the part of the brain that lights up for prosody. Where Zola's valuation of science may once have seemed inflationary, now he might seem entirely consistent with the neuroscientific tenor of the age. Perhaps after all it is the plays and novels that sloppily fail to conform to the bracing rigour of the theory.

The Zola of my undergraduate years and the Zola of today both seem to involve a contradiction between theory and practice. In this essay, I want to suggest that this may not be the best way to understand Naturalism and that, in some ways, the failures of its theory or its practices, may be the source of its philosophical success. To make this argument, the philosophical resources I want to draw on are almost exclusively from the Anglo-American analytic and post-analytic tradition. There are two main reasons for this. First, I think the analytic tradition has produced more substantial and precise work on the nature of consciousness than the continental tradition (phenomenology notwithstanding). That may be a matter of taste, of course, but, second, it is evident that Émile Zola's intellectual commitments are precisely in the spirit of the early analytic movement: quasi-scientific, skeptical of metaphysics, drawn to the verifiable and empirical. While much of the work I use here is unconvinced by that scientism, it is at least engaged in the same debate. There is a case, of course, for reframing the debate entirely but my approach, I hope, has at least the virtue of helping to elucidate some of the complexities and fine distinctions within Zola's project, rather than dismissing it from outside. As Zola himself once wrote at the end of an essay responding to another attack: "I am still awaiting an adversary who will agree to meet me on my own ground and to engage me in combat with my own weapons" (1880: 296).[1] As will become clear, my attitude to Zola is more sympathetic and less combative than this.

There is a broader question, though, about whether philosophy is an appropriate method here at all. In the same essay, Zola implored his opponent not to "stray onto philosophical territory, which is very infirm underfoot; let's place ourselves instead on scientific territory [...] because here we have certainty" (1880: 292). A few years ago, I chaired an event designed to generate more scientifically-engaged writing for theatre, funded by the Wellcome Trust; the room was full of playwrights and neuroscientists. One of the first speakers was

1 This and all subsequent quotations from *Le Roman Experimental* are translated by me.

the distinguished philosopher of mind, Tim Crane. I was surprised that several of the neuroscientists were bewildered, even affronted, that they were being addressed by a philosopher, one of them explaining that the task of explaining the mind has passed decisively from philosophy to science, that the philosophy of mind is at best an irrelevance and at worst an impertinence.

Well, not so fast. I do not think it evident that philosophy should just get out of the way and leave the task of accounting for the mind to the scientists. On the contrary, I argue that philosophy has an important role in checking the validity of some statements by scientists, who I think have been wildly overconfident in extrapolating from no doubt major discoveries to unwarranted conclusions. As may be obvious, I am not wholly convinced by the neuroscientific picture, at least in the radical forms in which it is claimed to have found complete physical explanations for all mental phenomena, systematically reducing mind to brain, and perhaps eliminating talk of the mind altogether. As Thomas Nagel says, talking about the brain without mentioning mental contents is like writing about Picasso without mentioning the paintings (1995: 88). It is true that I am not a neuroscientist and do not have access to detailed knowledge of the brain, there are problems to which we all have access irreducibly associated with the kind of physicalist reduction that some neuroscientists like to espouse.

This is perhaps the moment to state clearly that what interests me about Naturalism is to see the ways that the faultlines and contradictions in the physicalist position might be seen playing themselves out in the plays and performance practices. It makes Naturalism, in my view, *more* interesting, not *less*, that its theory so incompletely manifests itself in performance and that, in the spirit of Naturalism's interest in scientific method, this experimental finding might be the point.

The Libet Experiment

Before I get to that, I want to demonstrate ways in which philosophical debate (and a piece of popular comic writing) can provide resources to critique 'hard science'. Some of the most widely-cited experiments in contemporary brain science are those carried out by Benjamin Libet, which, it is said, demonstrate the non-existence of free will (see Libet 1999: 47–58).

Libet attached electrodes to a series of subjects, placed them in a room, and asked them, at some moment of their own choosing, to flex their wrist. There was a large clock on the wall and they were asked to note exactly the moment at which the decision to flex their wrist was taken. At the same time, the electrode measured any electrical activity in those centres of the brain that control movements of the wrist.

What he and his team discovered was very surprising. They found that if the action of moving the wrist takes place at point 'S', in general the will to move was consciously experienced 200 milliseconds before 'S'. That seems fair enough. But the electrode picked up brain activity related to moving the wrist 350 milliseconds before that. In other words, the brain appears to have decided to move the wrist over half a second before the mind has. Or, as Libet puts it, "the brain was evidently beginning the volitional process in this voluntary act well before the activation of the muscle that produced the movement" (1999: 49).

This has radical and disturbing implications for our belief in free will. If we want to believe that our minds control our actions we will need to explain how it is that the action has already been decided by the brain.

This is not in itself decisive. We might think that 300 milliseconds is not very much. We might think that it offers a margin of error; is it not possible that there is a fractional lag between deciding to act and looking at the clock? Indeed, this argument has been made. However, Libet's electrodes have been superseded by the kind of inspection offered by fMRI machines. John-Dylan Haynes, a scientist working at the Bernstein Centre for Computational Neuroscience in Berlin, conducted a similar set of experiments, but, while Libet was registering activity in the supplementary motor area, which may well be rather late in the decision-making chain, they monitored activity in the prefrontal and parietal cortex. And their experiments suggest that the decision has generally been taken a full second before the decision becomes available to the conscious will – and sometimes as much as 10 seconds before (Soon et al. 2008a: 544). This would seem to remove the margin of error. Even more strikingly, Haynes asked his subjects to choose whether to move the right or left hand; different nerve centres in the brain operate the right and left hands, so the scientist is able to see clearly *what* decision has been made before the subject knows they have decided. The brain chooses before the mind. It would seem that free will must be an illusion.

Now, in fact, Libet and Haynes do not agree on this point. Libet believes that we do have free will, though it is a rather particular type of free will. He looks at the gap between the conscious decision and the action. He says that there are 50 milliseconds where the primary motor cortex has activated the spinal motor cells and this process cannot be stopped. However, he believes that in the 150ms before that, the will can intervene to veto the brain's decision. In other words, the brain might want to move the wrist, but we can intervene and stop that. In a sense, then, not moving the wrist at that moment is an outcome of a freely willed decision (Libet 1999: 51–53). It's not much but it's something.

But even so, do these experiments really demonstrate what they purport to demonstrate? I think not. In the Haynes case, observing patterns of activation in the prefrontal and parietal cortex is supposed to predict the decisions made. But

if we look in more detail we discover plenty of room for skepticism about the larger claim. First, this predictive method only gets it right around 55 % of the time (Soon et al. 2008b: 10). *Guessing* whether the subjects would move right or left would on average yield a 50 % success, so immediately that looks like a much less impressive success rate. Second, the '10 second' figure, which has been all over the popular science press (e.g. Smith 2011) is very misleading. In fact, the earlier the predictive indicator the less accurate it is (Soon et al. 2008a: 544). Third, in the Libet case, the statistical results are an average of 40 different experiments, so this image does not represent an actual model of a particular decision. It opens that experiment up to the distorting factors of outliers and subjects who have misunderstood the instructions or have misinterpreted what constitutes a decision. And finally, it seems that some scientists have found it difficult to duplicate these findings.

These are technical problems with the operation of the experiment and the framing of the results but there are more significant conceptual problems here. A very significant one is what Libet says he has detected. He describes the pattern of electrical activity in the supplementary motor area as 'readiness potential'. He says: "the brain process (RP) to prepare for this voluntary act began about 400 msec. before the appearance of the conscious will to act" (Libet 1999: 51). But this is illegitimate. He has not observed an unconscious decision being made; he has observed a pattern of electrical activity. He has just *decided* it is a decision.

I am happy to accept that there would be some kind of correlation between activity in the mind and activity in the brain. It would, after all, be absolutely astonishing if Libet had discovered that when someone makes a mental decision there is absolutely nothing going on in the brain. So it is not particularly striking that there is some kind of electrical activity that precedes a decision. I am sure that a range of different brain processes go on, conscious and unconscious, when I am seeking out the basis for making a decision. It may be that before I decide to perform an act I mentally check out whether I am capable of performing that act and once I have had feedback from the supplementary motor area that all systems are go, I decide to act. In other words, the electrical activity may be part of the deliberative process that precedes the conscious decision.

Haynes's experiment appears to be more difficult to refute because it is lateralized; that is, by giving the subject to choose to move the left or right hand, which are associated with different neural areas, it appears to give more content to what the scientist observes. I would still maintain that just because one sees an area of the left motor cortex 'light up' in an fMRI scanner before the subject makes the decision to move the left hand does not mean that the left motor cortex is registering that a decision has been taken. It may be that we are testing to see if this decision is viable.

But there is a still more fundamental problem with this experiment, which is that it is not registering a *decision* in any significant sense. There is a wonderful book by the comic writers Douglas Adams and John Lloyd, *The Meaning of Liff*, first published in 1983, which takes the form of a dictionary of things that there are not words for yet. This is one definition:

> **Deventer** (n.) A decision that's very hard to take because so little depends on it – like which way to walk round a park. (Adams and Lloyd 1990: 26)

This minor comic observation seems to me rather significant. What is free will? Is it the freedom to make arbitrary decisions like whether to clench your left or right hand? Or is it the freedom to act on the basis of deliberation and reasoning? Surely it is the latter. Any variety of free will worth wanting, to use Daniel Dennett's (1984) phrase, will involve me bringing my beliefs, attitudes, and desires to bear on the decision. I think if we reflect on the task that Libet and Haynes have set their experimental subjects, I think we can say that they are barely 'decisions' at all. Because on what deliberative basis could you 'decide' to flick the right wrist rather than the left? There is surely no rational process that can lead you to choose to flex your wrist at 14 seconds rather than 15? Libet advises his subjects "let the urge come on its own, spontaneously" (Libet 1999: 48), but an urge is different from a decision.

We do sometimes have to choose between two perfectly balanced and essentially trivial options. For that reason, it seems plausible that the mind, rationally observing that it would be unhelpful for us to remain stuck, like some latter-day Buridan's Ass, overwhelmed by choice and unable to walk round a park, hands the decision over to the brain. It also seems plausible that the brain has some kind of unconscious mechanism to produce these apparent decisions, like the hidden algorithm in a scientific calculator that allows it to produce apparently random numbers. How else could the decision be taken? So we could grant Libet and Haynes that maybe they are right; in this narrow group of non-rational decisions, the brain does the arbitrary work and the will becomes aware of it only afterwards.

We start to see how limited the experiment is when we imagine a more complex choice: to quit your job and become a volunteer in the developing world, or to leave your husband, for example. Could Libet ever identify a *moment* at which the conscious decision is taken? Proper decisions like this are a long process of deliberation, thought, weighing up pros and cons, talking to friends and colleagues, entertaining numerous possible scenarios and outcomes. We slowly come to these decisions with many false alarms along the way. Sometimes you become aware that you have already made the decision; not as some unconscious mechanical process but because you realize you have known for a while what you have to do and your continued thinking is just a search for a more painless route.

And just as it seems absurd to think that the conscious mind could make arbitrary decisions, it seems, to me anyway, implausible that 3lbs of damp grey neural matter could make a decision as complex as whether to leave your husband or throw in your job in academia to become an aid worker.

Zombie Problems

A particularly important philosophical tool in debates about the nature of mind has been the thought experiment. It is an area where philosophy comes close to more explicitly creative writing and to making theatre. Thought experiments are imaginative stories and scenarios in which we expose the very structures of thought and the possibilities of the universe. As such, so their advocates claim, they offer rigorous insights into the nature of things, provide clear evidence for our conclusions about the world. It may seem as if this is where they differ from something like a play, which we are less accustomed to thinking of providing propositions and logical proofs, but we might also recall Aristotle's remark that tragedy is more philosophical than history; although the latter is grounded in fact, it is only able to report what did happen; tragedy tells us what *must* happen (Aristotle 1996: 16). What he means is that the action is tested against the audience's sense of the plausibility of action; in a successful tragedy, we feel at each point that this is what would happen. Tragedies, in Aristotle's sense, are elaborate thought experiments.

Several thought experiments seem to show problems with physicalism and other kinds of reductionism. An early-modern example is Gottfried Leibniz's comment in section 17 of the *Monadology*, which bears directly on the limitations of physical descriptions of the mind:

> Imagine there were a machine which by its structure produced thought, feeling, and perception: we can imagine it as being enlarged while maintaining the same relative proportions, to the point where we could go inside it, as we would go into a mill. But if that were so, when we went in, we would find nothing but pieces which push one against another, and never anything to account for a perception. (Leibniz 1998: 270)

Leibniz is noting that there appears to be a gap between a physical object and phenomena like thought and sensation. In other words, no matter how meticulous your observation and description of the brain will be, it will never account for mental experience.

One might counter that this is to confuse the third-person nature of observing a mechanism and the first-person nature of consciousness; who are we to say that Leibniz's 'Mill' does not have consciousness? Thomas Nagel's famous essay

"What Is it Like to Be a Bat?" takes this on by asking whether it would ever be possible for a physical description to account for mental experience. To do so, he thinks about being a bat and suggests that, given the sophistication of their interaction with the world, bats probably do have some kind of conscious experience: "there is something it is like to be a bat" (Nagel 1974: 438). But, given the distance between our system of perception and bat sonar, it is impossible for us to say what that experience would be like. We cannot understand bat 'qualia', that is, the texture of the bat's experience of the world. This tells us that consciousness is irreducibly subjective, whereas physicalist accounts are irreducibly objective, and it is impossible to turn one into the other. Nagel concludes that this does not mean physicalism is false, just that it is "a position we cannot understand because we do not at present have any conception of how it might be true" (1974: 446). Such a thought experiment suggests what Joseph Levine (1983) has called an 'explanatory gap' that cannot be bridged between the mental and the physical.

Functionalists have argued that this is an unnecessary mystification of the mind. Rather than treat the mind as some mysterious Cartesian essence locked away in private space, they claim that the mind is just what it *does*. The most famous version of this is the 'Turing Test': Alan Turing tried to cut through some of the complexities of consciousness by saying that, basically, if you communicate with a computer for a reasonable length of time and it seems indistinguishable from a human, then we should consider that computer to have consciousness (1950: 433–434). After all, we assume this about other people without being able to see the internal workings of their minds. (I will return to this computational model of the mind later.)

Ned Block, in "Troubles with Functionalism", pushes this to the limit of plausibility. For functionalists, if a mind is simply a set of processes that take one input and produce another output, the mind can be multiply realized: any system that takes the same inputs and produces the same output is functionally equivalent, is functionally the same mind. So he conducts this thought experiment:

> Suppose we convert the government of China to functionalism, and we convince its officials that it would enormously enhance their international prestige to realize a human brain for an hour. We provide each of the billion people in China [...] with a specially designed two-way radio that connects them in the appropriate way to other persons and to [an] artificial body [...]. Surely such a system is not physically impossible. It could be functionally equivalent to you for a short time, say an hour. (Block 1978: 279)

However, it seems clear that although it might be possible for such a network to be functionally equivalent to your brain or mine, the system will still not be *conscious* in the way that we are. The system experiences no qualia, has no experience. Consciousness remains unexplained by functionalism.

Frank Jackson offers another variant of this line of attack on functionalism and other kinds of reductionist thinking in an essay called "Epiphenomenal Qualia". His thought experiment involves a scientist called Mary who lives and works in a black and white room. She has always lived there and we are to imagine that she and her colleagues have pigmentation adjustments that mean she has never seen a colour. However, Mary is a specialist in the neurophysiology of vision and she understands what colour is, how it strikes the retina, how that information is passed down the optic nerve and processed in the brain. She has all the physical information about seeing colour. The question Jackson asks is, "What will happen when Mary is released from her black and white room [...]? Will she learn anything or not?" (Jackson 1982: 130). It seems obvious that she will: she will learn *what it is like* to see colour. If she has learned something new about colour but she knew everything about the physical processes, then knowledge of the physical processes does not give you a complete picture of the mind. So physicalism must be incorrect.

All of these thought experiments are different attempts to defend the irreducibility of consciousness. In doing so, they are capturing what the Australian philosopher, David Chalmers, first described in a conference paper in 1994 as the "hard problem of consciousness" (2010: 5). There are, he says, some easy problems of consciousness, including the difference between wakefulness and sleep, the reportability of mental states, etc. These and others are pretty amenable to scientific description and analysis. But what neuroscience cannot capture or account for is 'experience', not just detecting the world around us but *what it is like* to experience the world. This a problem, because it is unclear how physical processes can give rise to these experiences and why they do: "It is widely agreed that experience arises from a physical basis," he writes, "but we have no good explanation of why and how it so arises. Why should physical processing give rise to a rich inner life at all? It seems objectively unreasonable that it should, and yet it does" (Chalmers 2010: 5).

Chalmers goes further, arguing that the irreducibility of conscious experience to physical processes makes the physicalist or materialist picture of the world incomplete. In his book, *The Conscious Mind*, he makes use of a simple (or apparently simple) thought experiment: the zombie problem. He asks us to imagine his zombie twin; this is "someone or something physically identical to me [...] but lacking conscious experiences altogether" (Chalmers 1996: 94). That is, while they are, atom for atom, identical to the original, and they can perform exactly the same actions, speak the same words, and so on, they have no experience: as Chalmers puts it, "There is nothing it is like to be a zombie" (1996: 95).

The philosophical zombie is evidently not like the foot-dragging, brain-eating monster of popular film. However, it does have remarkable destructive power, in

its capacity to lay waste to reductive physicalism. One way of setting out the zombie argument against physicalism is as follows.

1. If consciousness were physical, then all the physical facts would include all the facts about consciousness.
2. It would not be possible for the physical facts not to include all the facts about consciousness.
3. But it is possible, because zombies are conceivable.
4. Therefore, consciousness is not physical.

The key step in the argument is (3). It may seem curious to suggest that philosophical zombies are possible, just because we can conceive of them. To clarify this, it is important to make distinctions between different kinds of possibility. Things that are practically possible in our world may be called 'temporally possible'. For example, it is possible that I might get a pet dog and call it Zola. Things that are actually impossible but possible within the laws of nature may be called 'nomologically possible'. It is nomologically possible for there to be a pizza the size of Birmingham, though we know this is impossible in practical terms. It is nomologically but not temporally possible. Things are 'metaphysically possible', if to conceive them involves no contradiction, even if they are temporally and nomologically impossible. For example, it is metaphysically possible that someone might travel faster than the speed of light, though we cannot do it and indeed know it to be forbidden by the laws of physics. Philosophical zombies are almost certainly temporally impossible (though I have suspected some acquaintances of being all dark inside) and they are probably nomologically impossible too. But they seem to be metaphysically possible, because it is not logically contradictory to imagine a being with the same kind of body as you or me, but without having conscious experience. It will seem strange and perplexing in some ways, but there is no logical contradiction in the idea as there is in the idea of a married bachelor or a round square.[2]

2 The relationship between conceivability and possibility is much debated. Even among those who believe one can make sense of the idea that conceivability entails possibility, there is much debate about what counts as conceiving something; after all, one could describe it in many ways; is conceiving 'P' the same as imagining (that) 'P'? Or being unable to rule out 'not-P'? Or understanding the proposition that 'P'? Pretending that 'P'? Or something else? It may be that certain things are inappropriate subjects for conceiving; or that certain types of conceiving are inappropriate; or indeed that different kinds of conceiving are only appropriate for certain kinds of subject matter. For an excellent overview, see Gendler and Hawthorne (2002), in particular the editors' introduction. These concerns and questions may be raised for any thought experiment, of course. As it happens, I do not believe my conclusion as regards Naturalism depends on

So the very fact that zombies are conceivable means that they are metaphysically possible. And if they are metaphysically possible, it seems that the physical facts cannot account for consciousness.

Naturalist Zombies

In a rather different essay on the philosophy of zombies, Craig Derksen and Darren Hudson Hick dismiss Chalmers's imagined beings as "zombies in name only" (2011: 22). But while it is true that Chalmers is not drawing on folkloric or popular culture images, in many respects his philosophical zombies do connect profoundly with the troubling manifestations of the walking dead.

The zombie's appearance in western culture coincides very precisely with Naturalism in the theatre. In July 1887, four months after André Antoine founded the Théâtre Libre, the first successful Naturalist theatre, the journalist Lafcadio Hearn arrived in Martinique, where he would stay for two years, eventually writing *Two Years in the French West Indies*, a book in which he recalls many fearful conversations about the mysterious figure of the *zombi*. In Hearn's account, the *zombi* has many features: they may look like innocent strangers, even friends, but are always malevolent (1890: 370); sometimes they are toweringly tall (1890: 188), can be seen by children who have not been baptised (1890: 369), can put on or take off their skin at will (1890: 185), may be sent by witches and wizards to do harm (1890: 371), alongside the aspects that came to dominate the popular imagination: that they are the walking dead (1890: 188) and they eat the living (1890: 370). Being both living and dead, the *zombi* has a peculiar in-betweenness. As Roger Luckhurst notes, this becomes a source of frustration for Hearn, who cannot get a clear definition of this mythical creature: "[the] *zombi* constantly slips away from definition, leaping from noun to noun, lost in translation" (2015: 19). The *zombi* does not merely trouble definition but the supernatural beliefs on Martinique threaten the whole apparatus of rational understanding: "even now things are done that would astonish the most sceptical and practical physician" (Hearn 1890: 374). The passage from which this comes suggests that the author is undecided between cunning and magic in his explanation of what he has seen.

The *zombi* occupies that area of experience that we sometimes call 'the uncanny'. Indeed, the word took on this sense of bringing together the strange and familiar in precisely this era. The *OED* says this meaning became common

conceivability entailing possibility or even that the zombie argument (or any other anti-physicalist thought experiment) is sound, just that the question can be meaningfully asked.

after 1850 and the term was first theorised in this sense by Ernst Jentsch in 1906, and then, more famously, by Sigmund Freud in 1919. Freud's essay pays particular attention to E. T. A. Hoffmann's story "The Sandman", in which the figure of a living girl and an inert doll are chillingly overlaid (cf. *OED* s.v. *uncanny* adj.). Uncanny images of dolls and puppets proliferated in the century between Hoffmann and Freud. In 1853, Charles Baudelaire wrote that some children play with dolls out of a "first metaphysical tendency" to "get at and see the soul of their toys", while others destroy their toys for similar reasons: "a superstitious passion against these tiny objects which imitate humanity" (1995: 203–204). Mark Sandberg has written extensively on the spread of waxworks and mannequins in the Naturalist period, noting their ubiquity "in storefront windows, at international exhibitions, and in several interrelated forms of popular museum display" requiring spectators "to sift through the mannequin's sometimes inconspicuous, finely nuanced ontological distinctions between the living and the dead" (2002: 4–5). One might also think of the way photography emerged in the nineteenth century. At first, photographs of Paris Street by pioneers like Charles Marville were apparently empty of people, because the slow shutter speed meant that while buildings were captured, moving figures left less than a blur on the film. While exposure times reduced towards the end of the century, human subjects would have to stand still for long periods, giving the resulting photograph an uncannily mannequin-like quality.

The spread of such images is only in part to be explained through the advent of new technologies; these images seized the public imagination because they responded to deep cultural anxieties about what it meant to be conscious. Sandberg suggests that mannequins are "tightly linked to the social context of commodified bodies and urban crowds in the late nineteenth century" (2002: 4).

It is unsurprising then to find zombies making an appearance in Naturalism. In *Thérèse Raquin*, Laurent regularly visits the City morgue, waiting for Camille's body to turn up and the descriptions of the bodies oscillate between life and death:

> But drowned corpses are always bloated; he saw enormous bellies, bulging thighs, and rounded, powerful arms. Unable to decide, he stood shivering in front of these greenish hunks of meat, which seemed to be mocking him with their leering grimaces. [...] [O]ne drowned man [...] was so softened and decomposed that the water running over the corpse was washing it away scrap by scrap. All of a sudden the nose collapsed and the lips came away, showing the white teeth below, and the drowned head split itself with grim laughter. (1992: 74–75)

In *The Ladies' Paradise*, Zola describes the mannequins in the eponymous department store in a way that makes them seem undecidably alive and dead: "each one

had a little wooden handle, like the handle of a dagger, stuck in the red flannel, which seemed to be bleeding where the neck had been severed" (1995: 253). Sandberg's suggestion that urban crowds are a source for the uneasy popularity of the mannequin can be seen in the surge of customers on the first day of the sales; one woman remarks, "My goodness! I've never seen anything like this. You're just carried along" (1995: 240). This and related moments in *Germinal* and *The Masterpiece* are images of will-less movement, of zombie crowds. Two of Ibsen's plays have faintly zombie titles: *When We Dead Awaken* is one and the other is *Ghosts* – the original title of which is *Gengangere*, which more literally means 'something that walks again'. Is there perhaps something uncanny about the way Nora, in *A Doll's House*, begins the play as a doll-wife before coming to eerie life?

In *Staging Consciousness*, William W. Demastes argues that the theatre can be a form of resistance to the dominance of scientistic attitudes to the world, precisely in its ability to manifest consciousness on stage. He is sharply critical of mechanistic models of the universe, of reductionism and physicalism, which he accuses of neglecting consciousness and the "mindful hand" (Demastes 2002: 7) giving "meaning and purpose" to the "dumb mechanistic luck and blindly random selection" of the physical universe (Demastes 2002: 17). If this sounds a little mystical, that's because it is. Demastes has frequent recourse to images of "magic" (2002: 9), "soul" (2002: 81), as a way for "dreamers and visionaries" (2002: 5), to repair the "spiritual devastation" (2002: 3) wreaked by science. I am sympathetic to Demastes's sense that reductive physicalism does not give you consciousness, though I am unconvinced that this leads us to this kind of spiritualism.

Demastes suggests that, if they are to get anywhere with consciousness, scientists must occupy "the realm once allocated to artists [and] religious figures" (2002: 5). Again, I should note a reluctance to see those figures elided like this, but it is a clear statement of Demastes's view of art and theatre as fundamentally concerned with consciousness ("The story of theater is quite literally the story of consciousness drawn out of material existence" [Demastes 2002: 42]). If this is fundamental to theatre, it explains why Demastes reserves special ire for Naturalism. Early on, he denounces those arts that "have conceded the dominance of the science, often merely imitating scientific enterprises by generally withdrawing to slice-of-life dissections" (Demastes 2002: 3). This sly reference to Naturalism is made more explicit later when he criticizes its belief that "anything beyond the inertly material must remain officially outside the realm of direct consideration" (Demastes 2002: 13), or that "the theatrical naturalists essentially bracketed out consciousness in their pursuit of understanding our physical 'place' in the external world" (Demastes 2002: 23).

But as we have seen, even if the Naturalists did believe this, it was not consistently manifested in the books and plays. Demastes agrees, in particular noting of Chekhov that he was not a 'pure' Naturalist, citing Bert States remarking that "we can also see a lyrical, or impressionistic, or Maeterlinckian side of Chekhov" (States, qtd. in Demastes 2002: 24). He goes on to show that even in their most 'Naturalistic' plays, the work of Ibsen and Strindberg "muscles out a vision of reality that pulls together the two apparently discrete realms of mind and matter" (Demastes 2002: 24). I might be tempted to say that if your definition of Naturalism excludes Ibsen, Chekhov, and Strindberg, you might want to look at your definition. In fact, Demastes seems both to want to maintain artificially strict definitions and to abandon them; for him, the theatre will always manifest consciousness, despite the efforts of Émile Zola, he being the only major Naturalist left standing.

But does Zola even believe what Demastes and others impute to him? Although one can find quotations from Zola that appear to suggest he is a thoroughgoing physicalist, there is also plenty of evidence that he has an open mind on this point. He repeatedly suggests that 'metaphysical man' has been replaced by 'physiological man' (e.g. 1880: 54, 292; 2003: 104), by which, in theatrical terms, he means that Romanticism has given way to Naturalism; but to believe that this commits Zola to Naturalist theatre being exclusively physiological, you would also have to believe he thinks Romantic theatre to be exclusively metaphysical, a notion that is implausible to the point of incomprehensibility. In *Le Roman Experimental*, he describes the Naturalist novelist "continuing, through observation and experiment, the work of the physiologist, who continues that of the physicist and chemist. In a way, we are conducting a scientific psychology, to complete a scientific physiology" (1880: 16). This does not suggest a reduction of psychology to physiology, but of two separate, but continous and complementary realms of knowledge.

Even more clearly, Zola describes his desire to imagine "character" as a "product of the air and the soil, like a plant", but then admits:

> it is true that we barely hold ourselves to this scientific rigor. All reaction is violent, and we are still reacting against the abstract formula of previous centuries. Nature has entered into our works with such impetuous force that it has filled them, sometimes drowning their humanity, submerging and carrying off characters amid a deluge of rocks and great trees. It was inevitable. We must allow the new formula time to find its own balance and come to its exact expression. Anyway, in these profusions of description, these excesses of nature, there is much we can learn, much we can say. (1880: 229–230)

This passage suggests that to think of the world in exclusively physical terms is a kind of thought experiment, a deliberate (or accidental) excess of thought from

which, nonetheless, insights can be derived. Demastes criticizes Naturalist theatre, speaking, as an example, of how the stage design is cramped and confined by its requirement to serve a narrowly deterministic function: "what is measurable is what is real, and nothing presumably can transcend the limitations observed: the parts necessarily adopt the whole, and the whole is a vastly reduced reality of virtually inert objects acting upon a generally inert life form" (2002: 14). But already one can see a certain hesitation here – "presumably" – alongside the tendentious determination to describe naturalism in such preposterous terms that it becomes easy to dismiss. But what if the uncongenial image of human behavior, far from being a flaw, is actually the point? What if its uncongeniality is not just an abrasive invitation to see human beings as mere physical objects but an invitation to ask how possible it is to conceive of human beings in this way? What if naturalism is asking deeper questions about its own scientific methodology than is generally supposed? The hard problem of consciousness has not been "overlooked or denied" (2002: 14) as Demastes suggests but is perhaps expressly the thing being investigated. To think of Naturalism in this way means we do not have to choose between theory and practice, between hardline scientific doctrine and subtle complex creativity; or, if you prefer, between clear-sighted materialist analysis and its failed attempts at artistic embodiment.

The Chinese Room

In "Minds, Brains and Programs", John Searle is seeking to undermine the claims of 'strong AI', the idea that a computer programme could not just be able to manipulate inputs and produce outputs, but actually understand what it was doing (1980: 417). In effect, the claim is that a computer could have consciousness. He tests this by imagining the following scenario. Searle is placed in a room with one letter box through which messages enter, another through which they leave, and a big book of instructions. These instructions tell him that whenever he receives Chinese characters, he is to consult the book, which will tell him which other Chinese characters he should reply with. Searle understands no Chinese at all, yet, to the person outside the room who does understand Chinese, who submits questions in Chinese and receives meaningful, appropriate replies in Chinese, it will seem as if the person in the room understands Chinese. As Searle says,

> As regards the first claim, it seems to me quite obvious in the example that I do not understand a word of the Chinese stories. I have inputs and outputs that are indistinguishable from those of the native Chinese speaker, and I can have any formal program you like, but I still understand nothing. For the same reasons, [a] computer understands nothing of

> any stories, whether in Chinese, English, or whatever, since in the Chinese case the computer is me, and in cases where the computer is not me, the computer has nothing more than I have in the case where I understand nothing. (1980: 418)

If Searle is right, this is a decisive refutation of strong AI, functionalism, and the computational theory of the mind (and, therefore, of the Turing Test). For that reason it was immediately the subject of numerous counter-arguments by advocates of all these positions.

One of the first and most interesting is what Searle calls the "systems reply" (1980: 419). This is that while it is true that Searle in the room does not understand Chinese, that does not mean the *room* does not understand Chinese. The understanding of Chinese is distributed amongst the operator and the instructions, the paper and pencils and the architecture of the room. One would not expect Searle to understand Chinese by himself, just as we would not expect the CPU of a computer to understand Chinese. Understanding is a feature of the whole system, not just the part. Searle's reply is very simple: the man in the room learns the instructions. Now there is no distinction between the operator and the system – and yet he still does not understand Chinese (Searle 1980: 419–420).

But having arrived at this point, we have also arrived at the theatre. Because what better cultural form can one think of in which people are placed in fake rooms and required to memorize certain phrases and actions that they are required to output in response to defined inputs? This is a description of a type of theatre (albeit a coldly abstract one). Indeed, it is a description of the theatre as a zombie. A physicalist might think the stage an excellent model for the human being: a mechanism into which writers and directors offer inputs which are then output by the actors. But theatre does not easily work like that and has a tendency to complicate things. The extent to which we might find this an uncongenial description of the theatre may also point up the extent to which we find implausible a physicalist vision of the world. The theatre is, in short, not a zombie.

The comparison also points up the subtlety of Searle's thought experiment; the 'systems' version of the Chinese Room, where it is all contained within a single operator, is not a zombie. This operator presumably does have consciousness (of the colour of the walls, the weight of the pen, a pain in his arm, etc.), just not consciousness of the meaning he is conveying. Rather than 'absent qualia', he has 'irrelevant qualia'. Similarly, it is evident that actors are not zombies, emptily mouthing the words they have been told to say, and carrying out moves as instructed by the director; actors have consciousness, but it can be technical consciousness of the best way to produce certain effects, where to lay the stress in a particular sentence, of the precise moment at which to make a particular gesture, or it can be entirely different (what they are having for dinner that

evening, and so on). Whether the actor's consciousness determines the consciousness embodied on the stage or whether it is autonomous from it remains an open question.[3]

In other words, Naturalist theatre, of the kind Zola envisaged, was a crucible in which to conduct a kind of three-dimensional thought experiment about the very possibility of a physicalist view of consciousness. When Zola failed in the theatre, which he did often, it may be regarded an experimental success, evidence for the persistence of consciousness even in the zombie world of the theatre. Bernard Dort argues that we are wrong to think of Zola as simply attempting to put in place "a fallacious and impossible identity between theatre and reality" but instead to create the conditions for "a reflection on the very phenomenon of representation" (2003: 21, my translation). Demastes's book offers a vision of how the new discoveries about consciousness might not simply become another subject matter for theatrical dialogue, but transform the entirety of the theatrical experience. It would be wiser to think of naturalism – intentionally or not – as embodying through the most fundamental aspects of its stagecraft a genuine experiment in the possibility of fulfilling its own scientific imaginary. Rather than take their hypotheses as final statements of their beliefs, is it not time to take the Naturalists' claim to have launched an *experimental* theatre seriously?

Works Cited

Adams, Douglas and John Lloyd. 1990. *The Deeper Meaning of Liff*. London: Pan.

Aristotle. 1996. *Poetics*. Trans., with an Introduction and Notes by Malcolm Heath. London: Penguin.

Baudelaire, Charles. 1995. *The Painter of Modern Life and Other Essays*. Ed. and trans. Jonathan Mayne. London: Phaidon.

Block, Ned. 1978. "Troubles with Functionalism". *Minnesota Studies in the Philosophy of Science* 9: 261–325.

Brown, Frederick. 1996. *Zola: A Life.* 2nd ed. London: Macmillan.

Chalmers, David J. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.

Chalmers, David J. 2010. *The Character of Consciousness*. Philosophy of Mind Series. Oxford: Oxford University Press.

Demastes, William W. 2002. *Staging Consciousness: Theater and the Materialization of Mind*. Theory/Text/Performance. Ann Arbor, MI: University of Michigan Press.

3 Stanislavsky's determination to have actors align their conscious experience closely with that of the character might be understood as establishing a powerfully intuitive channel for creating all of those effects with particular intensity, or, as David Mamet argues, "as useless as teaching pilots to flap their arms while in the cockpit in order to increase the lift of the plane" (1998: 12).

Dennett, Daniel. 1984. *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge, MA: MIT Press.
Derksen, Craig and Darren Hudson Hick. 2011. "Your Zombie and You: Identity, Emotion, and the Undead". In: Christopher M. Moreman and Cory James Rushton (eds.). *Zombies Are Us: Essays on the Humanity of the Walking Dead*. Jefferson, NC: McFarland. 11–23.
Dort, Bernard. 2003. Préface. In: Émile Zola. *Le Naturalisme au Théâtre*. Ed. with Notes by Chantal Meyer-Plantureux. Preface by Bernard Dort. Le Théâtre en question. Brussels: Complexe. 13–25.
Gendler, Tamar Szabó and John Hawthorne (eds.). 2002. *Conceivability and Possibility*. Oxford: Oxford University Press.
Gobert, R. Darren. 2013. *The Mind-Body Stage: Passion and Interaction in the Cartesian Theatre*. Stanford, CA: Stanford University Press.
Hearn, Lafcadio. 1890. *Two Years in the French West Indies*. New York: Harper.
Hume, David. 2001. *A Treatise of Human Nature*. 2nd ed., 2 vols. Eds. David Fate Norton and Mary J. Norton. Oxford: Oxford University Press.
Jackson, Frank. 1982. "Epiphenomenal Qualia". *The Philosophical Quarterly* 32: 127–136.
Jentsch, Ernst. 1997. "On the Psychology of the Uncanny (1906)". *Angelaki* 2.1: 7–16.
Leibniz, Gottfried Wilhelm. 1998. *Philosophical Texts*. Eds. R. S. Woolhouse and Richard Francks. Oxford: Oxford University Press.
Levine, Joseph. 1983. "Materialism and Qualia: The Explanatory Gap". *Pacific Philosophical Quarterly* 64.4: 354–361.
Libet, Benjamin. 1999. "Do We Have Free Will?" In: Benjamin Libet, Anthony Freeman and Keith Sutherland (eds.). *The Volitional Brain: Towards a Neuroscience of Free Will*. Exeter: Imprint Academic. 47–58.
Luckhurst, Roger. 2015. *Zombies: A Cultural History*. London: Reaktion.
McConachie, Bruce. 2013. *Theatre & Mind*. Theatre And. Basingstoke: Palgrave Macmillan.
Mitterand, Henri. 1999. *Zola: Tome I. Sous le Regard d'Olympia 1840–1871*. Paris: Fayard.
Nagel, Thomas. 1974. "What Is It Like to Be a Bat?" *Philosophical Review* 83.4: 435–450.
Nagel, Thomas. 1995. "Dennett: Consciousness Dissolved". In: Thomas Nagel. *Other Minds: Critical Essays 1969–1994*. Oxford: Oxford University Press. 82–85.
OED = The Oxford English Dictionary. 2000–. Online Version 2nd ed. Oxford: Oxford University Press. <http://www.oed.com/view/Entry/210106?redirectedFrom=uncanny&> [accessed 11 December 2017].
Sandberg, Mark B. 2002. *Living Pictures, Missing Persons: Mannequins, Museums, and Modernity*. Princeton, NJ: Princeton University Press.
Searle, John. 1980. "Minds, Brains, and Programs". *Behavioral and Brain Sciences* 3: 417–457.
Smith, Kerri. 2011. "Neuroscience vs Philosophy: Taking Aim at Free Will". *Nature: The International Weekly Journal of Science* 477: 23–25.
Soon, Chun Siong, Marcel Brass, Hans-Jochen Heinze and John-Dylan Haynes. 2008a. "Unconscious Determinants of Free Decisions in the Human Brain". *Nature Neuroscience* 11: 543–545.
Soon, Chun Siong, Marcel Brass, Hans-Jochen Heinze and John-Dylan Haynes. 2008b. "Supplementary Information: Unconscious Determinants of Free Decisions in the Human Brain". *Nature Neuroscience* 11: <https://www.nature.com/neuro/> [accessed 21 March 2017].
Turing, Alan M. 1950. "Computing Machinery and Intelligence". *Mind* 59: 433–460.
Zola, Émile. 1880. *Le Roman expérimental*. Paris: Charpentier.
Zola, Émile. 1992. *Thérèse Raquin*. Trans. Andrew Rothwell. Oxford: Oxford University Press.

Zola, Émile. 1995. *The Ladies' Paradise*. Trans. Brian Nelson. Oxford: Oxford University Press.
Zola, Émile. 2003. *Le Naturalisme au Théâtre*. Ed. with Notes by Chantal Meyer-Plantureux. Preface by Bernard Dort. Le Théâtre en question. Brussels: Complexe.