

---

## **Playing with Metaphor: The effects, problems, and appeal of using a theater metaphor to frame a neurological understanding of consciousness**

*Melody Jue*

Although our experience of consciousness figures as a defining quality of what it means to be human, it remains one of the few mysteries that science cannot entirely explain, continuing to elude all attempts at complete and precise definition. In pursuit of better explanations, many scientists have used the framework of metaphor to conceptualize and understand consciousness. George Lakoff and Mark Johnson suggest that “Metaphor is one of our most important tools for trying to comprehend partially what cannot be comprehended totally: our feelings, aesthetic experiences, moral practices, and spiritual awareness.”<sup>1</sup> Because of the difficulty in conceptualizing consciousness, metaphor figures as both an appropriate and invaluable resource in our effort to understand its nature.

In this paper I examine the use of theater as a metaphor to explain and structure a neurological explanation of consciousness, particularly the account provided by Bernard J. Baars. Baars’ use of the theater as a metaphor for consciousness provides a particularly rich reference point from which to enter into a discussion of the role of metaphor in science. I find it interesting that Baars chose to resuscitate the theater metaphor at a time (1997) when the computational model of consciousness had begun to dominate imaginative research agendas (as it still does today), participating in the longstanding trend of using the latest technologies available to serve as metaphors for consciousness.<sup>2</sup> In light of this trend, why has Baars specifically selected the potentially anachronistic theater metaphor to frame his research? Why has he chosen it, already aware of Dennett’s caustic criticism of the Cartesian Theater? Even though the theater metaphor presents certain obvious problems, might it be useful in conceptualizing components of the conscious experience in other respects? My goal is to evaluate how a theater metaphor might be used to structure a scientific explanation of consciousness, paying attention to what it obscures and what it elucidates, what it ignores and what it highlights, and why its use might be particularly timely at this moment in the history of theory and scholarship.

The use of metaphor in science is problematic. Metaphors describe one thing in terms of another, showing how two things share points of similarity; generally, the more points of similarity there are, the stronger the metaphor. However, this fit cannot be perfect unless one goes so far as to actually say that one thing is the other. Instead, metaphors are always partial; the same metaphor can be used in a variety of

ways, none of which is perfect, but all of which have a utility value under specific circumstances.<sup>3</sup> In this way, metaphoric language seems to produce the opposite effect that scientific language strives for, allowing for the very play of meaning that scientific language seeks to root out in its quest to establish stable and unambiguous meaning.

Several scholars, however, suggest that metaphor cannot be separated out from language at all. Lakoff and Johnson posit that there is no “choice as to whether to think metaphorically. Because metaphorical maps are part of our brains, we will think and speak metaphorically whether we want to or not.”<sup>4</sup> Their research suggests that metaphor is biologically hardwired into our brains as a tool for organizing and understanding information we collect about the world around us. If this holds true, then producing unbiased, metaphor-free discourse may be impossible after all. Laura Otis’s research on metaphorical thinking is one example of recent scholarship that corroborates Lakoff and Johnson’s work. Otis suggests that metaphor pervades scientific thought by structuring the framework of research agendas, and examines how political metaphors of invasion, resistance, and barriers affected 19<sup>th</sup> century science. Otis focuses on the development of German physiologist Theodore Schwann’s cell theory alongside the prevalent political phobia of invasion and suggests, “While Schwann’s achievements reflect biological reality, technological improvements, and his persistence as a scientist, the theory developed in part because Schwann *wanted* to see cell membranes, a cultural as well as a scientific desire.”<sup>5</sup> The metaphorical preconception of “the barrier” structured the kind of questions Schwann asked and what kind of answers he looked for. Otis suggests that, “Metaphors do not *express* scientific ideas, they *are* scientific ideas.”<sup>6</sup>

However, one problem with using metaphor in science is learning how far to take the metaphor. In science, there is always the urge to fill in the metaphor and fully map it onto the thing it intends to describe. For example, when employing a theater metaphor to explain consciousness, the scientist’s natural inclination is to determine which parts of the theater correspond to which parts of the conscious experience. Yet the more thoroughly metaphors are explicated, the more they open themselves up to criticism.

The theater metaphor has perhaps received its most caustic criticism from Daniel Dennett, who specifically critiques the “Cartesian Theater” model.<sup>7</sup> Dennett stresses that one of the major problems with any theater model is the notion that consciousness “happens” at a fixed location in the brain, a central stage. In the past, certain brain structures have been suggested as stages, such as the pineal gland named by Descartes.<sup>8</sup> Dennett points out that contemporary neuroscience research has definitively disproved the fallacious assumption that consciousness “happens” in a fixed location in the brain. Instead, most neuroscientists agree that the brain gives rise to consciousness through some kind of integrated network, where multiple brain structures interact to produce a kind of mapping of the self and the world.<sup>9</sup>

In place of a theater model, Dennett posits a “Multiple Drafts” model of consciousness in which the “content” of consciousness is continually edited by various processes distributed throughout the brain. Yet this “model” is also a metaphor, and it draws attention towards the way in which the content of one’s consciousness is like a manuscript that the brain continually revises: “[A]t any point in time there are multiple ‘drafts’ of narrative at various stages of editing in various places in the brain.”<sup>10</sup> Dennett’s Multiple Drafts model avoids some of the problems of the theater model, such as the fallacy of assuming that consciousness takes place on a central “stage.”

Interestingly, the 18<sup>th</sup> century philosopher David Hume used the theater metaphor in a way strikingly similar to Dennett’s “Multiple Drafts” model. Hume posited that consciousness is

...nothing but a bundle or collection of different perceptions, which succeed each other with an inconceivable rapidity, and are in a perpetual flux and movement....The mind is a kind of theatre, where several perceptions successively make their appearance; pass, re-pass, glide away, and mingle in an infinite variety of postures and situations.<sup>11</sup>

Hume's "collection of different perceptions" and Dennett's "multiple drafts" both emphasize how multiple activities can be simultaneously sustained within the scope of one's consciousness, providing a plurality of perceptions at any given moment. Although Dennett would not explicitly make the connection, his draft-editing metaphor is reminiscent of the passing and re-passing of perceptions that Hume describes through the framework of a theater. Perceptions enter consciousness, coming and going; similarly, drafts are established, and then are revisited at each step of the revising process.

Hume makes an effort to state the conditions under which his metaphor will work, agreeing with Dennett that consciousness should not be conceived of as occurring in a single location in the brain. Hume asserts, "The comparison of the theatre must not mislead us. They are the successive perceptions only, that constitute the mind; nor have we the most distant notion of the place, where these scenes are represented, or of the materials, of which it is composed."<sup>12</sup> Hume's distinction between the part of the metaphor to which he alludes (successive perceptions) and the part that he wants to exclude (location) illustrates the point made by Lakoff and Johnson particularly well, that metaphor is always partial.<sup>13</sup> Hume finds the theater metaphor helpful insofar as it illustrates a single concept, but not in other respects—this is why his deployment of the theater metaphor must be followed by several caveats. Metaphor may be used to communicate aspects of certain ideas, but a reader's interpretation of metaphor is unpredictable and therefore must be guided to avoid the risk of mis-communication.

This is a risk Baars is willing to take. Baars suggests that metaphor can be a useful tool to frame an understanding of science. His book, *In the Theater of Consciousness*, is one of the most sustained explications by a neuroscientist of the theater metaphor. The reason he chooses this metaphor seems contingent on his purpose in writing the book—to narrate the "story" of consciousness provided by cognitive neuroscience. His target audience is the uninformed public: "Only distant hints of the current ferment in scientific consciousness research have reached the public; some of the most fascinating findings and ideas have simply gone unnoticed."<sup>14</sup> Baars feels it is his duty to provide his readers with a framework in which to understand the most salient aspects of cognitive neuroscience research: "What has not been done so far is to forge a working link between these great bodies of thought [human cognition] and the core issue of human experience. That is the aim of this book."<sup>15</sup> Baars' use of the word *forge* suggests that the connection between science and human experience does not effortlessly form, but instead must be effortfully created and articulated. Yet Baars emphasizes how he must be careful about using the metaphor, recognizing that the metaphor not only will "point to new openings, new questions that can now be asked more clearly," but also will be "useful only up to a point; we will keep track of its flaws as well as its uses as we go along. . . . [W]e will not treat the theater metaphor as theory, though all current integrative theories can be thought of as theaters. We will use it just to simplify the evidence."<sup>16</sup> Baars chooses the metaphor for its accessibility to a general public and for its capacity to orient and make sense of a large body of scientific evidence, but keeps open the possibility—indeed, the likelihood—that for some people, the metaphor will raise more questions about consciousness than it answers. This explicative humility alludes to the adaptive quality of metaphor; it can't be proven quite right, but it also can't be proven wrong—there always remains the promise of more perfectly interpreting the metaphor in the future. Perhaps Baars is willing to give the theater metaphor a chance not only because it might be used to "simplify" and frame a neurological understanding of consciousness, but also because the questions that the metaphor raises might lead to new insights.

Baars begins his explication of the theater metaphor by responding directly to Dennett's criticism of consciousness as happening at a fixed location. Baars recognizes that there is no literal "stage" upon which consciousness takes place. Although he rejects the notion of a single stage, he has posited the existence of five separate stages, one for each sense.<sup>17</sup> Despite this misfit, the theater metaphor remains convincing to Baars for other compelling reasons. Baars holds that the theater metaphor is an apt means of making sense

of the limited content that can be held in one's consciousness, or working memory, at one time—an idea that he first introduced in his Global Workspace Theory (GWT).<sup>18</sup> The content of one's consciousness is limited to what can be contained within a given workspace. This content is then broadcast to a variety of unconscious brain processes, all of which compete for access to the limited workspace.<sup>19</sup> Baars recognizes that GWT is not identical to the conscious experience, but considers it to be an effective model for understanding several characteristics of consciousness, such as its role in handling new situations, its limited capacity, and its sequential nature. The simultaneous advantages and limitations again exemplify how a metaphor/model can ever be only partially effective, as it is used to draw attention to certain aspects of the conscious experience without ever capturing the experience in its totality.

Baars' theater metaphor extends this notion of unconscious processing and limited workspace. In this metaphor, unconscious processing happens in two places: within the audience, and backstage. The theater metaphor also introduces a new concept, the notion of a spotlight, which describes attention as a "selective act that results in a conscious event."<sup>20</sup> This idea follows the scholarship of psychologist and philosopher William James, who distinguished between attention and consciousness, noting that attention involves selection, whereas consciousness is the resulting experience. The spotlight also represents the limited capacity of one's consciousness and the periphery of things that one is partially aware of, what Baars calls "fringe consciousness."<sup>21</sup> However, this aspect of the theater metaphor presents new problems. Who or what controls the spotlight, and how? Baars elides this question, suggesting that, "In the dark theater we cannot see who controls the spotlight or what decisions guide its movements. All that is hidden in darkness. We only experience the *results* of those decisions."<sup>22</sup> The "darkness" is a convenient substitute for an actual explanation of what controls the spotlight.

Traditionally, the theater metaphor has faced criticism from its reliance on the presence of a homunculus, the "little human" who is supposed to be sitting inside the brain "watching" the show produced by consciousness. If the experience of consciousness is like the experience of being in a theater, then what is the audience like? What is going on inside the minds of the audience? Would their minds also contain audiences watching them, in an infinite regression? Dennett's "Multiple Drafts" model avoids the assumption that a "subject" "watches" the stage performance of consciousness. Dennett suggests that, rather than being under the direction of a "master" editor, "all varieties of mental activity are accomplished in the brain by parallel, multi-track processes of interpretation and elaboration of sensory inputs. Information entering the nervous system is under continuous 'editorial revision.'"<sup>23</sup> Thus, information becomes edited and revised by one or more processing systems in the brain, not by a unitary "editor."

Dennett claims that in any theater model, information would have to be performed on a stage and "re-presented" to an observer or "master discriminator."<sup>24</sup> Interestingly, Baars' explication of the theater metaphor *does not* treat the audience like a homunculus. For Baars, current research has altered how the "audience" might figure into a theater metaphor of the mind. Baars suggests that the audience is constitutive of various *unconscious* processes which include memory, interpretive systems, and automatisms such as language, reading, and thinking.

All unified models of cognition today suggest some sort of unconscious audience, including unconscious memory archives and automatic routines that are triggered when their "calling conditions" appear in working memory. In the brain the audience seems to consist of functional networks and routines—collections of neurons that work together to perform some job. We can think of them as people sitting in the dark audience, unconscious but with great local expertise.<sup>25</sup>

In other words, unconscious processes are highly developed, but not actively paid attention to. For example, one isn't consciously aware of the mechanisms that make it possible to start speaking, although speech depends upon previous learning. In this way, new information from research about memory and interpretive systems necessitated a change in how Baars applied the theater metaphor to conscious experience. In Baars' interpretation of the theater metaphor, the audience members act not as decision makers but instead as a special kind of resource, a Swiss army knife of abilities that can be utilized as needed.

However, Baars' portrayal of the audience as a resource stands in conspicuous contrast to the assumption by modern audiences that their only job is to be entertained. Audiences pay to attend theater performances for pleasure, not for the explicit purpose of evaluating the actions on stage. Another difference between Baars' audiences and typical audiences is that Baars imagines a cooperative audience that together serves to record and interpret events that occur on stage during a performance. However, real audiences do not communicate during a performance, and even when they do, it is not as a whole. They may discuss the play with any family, friends, or lovers with whom they attended, but not with the entire rest of the audience. In contrast to this, all the individuals in Baars' audiences form a cooperative, synchronized network that together works to process the events on stage. This is a salient difference. At this point in his argument, Baars says that "it is terribly important to keep in mind that audience members are *merely* metaphorical."<sup>26</sup> The phrase, "merely metaphorical" belies his claim that the audience is the "raison d'être of the whole design."<sup>27</sup> Within the scope of his metaphor, Baars settles on no definitive explication of what the audience represents.

If Baars' purpose is to use his metaphor to include *all* of the information about consciousness that current cognitive neuroscience research has to offer, then it fails in several respects, as do all metaphors that attempt to encompass a totality. Baars' metaphor does a good job of elucidating certain aspects of the conscious experience, but it obscures equally important aspects about our consciousness that neuroscience has revealed to *not* be directly accessible by "experience." For example, neuroscience has shown how easily the self can be deceived by what it thinks it experiences. One of these deceptions involves the theater model's fictional depiction of time. In the theater, events are conceived of as occurring sequentially, scene after scene. Dennett regards this as a fiction, citing two phenomena to illustrate: the color-phi phenomenon and Libet's half-second delay. In the color-phi phenomenon, an experimenter sets up a red and green light next to each other. If they are set off a second apart, the lights appear to blink red first, then green. However, if they are set off within increasingly smaller time intervals, the red light *appears to fade into the green light*. This "blending" is what Dennett calls "a memory hallucination," an inaccurate recollection of one's original consciousness of two separate events.<sup>28</sup>

Libet's half-second delay echoes the implications of the color-phi phenomenon, depicting how readily the brain rewrites its observations. Libet showed that it takes the brain a full half second to register direct cortical stimulation. However, when the corresponding body part (in his experiment, the hand) was stimulated at the same time as the cortex, patients consistently reported that they felt the stimulation *before* the tingle produced by direct brain stimulation; the experience of feeling their hand touched was "automatically" "referred backwards in time."<sup>29</sup> This "back-dating" describes an "editing" of the conscious experience similar to an Orwellian rewriting of history in the novel *1984*. This phenomenon suggests that what we think we experience does not exactly correspond to the sequence in which it was registered in our consciousness. Libet notes that "one should not confuse *what* is reported by the subject with *when* he may become introspectively aware of what he is reporting."<sup>30</sup> The observing self and its senses can be tricked.

Perhaps the fundamental semantic disagreement between Dennett and Baars is over the nature of language. Dennett criticizes the theater metaphor for obfuscating certain aspects of the neuro-scientific explanation of how the mind works. For Dennett, anything less than a "flaw-less" trope to explain

consciousness should be discarded. In his own ambitious book *Consciousness Explained*, Dennett strives to explain consciousness in a language that cannot be misinterpreted, a language consummately precise and unambiguous, a language of “science” outfitted with all its armor. In this sense, he joins Robert Boyle in seeking an effective literary technology in which to convincingly explain himself.<sup>31</sup>

Despite the problems he points out of fully mapping consciousness onto a theater metaphor, Dennett cannot resist using a trope to describe how he understands consciousness to “work.” Upon closer scrutiny, Dennett’s own “Multiple Drafts” model depends on certain assumptions about reality, in the same way that Baars’ theater metaphor does. What Dennett’s account of conscious experience fails to account for is an understanding of time. Dennett’s metaphor conceives of perception as inscribed on material documents, drawing attention away from the possibility of their eventual disintegration and decomposition. It obscures the fact that people forget things, remember things partially, and repress other experiences. In contrast, the perceptions, or drafts, in Hume’s theater metaphor serve as entities that appear on a stage and engage the limited resources of the attention for a finite period of time before fading away. Dennett’s “Multiple Drafts” metaphor successfully draws attention towards the process of continual inscription in memory, but simultaneously elides the way these documents of memory face wear and decomposition, variously plagued by smears, tears, and lacunae. Essentially, the “Multiple Drafts” model ignores the mortality and fallibility of memory, thus displaying the same Achilles’ heel that all metaphors share—of being able to effectively describe one thing in terms of another in only a limited and highly select number of ways.

If, as Lakoff and Johnson suggest, it is impossible to think outside of metaphor, then perhaps it is also impossible to stop thinking in terms of our culture’s most prevalent metaphors—of life as a journey, of arguments as battles, or perhaps even of the mind as a theater. Our language is suffused with the theatrical language of acting and playing. If Aristotle is correct in postulating that tragedy is mimesis, the re-presentation of action, then perhaps theatrical language truly is inescapable—for, as human beings with memories, we continuously remember and “re-present” past events in the private space of our minds. We may replay the day’s events, muse upon particular “characters” we encountered, wonder how we could have directed events differently. Essentially, then, the art of the theater suffuses the process of making memory.

Yet the theater metaphor is a unique and complex metaphor, possessing an intrinsic disposition towards modification, revision, and addition. Baars celebrates the protean nature of a theater metaphor throughout his book. His goal is to provide a framework in which to present the most recent scientific information about consciousness, and if his framework doesn’t work for the reader, that’s just fine: he has an appendix available for consultation. Baars even titles this appendix, “Make your own theory: a summary of the evidence,” and presents pairs of conscious versus unconscious facts that “any complete theory must explain. Try to see if you can account for them—with a theater metaphor or in any other way!”<sup>32</sup> For Baars, scientific evidence is like a set of building blocks ready to be assembled into a structurally coherent theory.

What Baars doesn’t include is an appendix of all the possible theater metaphors one might also draw from. He neglects to consider that there might be more than one theater metaphor one could use, despite the fact that the theater is a cultural institution with its own history. It seems necessary to explore a few of the “variables” of the theater to which one might allude, since people’s ideas about what the theater even is are not consistent. I will compare three Western examples.

Greek theater was a fundamentally different institution from Shakespearean theater, both of which were drastically different from the kind of theater people generally attend in America today. One major difference among these three types of theater involves the “space” of the theater. In Ancient Greece the theater was more of an amphitheater or arena, an open-aired space in which performances took place

during the daylight with full exposure to the elements and minimal attention to scenery construction. Shakespearean theater was also performed in the open air, although it took place within a theater that was a more enclosed structure than that in Greece. In contrast, the theater in America and Europe today generally takes place indoors within a more controlled space, a boxed space similar to the cinema. These differences are important to note, since one's conception of the space of the theater influences one's understanding of the privacy and permeability of conscious experience.

Another major difference among these Western theater models involves the varying levels of acculturation and audience literacy. Unlike the varied plays performed in America today, the stories performed in Ancient Greece were drawn from a bank of familiar mythology. For example, each audience member would already be familiar with the story of Agamemnon, but not the particular way in which Aeschylus would narrate the myth and stage the dialogue. Thus, Greek audiences came to see plays knowing an outline of what to expect, sensitive to the nuances of each retelling. As Ian Storey and Arlene Allan note, "What matters for tragedy in particular is the variation from the pattern, not the pattern itself."<sup>33</sup>

The potential for interaction between audience members and actors is also a variable within the institution of theater. Theaters in the Elizabethan era, for example, were spaces of both performance and commerce; prostitutes would solicit customers within the audience, and rich audience members could "buy" a seat *on the stage itself*, placing them within close proximity of the actors. The rowdy audience members would shout out calls to actors, who in turn could choose to ignore or respond to them. Thus the potential for interaction between audience members and actors during the Elizabethan era was drastically higher than the interaction between actors and modern American audiences; today, our habits of watching having been transported from the cinema to the stage, and shouting out or walking onstage would be regarded as highly impolite. These actions would interrupt the individual experience of other audience members, a kind of individual "privacy" we regard as a right.

Cultural assumptions of functionality pervade all theater metaphors, coloring their interpretation by specific audiences. For example, is drama intended merely as entertainment? Is it a form of mass education, instructing citizens in proper conduct? Is it a space for airing political concerns? The answers to these questions might not be incompatible, but an important part of approaching them involves questions of access to theater: who could attend and how much did they have to pay? Theater in the Elizabethan era was, among many things, a form of entertainment, a raucous spectacle that took place on the same stage as bear-baiting. The Greek institution of theater, in contrast, was funded by the state. Greek drama was also the result of a contest, where playwrights would submit their work to be selected for performance at annual festivals. A handful of these playwrights would be selected to put on productions of their plays, and would compete for first prize. People attended the performance for both entertainment and the experience of catharsis [Κάθαρσις], a term generally agreed to describe an emotional cleansing. The functional connotations of what the theater of consciousness "does" could be drastically different, depending on the particular institution alluded to.

This list is by no means complete, but hopefully it provides an outline of possible differences within theater models. It is extremely important to be aware of different models, since each can be used to conceptualize consciousness with a different effect. For example, it has been suggested that the Greek model of theater, in contrast to the modern model of theater "in a box," might better describe a theater of the mind:

Modern audiences are used to the creation of reality in front of them; they expect visual and aural effects that make the dramatic atmosphere "real" and believable.... In the Greek theater the spectators had to do much of the work themselves, to imagine places and settings, import information and relationships from the mythical tradition, visualize in their minds the events occurring off-stage and narrated by others.<sup>34</sup>

In this way, using a Greek model of theater might better emphasize the importance of *context* in interpreting the action onstage. Appealing to other theater metaphors would open up new possibilities for understanding and modeling parts of the conscious experience.

However, any knowledge of historically different institutions of theater establishes the constant possibility of metaphorical contamination. When one refers to something as being “like” the theater, *which particular institution of theater is one referring to?* Even if the answer is specified, to the extent that a protean institution like the theater even can be specified, *to what extent might one’s understanding of a specified theater model be contaminated by one’s knowledge of how other models of theater operate?*

Consider this situation: a person decides to use the theater, as culturally experienced in 21<sup>st</sup> century America, to describe the experience of consciousness. In this metaphorical model, consciousness is like a private experience in a dark, enclosed room; only a select audience is able to observe it. However, another aspect one wants to emphasize is the experience of having a “Jiminy Cricket” type of conscience, an ever-present voice and moral guide. One might be tempted to liken this “conscience” to a Greek chorus, which was almost always present on stage and often advised the protagonist towards a certain course of action. Yet a “chorus” does not play a consistent role in American theater. Mixing the concept of a “chorus” into a contemporary understanding of the theater takes the concept out of its cultural and institutional context. It breaks the concept away, obscuring important details about Greek theater that might not fit well with the metaphor in its modern context. In Ancient Greece, for example, chorus members were selected from amongst the wealthiest Athenian citizens, whose participation was a patriotic duty.<sup>35</sup> The theater was provided to Athenian citizens as a service of the state. This experience of the theater was public and occurred outside, exposed to the elements in the open-air. To refer to this porous aspect of the theater in a modern American context, where conscience and consciousness are private and individual affairs occurring in the controlled space of a building, is to invoke a different metaphor altogether: “In a world of small cinemas and contained theaters, we cannot realize what the experience of the ancient outdoor civic theater was like.”<sup>36</sup> This context would be extremely important to consider when interpreting what it might “mean” to compare a Greek chorus to one’s conscience.

What is important to be aware of is that the act of borrowing concepts from another institution of theater always carries with it old connotations to which people may or may not be sensitive. Derrida understood this well, noting that borrowed concepts “are not elements or atoms, and since they are taken from a syntax and a system, every particular borrowing brings along with it the whole of metaphysics.”<sup>37</sup> When the limb of a word is broken from its surrounding body of text, the break is never completely clean; bits of the old context remain with it. Thus, the issue at stake with mixing metaphors is the immanent possibility of miscommunication, the constant risk of differential interpretation depending on differential understandings of past context.

In the practice of theater, however, one does have the power to mix the old and the new. The theater has always been a kind of imaginative technology, a space in which one can pursue possibilities beyond the physical capabilities of magic or technology. Perhaps what Donna Haraway calls “pleasure in the confusion of boundaries”<sup>38</sup> has already taken place in Baars’ construction of metaphor. Baars did not take his metaphor directly from history. Instead, his theater metaphor evolved first from AI research into “Blackboard Architecture.”<sup>39</sup> In Blackboard Architecture, a blackboard serves as a central repository for posting problems, which are then disseminated and made “visible” to a variety of “experts.” These experts include those using different logic systems, genetic algorithms, neural networks, or a variety of other systems. A mediator holding the “chalk” selects the most promising contribution from an expert to write on the board, continuing the process until the problem is solved. For Baars, Blackboard Architecture evolved into his GWT, and into the theater metaphor from there. Yet vestiges of computational language remain within Baars’ theater metaphor. Baars suggests that

If we think of the brain as a massive decentralized society of biocomputers, it may be more realistic to say that the motor system is learning something, based on conscious information that is made available to many unconscious local control systems. We are not doing the learning; they are.<sup>40</sup>

But technology is not just a vestige. Baars also imagines newly equipping each member of his “audience” with a telephone to represent how audience members, representing different neural networks, communicate with each other:

Automatic routines tend to be relatively separate, specialized, and autonomous; but in fact they work together with others to carry out the details of even the simplest action. Audience members, if you will, have a vast telephone network connecting each to each enabling them to carry out routine tasks without consciousness.<sup>41</sup>

To the extent that he infuses his metaphor with such new technologies, Baars creates a kind of cyborg theater, a confusion of metaphors and their boundaries. The theater metaphor demonstrates its potential for malleability on two major levels: on the one hand that of the cultural and diachronic, which mixes different institutions of theater over space and time, and on the other that of the cyborg, which modifies the physical space and participants of theater with technology. This potential for malleability is perhaps the most significant reason why the theater metaphor remains so appealing, despite the obvious problems it contains. Invoking the theater as a metaphor, for all its ability to morph and change, suggests the utopian possibility of finding more perfect ways to map one’s consciousness, if one were to “play” with the metaphor long enough.

There are other reasons why the theater metaphor may remain attractive to neuroscientists and literary scholars alike. One way to understand these reasons is by examining what issues the metaphor conveniently ignores and what issues it highlights for discussion. Sherry Turkle has written extensively about computers and human consciousness, and her insights provide a useful starting point from which to understand both the general use of metaphor in science and how the theater metaphor provides a particular understanding of consciousness. Turkle suggests, “The question here is not which theory [of the mind], the psychoanalytic or the computational, is true, but rather how these very different ways of thinking about ourselves capture our imagination.”<sup>42</sup> Because the theater metaphor, as Baars conceives of it, has problems as well as value, perhaps we should follow Turkle’s lead and not ask whether the theater metaphor of the mind is true or not, but instead ask why it has captured our imagination.

Turkle suggests that part of the reason why psychoanalytic and computational theories have captured our imagination so powerfully is because they allow us to “play with aspects of our nature that we experience as taboo.”<sup>43</sup> It is significant that Turkle uses the word “play” to describe how we interact with our own natures. Playing with ideas sequesters them in the safe space of an imaginary territory, where their truth value temporarily does not matter. This concept of “play” allows us to think about “highly charged materials” with impunity. According to Turkle, the “highly charged materials” that psychoanalysis and computers allow people to think about are sexuality and the machine-like aspects of our experience. Because they are taboo subjects, people are afraid to talk about them but also *want* to talk about them:

People are afraid to think of themselves as machines, that they are controlled, predictable, determined, just as they are afraid to think of themselves as “driven” by sexual or aggressive impulses. But in the end, even if fearful, people want to explore their sexual and aggressive dimensions; hence, the evocative power and appeal of...thinking about the self as a machine includes the feeling of being “run” from the outside, out of control because in the control of something beyond the self.<sup>44</sup>

The theories of psychoanalysis and of the mind being like a computer thus provide frameworks for a discussion of taboo subjects, providing a starting place for a larger dialogue on the issues surrounding agency. Turkle nails this point when she suggests that, “Exploring the parts of ourselves that we do not feel in control of is a way to begin to own them, a way to feel more whole.”<sup>45</sup> Confronting the fact that we might be controlled by either sexual impulses or mechanical controls is the first step in acquiring mastery over such controls. Knowing that they are in place is to be, in a sense, freed from their total control; it puts us in a position to ask how we can influence the controls that control us.

The questions that Turkle poses for psychoanalysis and computers can be directly applied to an analysis of the theater metaphor of consciousness. What aspects of our nature that we experience as taboo does the theater metaphor of consciousness allow us to explore, confront and play with? The taboo that Baars confronts on every level is the existence of a coherent self. His book is targeted on affirming this “self” at almost every level. He conceives of his book specifically as a tool for readers: “The meaning of ‘consciousness’ intended here *is best illustrated by your own experience* [emphasis added]. *Verifiable public report* is the key to scientific evidence, but *your experience here and now* is quite a good index to the evidence.”<sup>46</sup> In this way, the experience of the self as a unified whole, the reader’s experience, is the standard of truth by which Baars submits his account of consciousness for scrutiny. For Baars, the whole point of using a theater metaphor is to structure how the individual self experiences consciousness.

Yet at the same time, Baars’ argument has within it the possibility that the self is *not* a distinct entity. Baars does not describe a homunculus or Gilbert Rile’s “ghost in the machine” who “watches” the stage performance. Instead, Baars relies on a plurality of audience *members*. When considered for its deeper implications, this portrayal reveals a paradox. Consciousness is supposed to be a unifying experience. Baars also wants to tell a unified story. Yet how can a unified experience arise from the interaction of *different audience members*? Is our consciousness a collective experience? This notion of having multiple audience members has a considerable body of neuroscientific research to back it up, such as the Edelman and Tononi’s neural network theories. Consciousness seems to be produced by the interaction of multiple, specific systems in the brain. It seems that the same theater metaphor can be used to posit a collective consciousness (of an audience) even though the metaphor has historically been invested in reaffirming the identity of the liberal subject. This paradoxical quality of the theater metaphor to both affirm and destabilize the notion of a coherent “self” shows just how acutely the metaphor can be used to interrogate our deepest concerns with identity.

The theater metaphor may be appealing and promising for an entirely different reason: the fact that it coincides with the theatrical turn of theory. It may or may not be coincidental that renewed interest in the “theater metaphor” of consciousness, for all the potential it has to bring attention to the concepts of play, performativity and entertainment, has occurred concurrent with the “theatrical turn” in literary theory. In this respect, a theater metaphor has the potential to integrate and re-present the scholarship of a vast array of literary scholars, including Jacques Derrida, J. L. Austin, Judith Butler, and Mikhail Bakhtin, into a discussion of consciousness and the metaphysics of presence.

There is also something attractive about reuniting theater and theory, two concepts that have engaged in what Plato called an “ancient struggle.”<sup>47</sup> Both stem from the same Greek word *θεωρία*, with relations to the words for sight [*θέα*] and god [*θεός*]. In this way both theater and theory are intimately related to questions of epistemology and ontology. Martin Puchner suggests that

[T]his turn changes the way we think about the history of theory: theatrical theory and the history of theater become interrelated systems so that changes in the one will cause changes in the other. This is not to say that the study of theater, literature, and theory are one and the same thing. But it is to say that an understanding of theatrical theory is only

possible through a knowledge of the contentious intimacy that has bound the theater and theory closely, if not always happily, together.<sup>48</sup>

Perhaps theater and theory will always be bound in a kind of dance, distinguishing themselves from each other, only to reunite in unexpected ways before separating again.

Since the study of consciousness is perhaps *the* essential part of the metaphysics of presence, cognitive science may benefit from considering Derrida's notion of play. Imagining consciousness as a space of play seems particularly appropriate because consciousness refuses to be pinned down and exactly defined. Perhaps we should stop trying to "map" consciousness onto the *structure* of a theater metaphor, actor for actor, operation for operation, and instead strip down the layers of culture that the theater has been dressed with, dealing only with a space of play. This space should not be interpreted as a localized, fixed place where the play comes together, nor as the closed-box of modern theater. Instead we can learn from older models of theater that spaces of play can be dynamic areas of exchange, taking place both within and through the designated borders of the theater. We might regard this space simply as a space, the absent center that Derrida finds so interesting.

However, this understanding of play directly contrasts with the agenda of traditional science, which seeks to establish a system of stable, consistent, and thus reliable knowledge *without* play. Derrida describes the utopian dream of this myth of meaning without play as a centered structure "constituted on the basis of a fundamental immobility and a reassuring certitude, which itself is beyond the reach of play."<sup>49</sup> This is exactly what most scientific research tries to do: bring fully to presence something not completely present, forgetting the sliding sensitivity to difference that defines our consciousness. The scientific study of consciousness might be seen as a massive quest to find "the center" of the self that continually eludes capture, whose agenda is still driven by a kind of humanism with man at the center.

This elusiveness might be attributed to either or both of the attitudes towards totalization that Derrida describes. On one level, a perfect understanding of consciousness is impossible because there is an excess of detail, and always more that one could say about the matter. Yet in another respect total understanding might be impossible because of the concept of substitution: "there are infinite substitutions only because [the field] is finite, that is to say, because instead of being an exhaustible field, as in the classical hypothesis, instead of being too large, there is something missing from it: a center which arrests and grounds the play of substitutions."<sup>50</sup> Consciousness might be such an absent center, since there seem to be infinite ways of adding to a study of consciousness, of describing its nature. Perhaps every facet of consciousness described is like the sign (or perhaps metaphor) which replaces and supplements it.<sup>51</sup> The study of consciousness, which lends itself so easily to supplement, might never be fully closed but instead continually retain the ability to surprise us with new insights about ourselves.

## Acknowledgments

This paper is evidence of the value of interdisciplinary collaboration, emerging only as a product of many lengthy discussions. I would like to thank two people who have been extremely generous with their time and have offered continual encouragement: my advisor, Professor Colin Milburn, for his excellent references and for helping me find the diamond in my rough list of potential research topics; and Professor Evan Fletcher, for many interesting discussions over the past three years. I would also like to thank Professor Bernard Molyneux for a valuable philosophical perspective. On a non-academic note, I would like to thank all my family and friends—you make me care about everything, and I wouldn't have written this paper without you.

---

## Notes

1. George Lakoff and Mark Johnson, *Metaphors We Live By* (Chicago: Univ. of Chicago Press, 1980), 192.
2. Friedrich Kittler. *Gramophone, Film, Typewriter*. (Stanford: Stanford University Press, 1999), 29-34.
3. Lakoff and Johnson, 52.
4. Lakoff and Johnson, 257.
5. Laura Otis, *Membranes: Metaphors of Invasion in 19<sup>th</sup> Century Literature, Science, and Politics* (Baltimore: John Hopkins University Press, 1999), 15.
6. Otis, 127.
7. Daniel Dennet, *Consciousness Explained* (Boston: Little, Brown & Co., 1991), 101-138.
8. Dennett, 34.
9. See, for example, Edelman and Tononi's "Dynamic Core" hypothesis. Gerald M. Edelman and Guilio Tononi, *Reentry and the dynamic core*, reprinted in *Neural Correlates of Consciousness*, Thomas Metzinger, ed., (Boston: MIT Press, 2000).
10. Dennett, 113.
11. David Hume, *A Treatise of Human Nature*, IV.6.
12. Hume, IV.6.
13. Lakoff and Johnson, 52.
14. Bernard J. Baars, *In the Theater of Consciousness*, (Oxford: Oxford University Press, 1997), vii.
15. Baars, 7.
16. Baars, 54.
17. Baars, 72.
18. "Global Workspace Theory," 21 March 2007, <[http://en.wikipedia.org/wiki/Global\\_workspace\\_theory](http://en.wikipedia.org/wiki/Global_workspace_theory)>.
19. "Global Workspace Theory," 21 March 2007, <<http://cogweb.ucla.edu/CogSci/GWorkspace.html>>.
20. Baars, 95.
21. Baars, 44.
22. Baars, 99.
23. Dennet, 111.
24. Dennett, 113.
25. Baars, 46.
26. Baars, 45.
27. Baars, 45.
28. Dennett, 121.
29. Dennett, 154-165.

30. Libet, 1985a, 599.
31. Steve Shapin, "Pump and Circumstance: Robert Boy's Literary Technology," *Social Studies of Science*, Vol. 14, No. 4 (Nov. 1984), 481-520.
32. Baars, 169.
33. Ian C. Storey and Arlene Allan, *A Guide to Ancient Greek Drama* (Oxford: Blackwell Publishing, 2005), 31.
34. Storey and Allan, 52.
35. Storey and Allan, 19.
36. Storey and Allan, 30.
37. Jacques Derrida, *Writing and Difference* (London: Routledge Classics, 2001), 355-6.
38. Donna Haraway, *Simians, Cyborgs, and Women* (New York: Routledge, 1991), 150.
39. Blackboard Architecture, 21 March 2007, <<http://www.nb.net/~javadoug/bb.htm>>.
40. Baars, 59.
41. Baars, 46.
42. Sherry Turkle, *The Second Self* (New York: Simon and Schuster, 1984), 299.
43. Turkle, 299.
44. Turkle, 299.
45. Turkle, 299.
46. Baars, 4.
47. Martin Puchner, "Theater in Modernist Thought," *New Literary History* 3.3 (2002), 521-532.
48. Puchner.
49. Derrida, 352-3.
50. Derrida, 365.
51. Derrida, 365.

**Bibliography**

- Baars, Bernard J. *In the Theater of Consciousness*. Oxford: Oxford University Press, 1997.
- Blackboard Architecture. 21 March 2007. <<http://www.nb.net/~javadoug/bb.htm>>.
- Dennett, Daniel. *Consciousness Explained*. Boston: Little, Brown & Co., 1991.
- Derrida, Jacques. *Writing and Difference*. London: Routledge Classics, 2001.
- Edelman, Gerald M. and Giulio Tononi. "Reentry and the dynamic core." Rptd. in *Neural Correlates of Consciousness*, Thomas Metzinger, ed. Boston: MIT Press, 2000. 130-142.
- Global Workspace Theory. 21 Mar. 2007. <<http://cogweb.ucla.edu/CogSci/GWorkspace.html>>.
- Global Workspace Theory. 21 Mar. 2007. <[http://en.wikipedia.org/wiki/Global\\_workspace\\_theory](http://en.wikipedia.org/wiki/Global_workspace_theory)>.
- Haraway, Donna. *Simians, Cyborgs, and Women*. New York: Routledge, 1991.
- Hume, David. 1739. *A Treatise of Human Nature*, IV.6.
- Kittler, Friedrich. *Gramophone, Film, Typewriter*. Stanford: Stanford University Press, 1999.
- Lakoff, George and Mark Johnson. *Metaphors We Live By*. Chicago: Univ. of Chicago Press, 1980.
- Otis, Laura. *Membranes: Metaphors of Invasion in 19th Century Literature, Science, and Politics*. Baltimore: John Hopkins University Press, 1999.
- Puchner, Martin. "Theater in Modernist Thought." *New Literary History* 33.3 (Summer, 2002): 521-532.
- Shapin, Steve. "Pump and Circumstance: Robert Boyle's Literary Technology." *Social Studies of Science*, Vol. 14, No. 4 (Nov. 1984): 481-520.
- Storey, Ian C. and Arlene Allan. *A Guide to Ancient Greek Drama*. Oxford: Blackwell Publishing, 2005.
- Turkle, Sherry. *The Second Self*. New York: Simon and Schuster, 1984.