Consciousness and Feedback
Explaining the coherence of content, and the integration of semantics into syntactic operations

A thesis
submitted in fulfilment
of the requirements for the Degree
of
Master of Arts in Philosophy
in the University of Canterbury
by

Alexander Francis Ferguson

University of Canterbury
2006
Thanks to my parents, my supervisor Graham Macdonald, and Alex S.
Table of Contents.

• Chapter One. Introduction and History.

• Chapter Two. The Problem of Thinking and Materialism.

• Chapter Three. The Language of Thought.

• Chapter Four. Heuristic Rules in Formal Systems.

• Chapter Five. Syntactic Ambiguity.

• Chapter Six. The Thought Experiment.

• Chapter Seven. Feedback Explained.

• Chapter Eight. Objections and Replies.
## Expanded Table of Contents.

1. Introduction and History.

   1. 1 The Process of Thinking
   1. 2 Plato
   1. 3 Descartes and the Dualists
   1. 4 Interaction and Parallelism

2. The Problem of Thinking and Materialism.

   2. 1 Producing a Chain of Thoughts
   2. 2 The Case for Causation
   2. 3 Strict Physical Mind
   2. 4 Content Coherence

3. The Language of Thought.

   3. 1 Thinking and Language
   3. 2 The Representational Theory of Mind
   3. 3 Problems with Similarity
   3. 4 Linguistic Representation
4. Heuristic Rules in Formal Systems

4.1 Heuristic and Algorithmic Rules
4.2 Heuristic Operations in Formal Systems
4.3 Semantic Heuristic Operations
4.4 Automatic Formal Lingual Systems

5. Syntactic Ambiguity.

5.1 Homonyms, Homophones and Ambivalence in Sentences
5.2 Interpretation Through Context
5.3 The Syntax and Semantics of Machine Language
5.4 Machine Language of Thought

6. The Thought Experiment.

6.1 Part One
6.2 Part Two
6.3 Interpretations
6.4 Implications for the Process of Thinking
7. Feedback Explained.

7. 1 The Function of Feedback
7. 2 Feedback and Heuristic rules
7. 3 Feedback in the Process of Thinking
7. 4 Other Examples of Feedback in Cognition

8. Objections and Replies.

8. 1 Natural Language of Thought
8. 2 Machine Language of Thought
8. 3 Including Consciousness
8. 4 Intentions of the Theory
Abstract.

“Despite the shift from dualism to materialism, philosophy of mind and
cognitive science still face the challenge of explaining the interaction of the
physical and the mental. The language of thought hypothesis, combined with
advances in computing offers a promising explanation of the aforementioned
interaction by capitalizing on the parallels between the syntax and semantics
of language. Unfortunately, the language of thought hypothesis is vulnerable
to arguments and objections that stem from syntactic ambiguity, semantic
poverty, and semantic causation, all of which stand in the way of creating a
working theory of mind. I will claim that these problems can be avoided by
incorporating feedback to regulate the semantic content in chains of
thoughts. The regulation of semantic content would allow the operations
performed by the psychological machinery responsible for the process of
thinking to be causally sensitive to the semantic content of thoughts. The
causal influence of the feedback would be heuristic, rather than algorithmic,
avoiding the explanatory pitfalls traditionally encountered in attempts to
integrate semantic content into strict, syntax manipulating mechanisms. The
inclusion of consciousness feedback also answers a solipsistic worry, the
syntactic zombie, as well as fitting more closely with our experience of
cognition.”
Chapter One. Introduction and History.

The purpose of this paper is to provide a possible solution to the problem of thinking. The problem of thinking arises when attempting to mechanize the forces or rules by which the process of thinking operates. Simply stated, the process of thinking is the progression from one thought to the next. The process of thinking is the progression from thought to thought, and it is one of the central claims made in this paper that thought-to-thought progression involves thought-to-thought causation. The idea of thought-to-thought causation necessarily assumes that there are causal forces to which thoughts are sensitive; and that thoughts themselves are at least partially responsible for causing other thoughts. The basis for the assumption of causation is provided later on in the paper, and until then the reader is asked to suspend judgment on the issue, at least for the sake of argument. To illuminate the causal forces responsible for the process of thinking, this paper will focus on and examine of the result of that process. The result of the process of thinking is the production of a chain of thoughts, with the ordered progression of one thought to the next being necessary for the production of such a chain. Examining a chain of thoughts is informative about the process, or processes that are responsible for its creation. In the same way, a piece of pottery, such as a teapot, can contain information about
what was involved in its creation. A study of the marks in the clay, cracks in the glaze, and so on betray the type of shaping tools and techniques, type of glaze, and even type and temperature of the kiln used to make the teapot.

The multitude of topics presented in this paper, while apparently diverse, all serve to explicate that which governs the process of thinking. As is the case with many, if not all topics of philosophical inquiry, the process of thinking is not conceptually isolated. As an aspect of the philosophy of mind in general, the topic of the process of thinking and the problems and questions associated with it are interrelated with several other topics in the philosophy of mind. This relationship with other topics, issues and problems goes beyond the philosophy of mind, influencing and being influenced by topics in epistemology, and metaphysics, as well as so-called hard sciences. As this is the case, a certain amount of attention must be paid to the way in which these interrelated aspects influence the topic of the process of thinking itself. Therefore, in addition to the paper’s attention to the central topic of thought-to-thought causation, there will also be a concise explanation of the related topics, concepts and arguments involved. A brief introduction to theories of dualism for example, is vital for a proper understanding of the history of the philosophy of mind. Providing a concise history couches the topic of the process of thinking within the larger issue of the mind/body problem. For example, one has to decide what substance is involved and how it affects the process of thinking; one’s view on this is crucial to a proper understanding of the process of thinking and the problems associated with it as a whole.
There are various theories concerning the nature of the mind, the vast majority of which fall into two categories. The first category includes views that hold that there are purely nonmaterial things, and that the mind is one of those things. These theories commonly assert that there exist both material and nonmaterial things; hence they are called dual substance, or dualist theories. The second category is materialist monism, which contrasts with dualism by denying the existence of nonmaterial things. Materialist monism, or more simply, materialism, argues that the mind, or at least mental properties, can be said to exist without appealing to a nonphysical substance. Although materialism is now the more popular view, this was not always the case.

A classical version of dualism is demonstrated in Plato’s *Phaedo*. In *Phaedo*, Plato (speaking as Socrates) argues for the existence of a soul that is separate from the material world, as well as the senses. This distinction between the soul and the body is what gives Socrates his resolve in the face of death.

“And is (Death) anything but the separation of soul and body? And being dead is the attainment of this separation; when the soul exists in herself, and is parted from the body and the body is parted from the soul—that is death?”.

Plato’s arguments for a nonmaterial soul are intended to answer certain epistemological questions about knowledge. According to Plato, the nonphysical soul could interact with a similarly nonphysical world of Forms. The world of forms was non-physical, and distinct from the world that we know through sensory experience, which he called the world of flux. The inhabitants of the world of forms are perfect conceptual forms, which are
mimicked imperfectly in the world of flux. Every kind of thing that is manifested in the world of flux has a perfectly formed version of it that exists in the world of forms. As a soul, we experience the world of forms before we are born. Once our soul takes up residence in our body, we no longer have access to the world of forms and we recollect relevant knowledge we obtained in the world of forms as we perceive the world of flux through our senses.

The immaterial soul formed the basis for Plato’s conception and explanation of the intellect. The pre-birth access to the world of forms our souls had gave us an understanding of the forms matter takes here in the world of flux. For example, we can identify a horse we encounter here in the world of flux as a horse because it resembles the perfect horse form we know from the world of forms. That is to say, the flux horse is a horse because it exhibits horse-forms, or horse-ness. The horse-ness of the flux world horse is a transcendental property of the object. The same goes for any object in the world of flux. A baseball exhibits baseball-ness, roundness, whiteness and so on. If you consider any of these properties by themselves as divorced from the flux object that exhibits it, roundness for example, you are thinking of the form of roundness. This does not mean that the form of roundness is actually the idea of roundness you have in your mind. For Plato, the forms exist as separate from the objects in the world of flux and the ideas in your mind. The horse-ness of a horse, or roundness of a baseball exist as forms in the world of forms. The flux horse and the form horse are bound to be different, as nothing in the world of flux is perfect in form. The world of flux is imperfect, and according to Plato, is therefore subject to change. Since the world of forms is perfect, there is no reason for any change to occur. The world of forms, although different from the world of
flux, was no less real. In fact, Plato considered the world of forms to be more real, or ultimately real; the world of flux being seen by Plato as mere copies or images of the objects in the world of forms. The relationship and reality of both the world of flux and the world of forms is outlined in Plato’s famous story of the cave. “

The story of the cave demonstrates the relationship between the world of forms and the world of flux. The story itself continues further than what is included here, although the remainder of the allegory deals more with the pursuit of truth through reason. Since it is both philosophically important and relevant, the allegory of the cave could have been included it in its entirety. However, the portion included in this paper succeeds to illustrate the division between the world of flux and the world of forms. The proposed division between the soul and the body, as well as the world of forms from the world of flux is historically important. Philosophically, it introduces the beginnings of the mind/body debate that eventually becomes central to the philosophy of mind.

After Plato, the next major step in the development of a dualist approach to the mind/body problem was brought about in 1641 AD. In an attempt at epistemic certainty, the French Philosopher Descartes presented a more modern version of dualism. Like Plato, Descartes argued that the mind is nonmaterial and therefore fundamentally different than the physical body. Unlike Plato, Descartes did not posit the existence of a world of Forms. Famously, Descartes held that the only thing that he could be certain about its existence was his own consciousness. Descartes thought that the consciousness is the mind, and that the mind is a non-physical entity. In
Meditations, Descartes argues that the mind and the body are two different substances. An example is Descartes argument from doubt. The argument from doubt begins with the cogito as the first premise. A brief outline of the argument is given below.

1. I cannot doubt my own existence.
   
   “Thinking is another attribute of the soul; and here I discover what properly belongs to myself. This alone is inseparable from me. I am--I exist: this is certain; but how often? As often as I think; for perhaps it would even happen, if I should wholly cease to think, that I should at the same time altogether cease to be. I now admit nothing that is not necessarily true. I am therefore, precisely speaking, only a thinking thing, that is, a mind (mens sive animus), understanding, or reason, terms whose signification was before unknown to me. I am, however, a real thing, and really existent; but what thing? The answer was, a thinking thing.” iii

2. I can doubt the existence of my body.
   
   “I suppose, accordingly, that all the things which I see are false (fictitious); I believe that none of those objects which my fallacious memory represents ever existed; I suppose that I possess no senses; I believe that body, figure, extension, motion, and place are merely fictions of my mind.” iv

3. Therefore, I am not identical with my body.
   
   “I am conscious that I exist, and I who know that I exist inquire into what I am. It is, however, perfectly certain that the knowledge of my existence, thus precisely taken, is not dependent on things, the existence of which is as yet unknown to me: and consequently it is not dependent on any of the things I can feign in imagination. Moreover, the phrase itself, I frame an image (efffingo), reminds me of my error; for I should in
truth frame one if I were to imagine myself to be anything, since to imagine is nothing more than to contemplate the figure or image of a corporeal thing; but I already know that I exist, and that it is possible at the same time that all those images, and in general all that relates to the nature of body, are merely dreams [or chimeras].”v

This argument turns on Leibniz’s law. If Descartes has a property that his body lacks, then he cannot be identical with his body. If Descartes is not identical with his body, then materialist monism must be false. This and other arguments in meditations created the Cartesian version of substance dualism. Unfortunately, this argument contains a fallacy that falsifies the outcome. The fallacy which Descartes’ argument is guilty of is confusing an epistemic property with a metaphysical property. In other words, doubting an objects existence does not affect the metaphysical properties of that object. For example, suppose I spend an entire day sitting on my front porch and watching cars drive past on the road. During my day spent watching traffic, a red corvette drives by my house twice. I notice this car each time it goes past, but I mistakenly identify it the first time it drives past as a Camaro, and correctly identify it as a corvette when it passes the second time. Obviously, my believing the car to be a Camaro does not change the fact that it is a corvette. My epistemic belief does not change the metaphysical properties of the corvette. In the same way, my doubt that my body can exist is not a property of my body. Hence, there is no metaphysical property difference between my existence and my body’s existence that is affected by epistemic certainty or doubt.
For Descartes, the mind has to be immaterial in order to explain mentality and rationality. There is nothing we have encountered in the physical world that behaves in accordance with rationality, unless it is also endowed with a non-physical mind. The physical world operates according to natural law. For example, water boils at one hundred degrees Celsius. Water does not boil at one hundred degrees Celsius because it believes that it is the right thing to do, or that after some deliberation decides to boil. Water does not boil because it ought to, or it feels like it. Water boils because it is subject to natural laws that dictate its actions. On the other hand, human beings do behave in ways that do not appear to be directly dictated by natural law. We often explain our actions with statements like “It was the right thing to do.” Or “I was very thirsty.”. Almost all of our behavior is attributed to our mind. We do the same when we try to explain or predict the behavior of others. Unlike substances under physical laws, the mind can act according to rationality. Cartesian dualists claim that if the mind is governed by natural laws, it could not act rationally. Since all physical things are subject to natural law, the mind must not be a physical thing.

According to substance dualism, therefore, we are composite beings who consist of an immaterial mind and a material body. Since the mind and body are separate entities, one can exist without the other. There are examples of bodies existing without minds, such as rocks or mud puddles. It is also possible to conceive of minds existing without bodies, the famous “Evil Demon” argument suggests just such a possibility. According to Cartesian dualism, mind and body interact. A Dualist theory in which the mind can causally affect the body, and the body can causally affect the mind is known as interactionism.
The interaction of the mind and body is one of our common sense beliefs about the world. The physical world affects consciousness through the senses. Being pricked with a pin produces the conscious experience of pain, and so on. It also seems that our mind affects our behavior. When I feel thirsty, I get up and get a drink. However, for this to occur, there had to be a way for the physical and the mental to affect one another. Descartes believed that the pineal gland was the point at which the mind and the body met.\textsuperscript{vii} This turned out to be false, and it was eventually to be the end of substance dualism. That the mind and body do not interact in the pineal gland is not so much a failure regarding where the interaction occurs, but how it could occur. Based on our common sense and scientific understanding, there appears to be no good reason to accept that there are wholly immaterial things in the world. If minds exist in the manner that substance dualism claims they do, we could never detect their existence. Even if there are such things as non-material minds, it is difficult to imagine how something without any physical properties might affect things in the physical world. If non-physical events affected physical events, the science of physics and the theory of materialism would be in serious trouble, as they both assume that there is what is called the “causal closure” of the physical world. Causal closure of the physical requires that the cause and effect of every physical event be entirely explicable through purely physical means. In other words, physical events can only cause, and be caused by other physical events. “Perhaps the point is made without using the term ‘physics’ at all. Consider the following argument schema:

All X effects are due to X causes. So anything having X effects must itself be X.” \textsuperscript{viii}
Using the argument above, one can see how the mental and the physical must be identical if they are to be responsible for physical effects. All physical effects are due to physical causes, therefore all mental causes must be physical causes if they instantiate physical effects. The inability to explain how minds can affect a causal difference in the world spelled the end for interactionist substance dualism.

Although Interactionism was largely abandoned, dualism as a whole still had other options. One such option was parallelism. The downfall for Interactionism was that it could not explain how the mind and body could causally affect one another. Parallelism is the view that the mind and the body do not affect one another in any way. The mind and the body appear to interact, but this it only because of a harmony that exists between the two. That such a harmony might establish itself randomly was incredible, so parallelism appealed to divine intervention. Malebranche theorized\(^x\) that God intervened on every occasion in which the mind and body appeared to interact. Hence, Malebranche’s version of parallelism is also known as occasionalism\(^x\). If you stubbed your toe, God made certain that you felt a corresponding pain. Likewise, if you felt like shouting “Hurrah!”, God would intervene and your body would shout. Not surprisingly, this view had several problems, the most glaring of which was that it was just a slightly different form of interactionism, where God replaces the pineal gland. Another problem with Malebranche’s parallelism is that God would have to be unnecessarily busy, insuring that the mental and the physical were kept in order with one another by affecting the mental or the physical each and every time they appeared to interact. This lead philosophers like Leibniz to propose a pre-established harmony between the mental and the physical.
Leibniz argued that God had created a harmony between the mental and the physical so that the two would always match up without requiring any further intervention. Any time you stubbed your toe and felt a corresponding pain, you felt the pain not because the physical act of stubbing your toe causally affected your mental state (of being a toe-stubbing pain), or because God intervened on that specific occasion. The reason you felt the toe-stubbing pain was that God had set the physical and mental in motion so that physical events would correspond perfectly with mental events and vice versa. In doing so, pre-established harmony parallelism preserves the substance dualist distinction between the mental and the physical. Unlike interactionism and occasionalism, parallelism avoids the complications that occur when attempting to explain causal interaction between the mind and body by denying any interaction between the two at all. The downside of the theory was that it required God to create the harmony between the mental and the physical. Unfortunately, for parallelism, without God to establish the harmony, the idea of pre-established harmony is untenable.

The failure of interactionism, and the implausibility of parallelism marked the end of the popularity of dualism. Parallelism was seen as implausible because it depended so heavily upon theistic intervention. Without God to intervene or set things in a perfect harmony, parallelism fails to deliver a plausible account of mind/body interaction. As it neither required the existence of a non-physical substance, or the assistance of God to explain the mind, materialism became dualism’s eventual successor.

However, although dualism has been largely abandoned in favor of materialism, the burden of this thesis is that the problems which dualism set out to solve remain unanswered. Materialism, although free from having to
explain the logistics of the interaction of physical and non-physical substances, has problems of its own, some not dissimilar to those faced by its forbears. One such problem, one on which this thesis focuses, has to do with the causal properties of thoughts. The difficulty occurs when one attempts to explain our experience of thinking.

Chapter Two. The Problem of Thinking and Materialism.

Thinking can be seen as both the production of new thoughts, and the progression from one thought to the next. For the most part, the production of new thoughts will be ignored and this paper will focus on the progression of thoughts, because the progression from one thought to the next is what forms a chain of thoughts. This progression of thoughts will hereby be referred to as the process of thinking. The process of thinking can be examined separately from the topic of the production of new thoughts, although it is almost certainly the case that the production of new thoughts is an aspect of the process of thinking. For the sake of simplicity, the paper will focus on the connection between two thoughts rather than the mechanisms responsible for the production of thoughts themselves. In other words, the paper will explore the progression from thought A to thought B and on to though C, rather than exploring the neural processes involved in the creation of thoughts A, B and C. The questions raised by the creation of new thoughts would only complicate an already complex topic, and can be safely sidestepped while exploring the process of thinking. While the
process of thinking certainly involves the creation of new thoughts, the question this paper is concerned with is “How are the thoughts in a chain of thoughts connected to one another?” or “How does one thought relate to the next?”, as opposed to questions like “What form does a thought take when it occurs?” or “Where in the brain are thoughts made?”. The focus is on the progression from one thought to the next; the mechanism of the production of thoughts is assigned to a thought producing module.

The process of thinking must operate according to a causal framework of some kind, otherwise the progression of thoughts would be random. Given the number of possible thoughts, it is almost impossible that a random string of thoughts would resemble what we experience when thinking. It is possible to think about anything, from oysters to deities. The options available to our minds for thoughts are essentially endless. Given this huge array of available thoughts, the odds against them randomly occurring in an order that resembles our experience of thinking are astronomical. In addition, our experience of thinking, the phenomenology of thought, gives us reason to believe that there is some kind of causal progression that brings order to what would otherwise be a chaotic progression. For dualist theories, the interaction between thoughts is easily explained. Thoughts are of the mind, and the cognitive processes of the mind create the connections between them. Just because the mind is a non-physical entity, according to dualists, there is no problem for the proposal that the causal relationship between thoughts is based on the content of the thoughts involved. Since the mind is a non-physical substance, it makes no difference if the processes that govern it are nothing like physical laws. The strength of dualism, therefore, is that thought-to-thought causation becomes relatively unproblematic: but
as we have seen, dualism fails critically in other areas, namely in explaining mind/body interaction. Thought-to-thought causation only becomes a problem when we embrace a materialist view of the mind.

The problem that materialism has with thought-to-thought causation can be called the problem of thinking, which arises when one attempts to explain the process of thinking using only physical laws. While both experience and theory tells us that there must be some causal forces at work in the process of thinking, this does not mean those forces are necessarily explicable using physical laws. The reason we believe there might be some causal framework to our process of thinking is that our experience of thinking exhibits an order of coherent progression. Some may object to a causal order of thoughts on the ground that while our thoughts are not randomly structured, that alone may not be sufficient to argue that they are necessarily causally ordered. While that may be the case, appealing to causal order is the best explanation available for a materialist theory. For the remainder of this paper, there will be a distinction drawn between a string of thoughts and a chain of thoughts. A string of thoughts refers to two or more thoughts that do not make sense, or in other words, display a lack of coherence of content with one another. The string of thoughts is unlike what we experience most of the time when we engage in the process of productive thought. A chain of thoughts is a series of two or more thoughts that exhibit content coherence and an organizing teleology. Let me explain what these are.

When we experience a thought, we experience the sense of being aware of the content of the thought. There is a difference between the physical existence of the thought in our brains, and our experience of the
thought. When I think, “It is cold outside today.” that is the only conscious experience I have of that thought. When I think the sentence “It is cold outside today.”, that is the content of the thought. Presumably, the sentence “It is cold outside today.” does not physically manifest itself in my brain in the same, nor even a similar form as it does on this page. The physical manifestation of the thought is probably a pattern of neurons and synapses firing in my brain. I have no conception about what neural pathways are firing when I think that thought, nor any experience of the thought as it occurs in my brain other than, “It is cold outside today.”. Therefore, I have no direct knowledge of the physical form of the thought as it manifests in my brain; I only have knowledge of the content of the thought.

When we say that the process of thinking must have a causal framework, it is essential to emphasize that we are talking about a causal framework that is sensitive to content. Our experience of the process of thinking is an experience of their being a connection between the content of the thoughts involved. It is because we experience the content of the thoughts as being somehow connected that we claim that there must be some order to our thoughts. Therefore, the causal framework responsible for the process of thinking respects or operates according to the content of thoughts.

The idea of a causal framework that operates according to content is problematic for materialism. As materialism discards the idea of a non-physical substance to account for mind, it appears that it must make do with a purely physical account. Nevertheless, it looks as though the content-based causal forces at work in the process of thinking do not conform to purely physical laws. Arguably the central problem that an explanation of mental causation faces is that physical laws, or rather the causal processes governed
by the laws, do not operate with respect to mental content. If they did, explaining the interaction of the mental and the physical would be no trouble at all. Returning to the problem of thought, it is perfectly conceivable that a string of thoughts that were caused by forces governed by physical laws would be incoherent, and the experience of that string of thoughts would be no different from a series of random thoughts. The problem is that content and physics appear to be incompatible. There are no laws or equations in physics in which the content of thoughts play any role, and it is this that makes us think that physical causes are not sensitive to intentional content. As we saw earlier, water boils at one hundred degrees Celsius with no mention of whether or not the water is in a rational state of mind, if it happens to be thinking about boiling, or even if anyone is thinking about it boiling... Using the laws of physics to explain the process of thinking, with no reference to thought content, would be almost as bad as having no causal explanation at all. Having causal forces at work that operated with no respect to content could produce a chain of thoughts that would be indistinguishable from a string of thoughts assembled at random. This is not to say that the content of thoughts is wholly removed from physical causation. If a pin pricks you, you still experience pain and think something like “Ouch”. Likewise, if you think, “It is dark in here.”, and reach to turn the light on, there are physical laws involved. This is a different aspect of mental causation than the problem of thinking. The process of thinking is concerned with the coherence of content between thoughts, rather than the causal interaction between being pricked with a pin and thinking “Ouch” or turning on a light. The pin and light-switch examples are both questions for the interaction of the mental and the physical. The problem of thinking is a problem that revolves around explaining the mental interacting with the
mental, using the physical as the basis for the explanation. Moreover, given that materialism asserts that the mind is physical, the mental effects (thoughts) are also physical, and so their causes have to be physical, given the causal closure of the physical.

Materialism needs a way to explain the coherence of content from one thought to the next, without that explanation involving an incompatibility with physics, or making it mysterious how such coherence is achieved. Coherence of content is another way of saying “makes sense”. If two or more thoughts exhibit a coherence of content, it means roughly that the thoughts make sense in regards to one another. What it is to “make sense” is difficult to define, because of the amount of flexibility allowed in strings of content that can exhibit coherence: there are all manner of ways in which the content of two thoughts can be coherent, or make sense to one another. For instance, suppose you were reading a murder mystery book and decided to speculate about the identity of the murderer. You might first think of all the suspects who are involved, and think about their various alibis. Then you think about the clues, murder weapon, time of death, and that sort of thing. These thoughts make sense to one another; they display a coherence of content. However, the reason we know they make sense is because we know their teleological destination as well, the ‘aim’ of this process. We know that the goal for this progression of thoughts is hazarding a guess at the identity of the murderer in our book. This teleological property of a chain of thought provides a context for attribution of coherence to the content.

The progression of thoughts from the initial thought about who the murderer could be, to the final reasoned guess about the murderers identity is an example of a chain of thoughts. There are two organizing forces that
govern the progression of the thoughts in the chain. A sort of relevant logical progression, as well as a teleological end or goal that effectively guides the logical progression, resulting in content coherence. Following this progression, the thoughts are first about incriminating clues. After the thoughts about various clues, come thoughts about possible suspects and so on. Of all the thoughts that it would be possible to have at any given time, it would be almost impossible that these thoughts arranged themselves in that order according to pure chance. The first organizing principle for a chain of thoughts is logical progression with relevance. The use of “logical” in this case is a loose sense of the word: the thoughts need not be deductively organized in order to manifest the kind of logical coherence intended. This logical progression is responsible for coherence of content. In other words, it should make sense when one thought follows another. The following is an example of relevance between three thoughts. The three thoughts in the example are part of a chain of thoughts that makes sense.

Thought 1. “I wonder who killed Dr. Penrose?”
Thought 2. “It was probably someone who didn’t like Penrose.”
Thought 3. “Gavin hated Dr. Penrose, it may have been Gavin.”

The three thoughts above display a logical relevance to one another. For contrast, here is an example of three non-relevant thoughts.

Thought 1. “I wonder who killed Dr. Penrose?”
Thought 2. “Penguins don’t live in outer space.”
Thought 3. “My recipe for custard cream pie requires one cup of sugar.”

There are many other similar examples that one could construct to demonstrate relevance between thoughts. The examples above show relevance to be a necessary condition for chains of thought (as we
experience them) as opposed to strings of thought. It is certainly possible to imagine strings of thought in which the thoughts are irrelevant to one another and would have no productive ability. There is no point thinking irrelevant thoughts. The process of thinking is useful and productive when it serves a purpose. More specifically, thinking is useful when it achieves a goal. The goal itself is important, as teleology is an organizing property of a chain of thoughts. The purpose, or end point of a chain of thoughts is vitally important in judging whether or not the content of the thoughts involved “make sense”.
Chapter Three. The Language of Thought.

In the previous chapter, we outlined a problem for materialism. The process of thinking results in a chain of thoughts that displays an order that appears to be inexplicable from the perspective of purely physical causation. What materialism requires is a purely physical causal system, which respects and preserves content and coherence. A physical system that fits this description is described in the language of thought hypothesis.

The language of thought hypothesis is a theory about the nature of thinking. Whether or not it is intended to be regarded as an empirical theory is open to debate, although Fodor introduced his hypothesis in *The Language of Thought* as “the only game in town” according to our current theories in cognitive science. This means that whether or not the language of thought hypothesis is an empirical theory itself, it is at least intended to be compatible with, and provide support for, the current empirical theories in cognitive science.

According to the language of thought hypothesis, thoughts occur as part of a mental language called Mentalese. What kind of language Mentalese will turn out to be is still a matter of some debate, but for now, it is easiest to explain it and its operations as a natural language. A mental language operates like a natural language in form and function, with syntax and semantics. Mentalese is physically realized in the brain, in keeping with a materialist viewpoint. The language consists of a formal system of representations with syntactic properties and semantic content. The use of the word “formal” in this case is different from what Plato meant by formal.
Rather than being “of the world of forms.”, formal refers to the rules and tokens of a type of system. Having a Thought, according to the language of thought hypothesis, is the tokening of a representation, one that has a syntactic structure with an appropriate semantics, in the brain.

The language of thought hypothesis views thinking as the manipulation of syntactic structure with a corresponding semantic content. This conception of thinking is a version of the RTM, or Representational Theory of Mind, the central claim of which is that thoughts have a formal structure which possesses a representational content. "LOT wants to construe propositional-attitude tokens as relations to symbol tokens. To believe that P is to bear a certain relation to a token of a symbol which means that P." xiii

Mental events, according to the representational theory of mind, are the result of the manipulation of information-bearing structures. In other words, RTM contends that thoughts are representational because they have representational content. The representational content of a thought is not identical with the structure that bears it. The relationship between the structure and the content in this case is not unique to thoughts: there are many non-mental examples of representation. One such example would be a painting that was of a famous figure or of a famous event - for example, a painting of Napoleon that consists of a combination of canvass and paint. The image on the canvass is of Napoleon the person, who was made up of flesh and bone. The painting is a representation of Napoleon the person, but the structure of the painting is not Napoleon the person, it is not flesh and bone. However, the content of the painting is not entirely divorced from the structure of the painting. If the painting looked nothing like Napoleon, and looked like George Washington instead, we would have a difficult time
saying that it still represented Napoleon. Therefore, structure and content are not identical, but are not completely independent of one another either. Another example of the connection between structure and content of representations can be found in language, which is foundational to the language of thought hypothesis.

The language of thought hypothesis is more specific about the content of thoughts, as well as the relationship that content has with the structure of the thought than the RTM is. The representational theory of mind presents thoughts as information bearing structures of one kind or another, which leaves a lot of room for interpretation. The language of thought hypothesis defines the information bearing structures and content of the RTM using language as a model. Thoughts occur in a mental language with syntax and semantic content. The representational theory of mind (or RTM) does not require thoughts to occur in a mental language. According to the representational theory of mind, thoughts could be representative in the way a painting is, and have no similarity to a language at all. The question is, what does modeling the representative structure of thoughts on a linguistic model of representation buy us?

As we saw earlier in the section on the problem of thinking, materialism needs a way to naturalistically account for content in thinking. The representational theory of mind gives us some idea of how this might be accomplished, but avoids going into specifics. The language of thought hypothesis takes it one-step further by, first, defining the representational structure of thought as a language, and then continues by drawing a parallel between the information-bearing structure of thought and its content, and the connection between syntax and semantics in natural language. Recent
advances in computing allow for this connection between syntax and semantics to be utilized in attempts to understand and model aspects of cognition. These advances set the pieces in place for the creation of a theoretical model for the preservation of content in the process of thinking. In *Psycho Semantics*, Fodor explains “…the recipe for mechanizing rationality: Use a syntactically driven machine to exploit parallelisms between the syntactic and semantic properties of symbols.”

The parallelism between syntax and semantics makes it possible for a formal system to preserve semantic content. The formalist motto regarding language is “If you take care of the syntax of a representational system, its semantics will take care of itself.”. Formal systems can be automated, and a computer is a perfect example of an automated formal system. Before we get into the automation of formal systems, it would be helpful to understand what a formal system is. “A formal system is like a game in which tokens are manipulated according to rules.” A popular example of a formal system is the game of chess. The definition of the formal system or game requires the specification of three things. First, it must be established what the tokens or pieces of the game are. In chess, this would be the pawns, bishops, queen and king and so on. In addition to the type of pieces/tokens, the total number of pieces in the game must be defined. After identifying the types of pieces/tokens involved, the next thing that must be specified is the starting position. Finally, the allowable moves must be specified as well. In specifying which moves are allowed, the allowable positions (given that it is a board game) will be included. Often, there is a goal position specified as well. In the case of chess, this would be placing the opposing players’ king
in checkmate. After these aspects of the game or system have been specified, there are other things that must be noted about formal systems.

The first thing to note about a formal system is that it is entirely self-contained. “Only its own tokens, positions, and moves make any difference to it, and these only insofar as they matter to the application of the rules. In other words, the ‘outside world’ (the weather, the state of the economy, whether the building is on fire, and so on) makes no difference whatsoever in the game.”. As Haugeland notes, there is no connection between the game and the world outside the game. The rules are fixed and unable to be changed by outside influences. This means that basic formal systems like chess have no means of representing anything outside of the system itself.

Another thing to note about formal systems is that every rule, allowable move, or position of the pieces in the game is definite. There are no ambiguities about possible moves or positions. Either a move is allowed or it is not, a bishop in chess can move diagonally and cannot move in straight lines. The same goes for whether or not a piece is in a given position, either it is at a certain space on the board or it is not. There is no “in between” allowed in a game of chess, no piece can be half on one square and half on another. The last noteworthy aspect of a formal system is that there are a finite number of possible moves from each position. In a game of chess, these moves would be quickly calculated given the position of, and allowable moves of the candidate piece. The position and type of all the pieces must be definite; a knight on a particular square can not also be a queen on a different square. The location and type of each piece, as well as all possible, allowable moves are all definite. This is not to say that all the possible, allowable moves are definite in the sense that those moves must be made, rather that they are definitely possible and allowable or definitely not.
Another way of saying that the pieces, rules and allowable moves are
definite is to say that they are digital. Any query about an aspect of the
formal system has a Boolean, yes or no answer; there are no “maybes”
allowed.

As a digital system, the same formal system can be realized in two or
more seemingly different physical structures. Any two formal systems that
follow the same rules and uses the same tokens, despite how different they
may seem, are essentially the same.

“This kind of essential sameness among formal systems is called
formal equivalence. Two formal systems are formally equivalent if they can
be translated back and forth in roughly the following sense:

1. for each position in one system, there is a unique corresponding
   position in the other system;
2. the two starting positions correspond; and
3. whenever you can get from one position to another in one system,
   you can get from the corresponding position to the corresponding position in
   the other system.”xvii.

These definitions given for formal equivalence between two formal
systems ensure that the two are essentially the same: not necessarily the
same in physical structure, but the same in operation. A basic example of
this formal equivalence is the variety of chess sets. It does not matter
whether the pieces of a chess set are made of marble, wood, or people
wearing costumes. The physical makeup of the pieces on a chess set is
unimportant, what matters is what type they are. The pieces in chess are
defined by the rules that apply to them, rather than their structure. Similarly,
the actual size, physical structure and location of the board is unimportant.
As long as the board is a square divided into 64 equally sized smaller
squares in a 8 by 8 allocation, it can serve as a chess board. This means a chessboard can be drawn in the sand, carved out of wood, or manifested on a cathode ray tube. Essentially, formal systems are equivalent if they follow the same rules. The importance of formal equivalence to artificial intelligence projects is that it allows systems to be built using computers rather than brains. The physical make up of a human brain is different from that of a microprocessor. A common intuition is that the flesh and blood that makes up a brain would function in a vastly different way to the silicon and metals that make a microchip, and thus any functional comparison between the two is impossible. Based on this tremendous difference between structures, it appears that any operations that could occur in the brain would never be mirrored in the microchip, and vice versa. However, if both the brain and the computer were realizations of formal systems that followed the same formal operations, they would be formally equivalent.

The game of chess is a good example of a simple formal system, and serves as an easily understandable introduction to formal systems in general. However, the formal system proposed by the language of thought hypothesis is more complex. Two examples of complex formal systems are formal logic and algebra. Each of these complex formal systems differ from simple systems in a manner that is based primarily on their use of tokens. The rules for these complex formal systems include various transformations that are allowable in regards to the tokens. The transformations dictate what new arguments, propositions, or outcomes are allowable. Like the simple formal systems, these systems have tokens that conform to and are identified by the rules they follow. However, unlike the simple formal systems, complex formal systems are capable of creating and operating with an infinite number of different compound tokens. In chess, on each side of the board there are
sixteen pieces (tokens), each piece being one of six different types. As we saw in the paragraph above, these pieces are tokens whose type or identity is defined by the rules. In formal logic or algebra, there are literally an infinite number of equations or arguments that can be constructed from a finite set of tokens and solved. The difference between the simple formal system and the complex formal system is that the complex system also includes compound tokens. The equation “a5 – b=c3” is a token in the system of algebra. The compound token “a5 – b = c3” is different from “a4 – b = c 1”, and “a139–b= c12” and so on and on, the list of different equations is endless. This infinite number of tokens appears to violate one of the principles of formal systems, namely that formal systems are definite. After all, if there are an infinite number of different available compound tokens, and a token is defined by the rules it follows; then it would follow that there would need to be an infinite number of rules for the tokens as well. All of this starts to look very unmanageable, until it is pointed out that the compound tokens are constructs of simple tokens and can be parsed down and further manipulated by the same rules that apply to their constituent simple tokens. An example of a simple token in a complex system is a letter or symbol in an algebraic equation. Like the king in chess, the symbol in the equation is defined by and follows definite rules. Using these simple or atomic tokens, complex formal systems can create compound or “molecular” tokens.

When seen as games, it may be difficult to understand the excitement about formal systems or see a connection between the rules for a game of chess and human cognition. The reason for the excitement is that simple and complex formal systems like chess and algebra can be automated.

“An automatic formal system is a physical device (such as a machine) which automatically manipulates the tokens of some formal system
Recently, there have been great steps forward in the creation and implementation of automatic formal systems. Arguably, the original theoretical model for an automatic formal system was devised by Alan Turing. The Turing machine, as they came to be called, consists of “an unlimited number of storage bins; a finite number of execution units; and one indicator unit. The indicator unit always indicates one execution unit (the active unit), and two storage bins (the ‘in’ and ‘out’ bins, respectively). Each storage bin can contain one formal token (any token, but only one at a time). Each execution has its own particular rule, which it obeys whenever it is the active unit. What that rule specifies will depend on what token is in the current in-bin; and in each case it will specify two things: first, what token to put in the current out-bin (discarding the previous contents, if any), and second, what the indicator unit should indicate next.”

The amazing thing about such a machine, which is responsible for much of the recent excitement in artificial intelligence, is that it can perform operations based on a set of instructions. The rules an automated formal system follows are either algorithmic or heuristic. Algorithmic rules differ from heuristic rules in their nature and application, although both are required to make a system that can play chess. An example of an algorithmic rule is “Bishops can only move diagonally.” Such algorithmic rules can be thought of as the defining rules of the game.
Heuristics are maxims or truisms that are incorporated into formal systems. Returning to chess as an example of a simple formal system, there are definite rules about which moves are allowable and which are not. They are the rules that you would need to include in a computer program which simulated a chess set. They are sufficient for defining the game of chess, but they say nothing about how chess should be played. From here on, the rules that define a formal system will be referred to as simple rules. If one wanted to create an automatic formal system that could play chess, one would need to include more than a list of the pieces and the rules of the game. That is to say, there is a difference between knowing the possible, allowable moves available and knowing which moves would be advantageous or disastrous. Knowledge of the latter variety is created using heuristic rules. In artificial intelligence, the term “heuristic” means essentially the same thing as a “rule of thumb”. Some examples of heuristic rules in chess would be “Exchanging a queen for a pawn is bad.”, “Moving the king early is bad.”, and “Loosing a knight to capture a queen is a good move.” Both heuristic rules and simple rules affect the functioning of the automatic formal system. Again, neither simple nor heuristic rules are influenced by anything external to the formal system. The largest difference between heuristic rules and simple rules is that heuristic rules are not definite in the way simple rules are. Heuristic rules have a ceteris paribus clause, which means that any heuristic rule, such as P, can always be phrased as “All else being equal, then P.”. This is just
saying that it is not always the case, given certain situations, that the heuristic rules should be applied. There may be cases in which a normally helpful heuristic rule is not applicable, or is over-ruled by another heuristic rule. For example, suppose the available moves to a pawn were to take the other players knight, or to take the other players bishop. Suppose that the automatic formal system contains heuristic rules to the effect of “Taking a knight with a pawn is a good move.” And “taking a bishop with a pawn is a good move.” Both of these rules apply to the situation at hand, but making a move with the pawn that complies with both is not possible. The decision about which heuristic rule should be followed, when there is a conflict or choice between two or more rules, like the ones above about the pawn’s possible movements, is a matter of how the system is implemented. You could have a chess-playing program that just randomly performed an allowable move in any given situation, but that would very likely produce a automatic formal system that wasn’t very good at chess at all. A better way to decide on which move should be made in a given situation is through a further analysis of the situation itself. The decision of which possible move to make depends on situational factors, such as what pieces the other player still has on the board, what pieces of yours remain on the board, how the pieces are arranged, whether either move will put the other player in a stronger position, and so on. Once all these factors are taken into account, the proper move is the one that is more likely to lead to eventual completion of a goal. In chess, this goal is to put the other player in checkmate and therefore the most applicable heuristic rule in any given situation in chess is the one that is more likely to result in attaining that goal. The teleological aspect of heuristic rules is what allows for gambits and traps to be utilized by chess playing automated formal systems.
With the algorithmic and heuristic rules in place, and the pieces and moves formalized, the computer is ready to play chess. The chess-playing computer is not in the realm of science fiction; such things are so commonplace that chess programs are often included on new computers for free. On the other hand, the computer that can think is still in the realm of science fiction.

This brings us back to the discussion of the language of thought hypothesis and the problem of thinking. According to the language of thought hypothesis, the brain can be seen as an automated complex formal system. The tokens and rules for the formal system proposed by the language of thought hypothesis are provided by the syntax of the mental language. Hence, the brain can be seen as a syntax driven formal system, similar in structure to the equations and formulas in algebra: thoughts that occur in the mental language are compound tokens. A sentence in language is made up of words, in the same way as an equation in algebra is made up of symbols and operators. Like an equation, there are rules that dictate what order the symbols can be arranged in to form a proper sentence. Syntax defines the word classes of language, as well as governing the grammatical rules for their combination. Whether or not a sentence is “proper” is determined by its grammatical structure, which is determined by syntax.

The combination of simple words/tokens into a sentence is governed by the grammatical rules of syntax. Unlike an algebraic equation however, sentences in natural languages have semantic content as well. Thus far, all the formal systems outlined in this paper have shared a lack of content. There is no aspect or property of the rules and tokens in a formal system like chess that had to do with anything outside the formal system itself. There are kings and queens and knights that exist outside of chess, but the chess pieces
do not represent those people. The only meaning a piece has in chess is in relation to the system itself. What it means for a piece to be a rook in chess is that it can move in certain ways, and cannot move in other ways. Natural language is different from chess, algebra and formal logic, precisely because it contains representational content. Take for example, the sentence, “There is an oak tree in my back yard.” As a compound token, it can be parsed down by an automated formal system to simple, atomic tokens. In virtue of semantic content, the sentence represents, or is about something that is external to the formal system; namely, an oak tree located in my back yard. As was noted earlier in the section about the representational theory of mind, thoughts are information-bearing structures with content. They exhibit a representational content, or an aboutness that is different from their structure. Going back to the representational theory of mind and the painting example, the physical composition of the painting of Napoleon is different from that which it represents (is about), which is Napoleon the person. The sentence is representational, but its representational content is not acquired through similarity. The sentence is not about the oak tree because it looks like an oak tree.

While the representational theory of mind is fairly non-specific about what form the representations take, the language of thought hypothesis insists upon a mental language. One reason for accepting that thoughts occur in a mental language is combinatorial syntax. Combinatorial syntax gives language a range of representation that exceeds the representational ability of similarity. It is possible to think about, or construct sentences about, things that it is impossible to represent with similarity. One example is the sentence, “It is going to snow next week.” You might be able to paint a picture of snow, but how do you paint a picture that represents “next week”?
There are sentences about impossible objects, such as the round square. So, using combinatorial syntax as a framework for mental representations explains the seemingly endless number of thoughts available to us. The number of possible thoughts is arguably infinite, and there is essentially an infinite number of sentences which it is possible to construct and understand using a natural language. By utilizing combinatorial syntax to explicate the form mental representations occur in, we are assured that any sentence we can form in a natural language is also a possible thought. In other words, it would be strange if it were possible to form and understand a sentence in natural language which it was not possible to express as a thought. That is to say, it is hard to believe we can think up sentences about ‘things’ (tomorrow, next week, impossible objects, etc.) which we are unable to represent mentally. The inability to account for much of the content of our thoughts is a serious problem for any theory that wishes to use similarity to explain representational content. This should not be mistaken for a claim that all thoughts it is possible to think must occur in the language of thought. It is possible that some thoughts derive their intentionality from similarity instead of semantic content. These thoughts would include visual-spatial representations and other thoughts that make use of the “mind’s eye”. While it is possible to think these thoughts, it is unlikely that they constitute the entirety of our mental lives, especially the “inner monologue” portion of our mental life. Most, if not all of our abstract reasoning takes place in an inner monologue, and it is this type of thinking that the language of thought attempts to explain.
Another benefit of the language of thought hypothesis is that it accounts for systematicity of thought. “Systematicity of thought consists in the empirical fact that the ability to entertain certain thoughts is intrinsically connected to the ability to entertain certain others.” This is to say, if it is possible for someone to think one thought, it is possible for them also to be able to think several other specific thoughts based on their ability to think the first thought. This can be made clearer by considering some examples of the systematicity of language. If you understand the sentence “Sam is taller than Joe”, you are also able to understand the sentence “Joe is taller than Sam.” This is because what is required to produce and understand certain sentences is also required in the production and understanding of sentences related in the manner demonstrated above. The Language of Thought hypothesis is able to account for the systematicity of thought because language is systematic. Unlike language, pictures or paintings do not demonstrate systematicity. Therefore, a theory of mind that utilized similarity to explain representation would have a difficult time accounting for the systematicity of thought.

Fodor endorses the computational theory of mind, or CTM. In his explanation of the computational theory of mind, Fodor clarifies exactly how he imagines the syntactic formal system he proposes to operate. According to the CTM, mental states and processes are computational. The definition of being computational is that “mental processes have access only to formal (nonsemantic) properties of the mental representations over which they are defined” This is to say, that the operations that constitute mental processes are purely syntactic operations. Fodor calls this the formality condition, further defining the computational theory of mind. Computational processes
are formal because they operate according to the syntax of the representations. As a computational process, a mental process consists of the manipulation of formal symbols according to their syntactic structure. Fodor explains mental processes as having two levels, the hard-science implementation level and the higher plane psychological level. The Computational Theory of Mind for Fodor shows how the connection between syntax and semantics can be exploited by a formal system, as well as drawing a line in the sand. The line in the sand divides semantic content from causality. The causal forces at work in mental processes are all syntactic, and are not causally sensitive to semantic content at all.

What we have so far appears to provide a solution to the problem of thinking. The solution is a materialist solution as well. “…[if] representations form a formal system, i.e. a language with its proper combinatorial syntax (and semantics) and a set of derivations rules formally defined over the syntactic features of those representations (allowing for specific but extremely powerful programs to be written in terms of them), then the problem of thinking, …, can in principle be solved in completely naturalistic terms”xxii. An automatic formal system that operates in such a way that it preserves the coherence of representational content appears to be a solution to the problem of thinking. As was stated earlier, the problem of thinking centered on the difficulty a materialist theory that must operate under physical laws has in accounting for content. For materialism, the goal is to avoid strings of thoughts that ignore the content of the thoughts involved. To attain this goal requires a theory that can account for how physical laws can preserve the representational content of thoughts. The language of thought hypothesis does this by exploiting the parallels that exist between syntax and semantic content. The explanatory use of the parallels in
language is very similar to the dualist theory of parallelism that was outlined in the section on dualism. Parallelism (rather than occasionalism) appealed to the existence of a pre-established harmony between the mental and the physical. God had set both the mental and physical on separate causal paths that gave the illusion of interaction, achieved through harmony. The illusion of interaction, according to the language of thought hypothesis, consists in the natural parallel between syntax and semantic content. This materialist recruitment of pre-established harmony appears to solve a number of difficult problems, yet it raises several more. That there is a harmony between syntax and semantics is taken to be uncontroversial. However, an explanation for how this harmony has come to be is still an issue of heated debate. Thankfully, this paper does not have to concern itself with theories of semantics. Questions about how words come to mean what they do, or how it is that we learn a language, while interesting, are able to be left unanswered for now. The lack of a theory of semantics may prove to be problematic for the language of thought. For our purposes though, it is enough to show that such a harmony exists. A harmony between syntax and semantic does exist, or at least a connection which can be exploited exists between them. How strong the connection between syntax and semantic content is depends on the language in question. Different types of language have differing strengths regarding the connection between their particular syntax and semantics. This thesis will concentrate on a potential major problem for the language of thought hypothesis. The problem of syntactic ambiguity relates to the language of thought’s dependence on the relationship between syntax and semantics.
Chapter Five. Syntactic Ambiguity.

Syntactic ambiguity is a term used to describe multiple possible semantic interpretations of a single token or symbol. In other words, the syntactic form of some symbols allows for several different meanings. This syntactic ambiguity can occur both on an atomic (simple) level, or on a molecular (compound) level.

Syntactic ambiguity exhibited on an atomic level results in a homonym. A homonym is a single word that has multiple possible semantic meanings. Since homonyms are simple tokens rather than compound tokens, they are not so much defined by syntax as morphology. That being the case, it makes sense to talk about the syntactic ambiguity of simple tokens, being single words rather than sentences, as morphological or simple-syntactic ambiguity. Therefore, a homonym is a word that exhibits morphological or simple-syntactic ambiguity. Some examples are the words “Bee”, “Ball”, “Bass”, “do” and many others. There are honeybees, spelling bees, work bees and quilting bees, to name a few. One can play a game with a ball, or go to a fancy dress ball. Finally, there is the musical instrument bass, and the freshwater fish bass and “do” can either be the first note in a diatonic scale or the present tense form of an action verb. The list of written homonyms is short in comparison to the list of spoken or aural homonyms, also called homophones. These are words which sound identical when spoken, but are spelled differently and have different meanings. An example would be “Ant”
and “Aunt”, or similarly, “Ante” and “Auntie”. The word “ant”, which refers to the tiny insect sounds exactly like “Aunt”, which is a familial relation of your parent’s sister. Homophones, while interesting, can be differentiated by the spelling, or morphology of the words involved. The multiple semantic meanings of the word “bee” or “ball” however, cannot be so easily specified. The semantic content of words like bee or ball can only be determined by reference to the context in which the word occurs. When you encounter the word “Bee” in a sentence, you can discern if it refers to a spelling bee or a honeybee using the other words in the sentence; or if that fails, the previous or further sentences. For example, “I was stung by a bee today.”, or “I heard the sounds of a bee in the village center. The grandmothers were all intently sewing quilts and talking shop at the quilting bee.” and finally “I am very gifted at spelling. I won the bee that was held at my high school.” In the first example, the semantic content of the word “bee” is determined by the other words that constitute the sentence in which it appears. In the second example, the semantic content is determined not by the contextual clues offered by the words in the sentence the word appears in, but in the following sentence. The third example shows a situation in which the meaning of a homonym depends on the previous sentence. In all three examples, the prescription for determining the meaning of a syntactically ambiguous atomic token (word), lies in an examination of the surrounding tokens (words that make up the sentence). In either the immediate sentence in which it appears, or in those directly related to it. This is called a contextual definition. This prescription for the semantic determination of syntactic ambiguity also applies to compound tokens. The contextual relationships for compound tokens (sentences) are expectedly more complex than those involved in simple tokens (words).
A compound token that exhibits syntactic ambiguity is a sentence that is properly constructed according to the rules of syntax and has multiple semantic interpretations for either the sentence as a whole or a word in the sentence. This is the second form of syntactic ambiguity, in which a word in the sentence has multiple semantic interpretations. While this situation may initially appear to be another instancing of a homonym, this is not the case. Take these two sentences for example, the first being “We fed the squirrels the sandwiches because they were hungry.” And, “We fed the squirrels the sandwiches because they were stale.” Notice that in both sentences the word “they” could mean either the squirrels or the sandwiches. The word “they” is not a commonly accepted homonym. There are not two or more meanings for “They” in the same way there are for “Bee” or “Bass”. In the case of a classic homonym, like “Bee”, the possible semantic interpretations are already fixed and understood. In a sense, a homonym is like two words that each have a fixed semantic interpretation, which happen to share the same morphological form. In the case of a homonym, the context of the surrounding discourse decides the semantic content to be applied to, or understood by its instancing in the particular sentence. In the case of the multiple possible semantic interpretations of the word “they” in the two sentences above, something slightly different is going on. The word “they” is a pronoun, which like “I”, “he”, “it”, and “them” is used to refer to a noun which has already been specified. Unlike a homonym, the possible semantic interpretations of a pronoun are not determined until the pronoun is instanced in a sentence. While the instancing of a homonym in a sentence is also required for the semantic content of the homonym to be determined, the possible semantic interpretations of a homonym are already defined. So
while both a homonym and a pronoun have their semantic content determined by the context of the discourse in which they are instanced, the possible range of semantic interpretations of a pronoun is far greater than that of a homonym.

In the example of two sentences above, the pronoun “they” exhibited syntactic ambiguity. In both sentences, either the pronoun referred to the squirrels or to the sandwiches, but given the context, a pronoun can refer to nearly anything. The only limitations to the list of meanings for pronouns are categorical. Some pronouns can only refer to certain categories of nouns, for instance the pronoun “Them” refers to plural nouns rather than singular nouns. In this case, plural nouns and singular nouns are taken to mean the plural and singular forms of a noun (Apples or an apple), rather than the actual instancing of one or many words. These categories do not define the meaning of the pronoun; they define the form (plural or singular) of the noun the pronoun refers to. The categorical definition of form does very little to pare down the list of possible nouns, or as is the case, possible semantic interpretations of a pronoun.

The discourse that surrounds a homonym is often sufficient to determine its meaning (semantic content). This is not the case with pronouns, which determine their meaning through the semantics of the surrounding discourse. In the example of the two sentences about the squirrels and the sandwiches, the syntactic structure of the sentence is not sufficient to determine the semantic interpretation of the pronoun “they”. In both sentences, “they” could refer to either the squirrels or the sandwiches. Only an understanding of the meaning of “hungry” and “stale”, and knowing what “squirrels” and “sandwiches” are is sufficient to determine
which noun that the pronoun “they” refers to. That is to say, one has to know that sandwiches do not feel hungry or that squirrels don’t go stale.

The form of syntactic ambiguity that affects an entire sentence, rather than a single word, resembles the ambiguity of the sort exhibited by pronouns rather than homonyms. That is to say, the ambiguity of compound tokens (sentences rather than words), cannot be resolved using syntax. A popular example of compound token syntactic ambiguity is the sentence “Time flies like an arrow.” The sentence, “Time flies like an arrow.”, can be interpreted in several different ways. It could mean that “Time progresses in a swift, linear fashion like the flight of an arrow shot from a bow”. Another possible interpretation of its semantic content would be “Measure the speed of flies as you would measure the speed of an arrow.” A third possibility would be “A variety of flies, time flies, enjoy arrows.”, as is evidenced in the word play on this sentence commonly attributed to Groucho Marx, “Time flies like an arrow, Fruit flies like a banana.”. In cases like this, the syntactic ambiguity extends to encompass the semantics of the entire compound token (sentence), rather than a atomic constituent (single word) of the compound token. Like the syntactic ambiguity of the pronoun, the semantic content of this sentence cannot be determined by its syntax alone. An understanding of the meaning of the words “time”, “flies”, and “arrow” is required to determine which of the possible meanings of the sentence as a whole is intended.

To get back to our topic, syntactic ambiguity is troubling for a syntactic engine that proposes to account for semantic content. A syntactic engine, as defined earlier, is an automatic formal system. As a formal system, a syntactic engine subscribes to the formalist theory of language.
The formalist theory of language can be summed up by the formalist motto, which is “If you take care of the syntax of a representational system, its semantics will take care of itself”. However, in the case of syntactic ambiguity, this is not the case. In situations like those demonstrated above, the syntactic structure alone is unable to account for the semantic content of pronouns or entire sentences. The syntactic engine is only able to overcome the problem of thinking (i.e. the apparent incompatibility of mental content and physical causal laws), as long as the parallels between syntax and semantics are as definite as the formalist motto says that they are. Syntactic ambiguity, as evidenced in the examples involving the pronouns and ambiguous sentences above, shows that the parallels between syntax and semantics are looser than the formalists insist they are. If the relation between syntax and semantics is not sufficient for syntax to determine semantics in all cases, and the examples above show that this is the case, then the success of a syntactic engine hypothesis in offering a solution to the problem of thinking is jeopardized. We have seen that the only way to determine the semantic content of a syntactically ambiguous simple or compound token, is using context. The effect of this is that if the brain is a purely syntactic engine, it is entirely plausible for it to create a chain of thoughts that do not exhibit content coherence. That is to say, the content of the thoughts would not make sense in regards to one another. If the formal syntactic system wrongly interpreted a syntactically ambiguous thought, the following thoughts in the chain of thoughts would be jeopardized in regards to their content coherence. Taking the “Time flies like an arrow.” example from above, a incorrect interpretation of that sentence could cause a chain of thoughts which wouldn’t make sense.
Certainly, a chain of thoughts created by a syntactic engine would be more likely to exhibit content coherence than a randomly assembled string of thoughts. Given the parallels between syntax and semantics, no matter how tight or loose they actually are, the chain of thoughts produced by a syntactic engine would exhibit more content coherence over time than random strings of thoughts. However, likelihood is not enough, given the incredible subject range and number of the thoughts we experience. To be able to explain the content coherence produced by the process of thinking requires more than probability. Otherwise, it could be argued that our thoughts are in fact random, and it is only by the most incredible coincidence that they appear to respect the content of one another in any way.

One reply to the problem of syntactic ambiguity in cognition is to say that the language of thought is not a natural language, but another kind of language instead. The problem with syntactic ambiguity in thoughts is that the connection between syntax and semantics in natural language is too loose to explain our experience of content coherence. This loose connection between syntax and semantics, allowing as it does for multiple semantic interpretations of the same syntactic structure, is a problem for a syntactic engine intended to preserve semantic coherence. One way around this problem is to deny that the language of thought is a natural language, and therefore does not suffer from syntactic ambiguity. Instead of being a natural language, the language of thought would be a language in which there is no possibility for multiple semantic interpretations of a single syntactic
structure. Instead of natural language, thoughts would occur in a language that exhibits a tight connection between syntax and semantic interpretation.

There are languages that demonstrate this tight connection between syntax and semantics. An example of such a language is mathematics. There are no homonyms in math, like there are in natural language. While “Bee” may refer to a spelling bee, or a honeybee depending on the context in which it occurs, the numeral “4” will mean “4” in all cases. External context plays no role in determining the meaning of the number four. In the language of mathematics, the semantic interpretation of “2 + 2 = 4” is fixed independently of outside factors: “2 + 2 = 4” is semantically evaluable regardless of context. This is because the division between syntax and semantics in languages like mathematics is almost non-existent. If the language of thought had the kind of connection between syntax and semantics that is evidenced by the language of mathematics, it would be free from objections about syntactic ambiguity. The only problem with the language of mathematics is that it is absurd to suggest that math would be able to account for the semantic content of our thoughts. The challenge then becomes to find an example of a language that is safe from syntactic ambiguity, as well as being able to account for the semantic range of our thoughts. In the face of this challenge, the commonly accepted candidate is machine language.

The most popular example of machine language is utilized by computers as the symbols that make up the instruction set for their actions. Machine language is a contender because like mathematics, it is a language that does not suffer from syntactic ambiguity. Furthermore, machine
language has an operational semantics, the syntax of machine language has actions and operations as its semantic content, which is why it is also commonly referred to as a ‘command language’\textsuperscript{xxiii}. Machine language is a command language because of its operational semantic content, the meaning of machine language for a computer is an instruction for what action to perform. It is operational semantic content that forms the basis for computers being capable of intelligent action. The \textit{machine-language} proponents version of the language of thought claims that Mentalese, the language of thought is not a natural language, but an innate biological machine language of the brain. There are two main arguments for accepting that Mentalese is a non-natural, machine language. The first is a comparison between human and animal cognition. Animals like dogs and apes appear to be capable of thought, but also appear to lack the use of natural language. If natural language was necessarily the domain of thought, then it would appear to exclude animals from being able to think. This is not a claim many people wish to make. While this is a valid objection to animal cognition occurring in a natural language, it is not obvious that it applies to human cognition. In fact, the apparently unique ability that humans have to use natural language may signal a telling difference between animal and human cognition. A possible response to this objection by those who champion natural language as the language of thought is to identify the type of thought in question. It is the case that we want to be able to say that animals can think, but we do not necessarily want to say that they think at human levels of cognition. The reply to the animal objection in this case is that animals think, but not at the same level as people do. The possession of both natural language and higher cognitive abilities is unique to humans, and that can be
taken to mean that the two are related. In this case, natural language is the medium in which higher level cognition takes place.

There is another objection against natural language being the language of thought that is essentially an appeal to simplicity. Natural languages are sprawling, complex systems when compared to the concise structure of machine language. To quote David Cole, “As far as we know, machines can potentially produce all the overt linguistic behavior of humans, and so provide all the evidence of understanding, cognitive capacity, and inference potency of humans….A grammar for English, and a vocabulary larger than any human English speaker, along with orthography and lexical definitions, can sit in an off the shelf laptop computer. The reason is this: we can build representations, models, of just about anything, including the syntax and semantics of English, from a very small set of logical primitives. Recursion makes simple operations go a (very) long way. So I think we can concede the nativist point, put this backhand way: Humans take linguistic strings as input and produce linguistic strings as output. They must have the native capacity to do that. They must have the native capacity to represent the rules by which those transformations take place -- that is, they must be able to produce the (possibly rule-governed) syntactic operations that convert linguistic input into output. But there is no evidence so far that this involves more than a handful of basic string manipulation operations -- just as in computers. All that must be represented is a series of syntactic steps that will transform string A into String B.”
The point is that natural language is unnecessarily complicated and large when compared with machine language. All else being equal, machine language is a simpler explanation for the mechanics of thought than trying to include some form of natural language. All else being equal, a simpler explanation that works equally well is preferable. From a Nativist standpoint, the psychological operations of the human mind can be explained with machine language and compilation. In other words, although it may appear that our mind operates according to a natural language of thought, it does not. The natural language in which we experience our inner monologue of thoughts is compiled into a machine language, which is actually doing all the work. An analogy can be drawn with differing levels of programming languages on computers. The most common method of writing programs on a computer is to use a high level language, such as C++ or Java, which are more verbose and intuitive. Rather than typing in a string of ones and zeros, the programmer can write a command statement that is later compiled into the appropriate machine code by a compiler. Only in the case of the mind, the thinker would experience their thoughts in the more verbose and intuitive natural language that would be compiled down into the brains own machine language. Machine language has the added bonus of having an obvious operational semantic content. If our materialist viewpoint has lead us to view the brain as a machine, then all the more fitting for it to operate on something resembling machine language.

Unfortunately for the machine language proposition, there are problems for this theory to overcome. While using machine language as a theoretical model for the language of thought gets us around objections raised by syntactic ambiguity and simplicity, we are not out of the woods
just yet, as there are also several arguments against machine language acting as the language of thought. To start with, there are objections that are raised by experimental, empirical observation. Psychological experiments on patients who suffer from aphasia and other cognitive abnormalities, as well non-affected patients show that events that affect or impair a persons natural language use similarly affect that person’s ability to perform mental tasks. In other words, there are many examples linking natural language with thinking. \textsuperscript{xxv} Either the results of these experiments are being wildly misinterpreted, or natural language is the domain in which higher level, abstract thought occurs. “direct tests of (limited forms of) our hypothesis – that natural language is the domain for thought- have now begun to be conducted. The most important of these is Hermer-Vazquez et al. (1999), which provides strong evidence that the integration of geometric properties with other sorts of information (color, smell, patterning, etc.) is dependent upon natural language.”\textsuperscript{xxvi}

To outright deny that the results of the psychological experiments are applicable is a possible option for the machine language proponent, but it isn’t a very attractive one.

Another objection to machine language acting as the language of thought stems from the semantic poverty of machine language. Unlike a natural language, which has a remarkable breadth of representational possibilities, machine language is severely limited in its ability to represent. The operational semantics of machine language are responsible for the semantic or representational poverty that the language suffers from. The root of the problem is that machine language can only represent internal registers of the machine itself. Take the word “Tree” for example. Suppose that a
computer has the word “Tree” in one of its internal memory registers, and for the definition of the word “Tree”, it has links to other internal memory registers. In the other memory registers you might find words like “Plant”, “Green”, “Bark”, “Wood”, “Roots”, “Leaves”, and so on, words which describe or relate in some way to the word “Tree”. The problem is that the computer does not represent or mean “Tree” when it uses the word ‘Tree’. The word “Tree”, once properly interpreted for the computer means “Internal memory address where the word ‘Tree’ is stored.” Rather than meaning “Tree” as you or I do. The reason for this is that although machine language does contain a syntax and semantic content, both are limited to commands for the machine. Machine language is a command language, words in machine language are commands for the machine to perform operations. Meaning in machine language begins and ends with commands and instructions. There is no foreseeable way for machine language to bridge the gap between syntax and semantic representations that are external to the machine itself. The semantic poverty of machine language makes it an unlikely contender to explain the functions and abilities of the human mind.
Chapter Six. The Thought Experiment.

Fortunately, there is another option available that is compatible with the empirical evidence from the experiments, and avoids the problem of syntactic ambiguity as well. The thought experiment below serves as an introduction to the concept of consciousness feedback by examining aspects of the relationship between consciousness and cognition.

The following thought experiment is designed to create an experience of thoughts that have separate causal origins. The two causal origins of the thoughts in the experiment are consciousness and sub-consciousness. The experiment is as follows.

“The first section of this experiment involves thoughts which are brought about by the conscious will of the thinker. It may help to turn off any music and close your eyes for this experiment, so as to allow yourself to focus on the experience without excess outside distraction. After you have cut down on the possible outside distractions, close your eyes and think the sentence ‘There is an oak tree in my back yard.’. Repeat this sentence to yourself a number of times, do not say it out loud, just think it a few times. Try to picture the words as you are thinking them, so you can see the sentence in your “minds eye”. It does not matter if you can’t picture the entire sentence at one time, as long as you can picture the words as you are thinking them. Finally, make sure that during the time you’re thinking “There is an oak tree in my back yard.” you pay attention to any aspect of the process that seems different. If there is any extra difficulty, or the thought itself is different to a normal thought, or anything along those lines
where this experience is not the same as what it is normally like for you to think.” That is the end of the first half of the thought experiment. It makes no difference if you repeat this portion of the experiment several times, or only do it once. As long as you have a solid grasp on any similarities or differences between this experience and everyday thought, you can move on to the second half of the thought experiment.

“The second half of the thought experiment involves thoughts which come about in spite of conscious effort. As in the last experiment, it may help to close your eyes and have any music turned off: Once you have closed your eyes, your goal is to stop thinking about anything at all. Try to keep your mind a complete blank for at least a full minute. This can be very difficult, so you may have to attempt it several times. When thoughts do arise, try not to follow them up with further thoughts, instead try to turn your mind back towards not thinking about anything at all. Do not picture anything, or think of what you have to do today or tomorrow, or even think about how this is different from everyday thought; just try and not think anything at all for at least a minute. Do not read any further in this thought experiment until you have attempted to keep your mind free from thoughts at least three or four times.”

To think about nothing for a full minute is a difficult task, and thankfully this thought experiment does not require you to be able to successfully keep your mind clear for a full minute. The important thing to notice about this portion of the thought experiment is that thoughts occurred without you consciously willing them to occur. In fact, thoughts occurred in spite of your conscious efforts against them occurring. This assumes that the
outcome of your participation in this section of the thought experiment resulted in thoughts occurring against your will.

The two halves of this thought experiment are intended to demonstrate the difference between thoughts with a conscious origin and those that originate unconsciously. Interpreting thought experiments like the one above can be difficult as there are multiple possible conclusions one can reach. For instance, it may be that the thought “There is an oak tree in my back yard.”, was caused by consciously willing it to. It could also be the case that the thought “There is an oak tree in my back yard.”, has a causal explanation which has nothing to do with consciousness.

The conclusion that can be drawn from the second half of the experiment is less debatable. Despite the fact that you were trying not to think of anything, thoughts continued to occur. While the effect of conscious attention can be debated in the first half of the example, it appears that little is open to debate in the second half. None of the thoughts which occurred during this portion of the thought experiment were the result of consciousness in the manner that thoughts which occurred in the first half were. There is a distinction between the possible outcomes of both halves of the thought experiment. For the first half of the thought experiment, the result is somewhat inconclusive. As we have only our experience to go on, there is no way which we can know for sure what the actual causes are for the thought “There is an oak tree in my back yard.”. The question is “Is the thought in the example caused by syntactic (formal, computational, unconscious) processes, or semantic (content based, conscious) processes?”.

Thankfully, there are only a few possible conclusions that can be attained
from this portion of the experiment. Dealing with ontology first, either the mind is physical or non-physical. If it is non-physical, then the processes that are responsible for the thought are equally non-physical. If this is the case, then the question of whether or not the causes are formal, physical causes does not apply. If the mind is physical, then the processes responsible for the thought in first half of the experiment are also physical. Proceeding with the theory of the brain being a formal symbol manipulator, this leaves four possible options available. The processes responsible for the thought in the first example are purely syntactic, purely semantic, a mix of both, or neither syntactic nor semantic processes. If the brain is a kind of formal symbol manipulator with semantic content, then we can safely eliminate the possibility of it being neither semantic nor syntactic properties responsible, if we take this to mean that the causally important properties are neither formal nor semantic. If the processes responsible are not syntactic, then the brain does not function as a formal system, and if they (the mental processes) are not influenced by semantic content they offer nothing to explain our experience of thought.

The three options remaining are purely syntactic, purely semantic, or a mixture of the two. The purely semantic option is incompatible with the idea of a formal system, as semantic content is not formal. The second half of the experiment also suggests that human cognition is not a process that operates purely on semantic content. If mental processes are governed purely by the semantic content of each thought, then we should have no trouble with stopping the production of thoughts. A thought with the semantic content “No thought should follow from this” or something similar would be sufficient to stop thinking if the process of thinking was sensitive to semantic content only. The goal of not thinking about anything would be far
more easily attainable, and no thoughts would occur against our will. More precisely, no thinking would occur that was not caused by the semantic content of previous thoughts. Furthermore, there would be far less of a phenomenal difference between our everyday process of thinking and the process of thinking experienced in the first section of the experiment. If the process of thinking is governed by the content of thoughts (being defined as our conscious experience of the thoughts), then our everyday experience of thinking should more closely resemble the first portion of the experiment.

The purely syntactic option is also problematic. As we saw above, a formal system operating on a purely syntactic basis must answer the problem of syntactic ambiguity. In answering the problem of syntactic ambiguity, it looks as though one must pick between the failings of a natural language and the failings of a machine language. Neither one is without its flaws, and therefore the problem of syntactic ambiguity remains a significant challenge for the purely syntactic formal system to overcome.

It appears that the theories that rely on purely syntactic, purely semantic, or the total absence of syntax and semantics, are all flawed. This leaves the theory of a mix of syntactic and semantic causation responsible for the process of thinking. The combination of syntax and semantics could be achieved using consciousness and feedback. Cognitive science has very little to say about consciousness. Where it is, and what it is, and what it is for are all still very much open to interpretation and debate. In fact, due to this lack of consensus, the topic tends to be avoided, especially when talking about artificial intelligence and cognitive systems. This is evidenced in one of the most famous tests for artificial intelligence, devised by Alan Turing,
called the Turing test. Roughly, to pass the test and be declared intelligent a machine must be able to converse with a human being at a level at which it is indistinguishable from another human. That is to say, a person has a conversation with the machine and cannot tell that she is not talking to another person, and is in fact talking to a machine. This test makes no mention of consciousness, nor does it claim to be able to detect it. In fact, consciousness is notoriously difficult to test for, some would say that it is actually impossible to devise a test for consciousness.

As we saw in our brief history of the mind-body problem, a central issue for philosophy of mind and cognitive science is the logistics of interaction. Although the language of thought hypothesis has shifted the concepts of body and mind onto syntax and semantics, the essence of the problem of interaction remains the same. If semantic content has a causal role to play, then what kind of rules or laws are up to the task of integrating content and causation? In *Psychosemantics*, Fodor presents several arguments against content based causation, by which I mean causation that is sensitive to content. In this instance, Fodor takes the phrases “intentional content” to mean essentially the same thing as “semantic content”, the meaning or “aboutness” of a mental state. “The moral is that even though it is true that psychological laws generally pick out the mental states that they apply to by specifying the intentional contents of the states, it doesn’t follow that intentional properties figure in psychological mechanisms.”\textsuperscript{xxvii} He provides reasons why semantic content cannot be included in the causal laws that are responsible for mental processes. The most convincing and applicable of the reasons he provides against content playing a causal role is the one he calls the technical reason. The technical reason is this, “If
thoughts have their causal roles in virtue of their contents per se, then two thoughts with identical contents ought to be identical in their causal roles. And we know that this is wrong; we know that causal roles slice things thinner than contents do.” xxviii. In other words, if the (intentional, semantic) content of thoughts was responsible for the causal role the thought played, there would not be enough variation in possible causal outcomes to account for our mental experience of thinking. If thoughts are causal in regards to their content, a thought with a particular content would always cause the same proceeding thought. That is to say, suppose that a thought with the content “My sister likes strawberry ice cream.” is causally responsible for the next thought, which is “Maybe I should get her some strawberry ice cream today.”. The mechanism responsible for the production of thoughts is causally sensitive to content only, and operates in such a way that when a thought with the content “My sister likes strawberry ice cream.”, is produced, the next thought produced is “Maybe I should get her some strawberry ice cream today.”.

In the example above, the content of the thoughts is responsible for the causal connection between them. Then all else being equal, a thought with the content “My sister likes strawberry ice cream.” will always cause a thought with the content “Maybe I should get her some strawberry ice cream today.”. Let us call the thought with the content, “My sister likes strawberry ice cream.” thought A; and the thought with the content “Maybe I should get her some strawberry ice cream today.” thought B. The causal relationship between thought A and thought B is that for every instance of A, it causes B. Yet we know from our own experience that this is not the case with thoughts. “The thought that - - P, for example, has the same content as the thought P on any notion of content that I can imagine defending; but the
effects of entertaining these thoughts are nevertheless not guaranteed to be the same.” xxix The causal relationship between thoughts is not that strict. There would be some very serious problems for mentality if they were. Suppose that thought A always causes thought B in virtue of their content based causal properties. Thought B would then cause another thought, we will call it thought C. Thought C causes thought D and so on, and on. This does not pose a problem until it is pointed out that we often entertain the same thought more than once, or at least we think several thoughts with identical content. For example, how many times have you thought “Where are my keys?”, or “I am late!”, or some other thought which you have thought on more than one occasion. If the relationship between thoughts is such that thought A always causes thought B, and B always then causes C, and so on, then thinking the same thought (or a thought with the same content) twice would create an inescapable loop of thoughts. Suppose that A causes B, which causes C, and thus a causal chain of thoughts is created. If this causal chain of thoughts causes thought A to occur a second time, the chain becomes a never ending loop of the same thoughts. We know that this is not the case, because we can think the same thought twice without being drawn into an eternally repeating chain of thoughts. As Fodor puts it, “Take a mental life in which the thought that P & Q(P → Q) immediately and spontaneously gives rise to the thought that Q; there is no guarantee that the thought that - - P & (P → Q) immediately and spontaneously gives rise to the thought that Q in that mental life.” xxx That is to say, our experience of thinking is not such that thinking a particular thought must always give rise to the same following thought in all situations. For example, take the thinking of thoughts A and B, the content of A being “I like pie.” And the content of thought B being “Pie is very
tasty.”. Suppose that in one instance, the thinking of thought A is followed by the thinking of thought B, but B following A is not a strict formula that applies to all situations. The next time thought A occurs, any number of other thoughts may follow it. For thought B to always follow thought A, the process of thinking would have to take place in a vacuum. That is to say, the process of thinking would have no input or influence from anything other than the process of thinking itself. We know this is not the case, for example if the thinker was struck with a stone as he was thinking thought A then it would be unlikely that his next thought would be thought B. It would more likely be thought Z, whose content is “Ouch.”.

Fodor’s argument against content causation is compelling, and makes a strong case against content playing a law-like role in the causal operations that govern the psychological mechanism responsible for mental processes. However, his argument is only against strict, causal, lawful connections between thoughts in virtue of their intentional content, and ignores the argument that the causal laws responsible for the production of thoughts would be sensitive to anything in addition to thought content. That is to say, while you may have thought the same thought a number of times, you have never thought that thought in exactly the same situation twice. Therefore, while you may have thought “I like pie.” on several occasions, the environmental variables in each occasion have been different from one another. Since the environment the thinker is located in changes continually, the causal determinates of the situation are not identical and we are not in danger of falling into the closed loop of thoughts. In other words, the semantic content of a thought may be identical with another thought, but because of the environmental variables that also factor into the process of thinking we are not in danger of the closed loop of thoughts. Since the
process of thinking is sensitive to environmental factors, like the thinker getting hit with a rock or seeing a beautiful sunset, or any number of other distracting events, the inclusion of content into the mechanisms of the process of thinking does not put the process in danger of a closed loop. Only the inclusion of content into the causal mechanisms of the process of thinking and the exclusion of all else threatens to create the thought loop.

The success of approaching a model of cognition using formal systems and language of thought rests upon the idea that there is a connection between syntax and semantics. As stated earlier, this approach appears to be a materialist rehash of parallelism that utilizes the harmony between syntax and semantics. The problems start to occur when one is pressed to say how closely the syntax and semantic content of the language of thought are tied together. This question draws us into the debate about whether the language of thought is a machine language or a natural language. The machine language supporters think that semantic poverty is less of a problem than syntactic ambiguity, and that syntax will ultimately trump semantic content. The natural language supporters think that syntactic ambiguity is not actually as big a problem as it appears to be, and that semantic content might hold some causal weight in the cognitive machine. Meanwhile the critics of the language of thought hypothesis in general think that both of the language options are wrong.
Chapter Seven. Feedback Explained.

The central claim of this paper, though it has taken some time to arrive, is that the answer to the natural/machine language of thought dilemma lies in a combination of machine and natural language in cognition. The combination of machine and natural language is accomplished by the causal weighting of both syntax and semantic content. The mechanism through which this is to occur is called feedback.

Feedback is a way of self-regulating a system. In our own bodies we have many feedback systems currently operating which are vital to keeping us alive. In biology, feedback is used primarily by living systems to achieve homeostasis. Feedback is essentially the return of a systems output into the subsequent input of the same system. An example of feedback in nature would be the temperature control of our bodies. Our bodies need to stay within a very small band of allowable temperatures to continue to function properly or survive. Due to the constantly changing temperatures of the environment, our bodies require a system that is sensitive to changes in both the internal and external temperature. A feedback system consists of receptors and effectors. The receptors sense a need to create a change in the system and send a message to the effectors, which are responsible for creating the necessary change. A simple example would be the thermostat on an oven. If you set your oven to 200 degrees in order to bake a cake, you expect it to heat itself to 200 degrees and remain at that temperature. In the case of the oven, the receptor would be the thermometer and the effectors would be the heating element. Until the thermometer reached 200 degrees, the heating element would continue to heat the oven. Once the oven reached
200 degrees, any fluctuation in temperature would be sensed by the receptors (thermostat) and corrected for by the effectors (turning the heating element on, or leaving it off completely). Without feedback, it would be impossible for the oven to self-regulate its own temperature.

In the case of feedback in cognition, consciousness would act as the receptor in the system. The effectors would be heuristic rules initiated by the conscious receptors. The purpose of including consciousness feedback in a formal system is to give some causal weight to the semantic content of a thought without using the syntactic, algorithmic rules. Attempting to integrate semantic content into algorithmic rules is what leads to the problems with syntactic ambiguity and machine language in the first place. Consciousness is arguably the only acceptable candidate for the role of semantic receptor in the cognitive system, as it has direct access to semantic content. Conscious experience is currently the only way that we can apprehend the semantic content of a thought.

For clarity and ease of description, this example assumes that the brain is modular. Although not modular in the strict, Fodorian sense, where a module in the mind is a domain specific, informationally encapsulated, data processor. I take “modular” to mean here that the brain has areas that fulfill specific functions and that can interface with certain other modules. A module would be an area of the brain, physical and/or conceptual, which performs a specific function. There are two modules involved in the production of a chain of thoughts, a thought production module and a module responsible for the conscious experience of thoughts. The production of a chain of thoughts begins with the module responsible for the production of thoughts. The module responsible for the production of
thoughts is an automated formal system, a symbol manipulating system operating with a formalized language of thought. The thought producing module produces thoughts without sensitivity to the semantic content of the thought being produced, it operates on a purely syntactic level. This module produces a thought, which is then experienced by the module responsible for the conscious experience of a thought.

The module responsible for the conscious experience of thoughts has access to the semantic content of a thought. The role this module plays in the production of a chain of thoughts is to fix or set the semantic content of any ambiguous thoughts. This is accomplished by interpreting the ambiguous syntactic structure with a definite conscious experience. As shown earlier, if a syntactically ambiguous thought is dealt with on a purely syntactic level, it has the danger of producing a chain of thoughts that does not exhibit content coherence. The conscious experience of a thought fixes its semantic content, so that it is no longer ambivalent. Once the semantic content of a thought has been fixed, the production of the next thought is influenced by heuristic rules that differentiate between possible thoughts to be produced based on the semantic interpretation provided by the conscious experience module. An analogy would be the operations taken by a chess-playing program in deciding the most appropriate move to make.

So the suggestion is that the chain of thoughts begins with the production of a thought. The production of thoughts is separate from the experience of thoughts. The experience of a thought occurs after it has been produced. The forces at play on the production of a thought are primarily algorithmic. Once the thought has been produced, the portion of the mind
responsible for conscious experience experiences it. The conscious experience of the thought serves to provide a single semantic interpretation of any syntactically ambiguous thought. While there may be many possible semantic interpretations of a syntactically ambiguous thought, conscious experience interprets an ambiguous thought into a single experience. The conscious experience of an ambiguous thought provides access to only one of the potential or possible semantic interpretations of a syntactically ambiguous thought at a time. While it may be possible to have multiple thoughts occurring at the same time, or the same thought occurring at different times, it is not possible to consciously experience the same thought with different interpretations of its semantic content at the same time. This is because the conscious experience of a thought is an experience of the semantic content of the thought, and is therefore an interpretation of an ambiguous thought. It may be possible to experience many different semantic interpretations of an ambiguous thought, one interpretation following the next very rapidly. For example, when one is trying to figure out the meaning of a curious sentence someone has uttered in conversation, or read a phrase in a book that does not make sense, then one might experience many possible interpretations of the sentence or phrase. However, this does is not the same as experiencing the same thought in several different ways at the same time. Which of the possible semantic interpretations of an ambiguous thought is correct is determined by context. Conscious experience has access to context, through the memory of the semantic content of previous thoughts. Deciding the semantic interpretation of a syntactically ambiguous thought allows the cognitive system to integrate semantic content into the production of new thoughts. As a thought is experienced, the semantic content of that
thought will be fixed by the experience of the thought. A thought which is ambiguous if processed by its syntactic structure alone, for example “Time flies like an arrow.”, is not ambiguous to conscious experience and context. Having access to the proper semantic interpretation of syntactically ambiguous thoughts allows the cognitive system to preserve content coherence in chains of thought. After the first thought is produced and then experienced, another thought must be produced. To preserve content coherence, the production of the new thought must respect the semantic content of the previous thoughts in one way or another.

In the production of the next thought, consciousness feedback operates in a similar way to a chess-playing program. As discussed earlier in the paper, a chess-playing program utilizes two kinds of rule, algorithmic and heuristic. In a chess-playing program, the heuristic rules dictate which move should be taken depending on the current location of the pieces on the board. In the cognitive system operating with consciousness feedback, the idea is that it is the heuristic rules that dictate which thought should be produced, this depending on the semantic content of the previous thoughts. Based on the layout of the pieces and the previous moves made, the heuristic rules of a chess-playing program decide what move would be best to make. A cognitive system utilizing consciousness feedback would use the semantic content of the previous thoughts to decide which of the possible thoughts would be the best to produce. For example, suppose you began a chain of thoughts with the thought, “Time flies like an arrow.”. For the sake of simplicity, lets say that there are three possible thoughts to be produced following that thought, possible thought 1, possible thought 2, and possible thought 3. Possible thought 1 is “Time fly bites don’t itch as bad as sand fly bites.”, possible thought 2 is “Timing a fly would be harder than timing an
arrow.” And possible thought 3 is “Time passes so quickly, where has the year gone?”. The possible thoughts all differ on the basis of the interpretation of the semantic content of the thought, “Time flies like an arrow.”. The three possible thoughts listed above are all progressions from differing possible semantic interpretations of the syntax of the already produced thought. Which of the three possible thoughts is produced next relies on the conscious experience, and therefore semantic interpretation of “Time flies like an arrow.”

To summarize: the semantic content of the current and previous thoughts provide the heuristic rules at work in the mind with the necessary information to decide on which possible thought to produce next. The actual production of the next thought as well as the possible thoughts to be produced would be a process governed almost entirely by algorithmic rules, rather than heuristic rules. The function of the algorithmic and heuristic rules at work in the mind is similar to the function of the algorithmic and heuristic rules at work in a chess-playing computer. The possible thoughts and possible moves are both the result of formal, algorithmic operations. The decision about which possible move/possible thought is the best in a given situation is the result of the heuristic operations. The conscious access to the semantic content of thoughts serves a similar role in the mind as access to the placement of the pieces on the board serves in the chess-playing program. A formal system that suffers from syntactic ambiguity cannot preserve content coherence between thoughts. Conscious feedback avoids syntactic ambiguity by giving a formal system access to definite semantic content. In much the same way, a chess-playing program cannot play chess without knowing where the pieces are on the board. The heuristic rules responsible for choosing the proper move need to know precisely where the
pieces are on the board. Likewise, the heuristic rules responsible for choosing the proper thought need access to definite semantic content of current and previous thoughts. Preservation of content coherence between thoughts is the result of the heuristic rules determining the proper (content coherence preserving) thought to be produced next, from a set of possible thoughts. Likewise, for a program, the playing of a successful game of chess is the result of heuristic rules determining the most advantageous/proper move from a set of possible moves. In both the chess playing program and the formal system with feedback, the combination of algorithmic and heuristic rules are absolutely necessary. Without the algorithmic rules, the system simply could not function. The heuristic rules are also necessary to provide a logical and teleological progression from one move/thought to the next.

The consciousness-receptor/heuristic-effector feedback would allow for some degree of conscious control over the process of thinking in the production of a chain of thoughts. The thought experiment earlier showed that at least some conscious control was present in the creation of thoughts, and this is consistent with common sense, naïve conceptions about how we think. That is to say, we like to think that we do have some conscious control over our own thought processes. The consciousness feedback system specifically ensures this conception, while pure syntax manipulating formal systems do not.

As I have said, this has mainly been conjecture about how the mind may work, but as it turns out feedback in cognition has been discussed recently in psychological journals. More specifically, consciousness feedback has been suggested as an explanation for schizophrenic thought
The claim is that a malfunction in the feedback mechanism of the mind plays havoc with the sense of agency that normally accompanies our thoughts. A schizophrenic suffering from thought insertion would believe that someone or something was planting thoughts in their mind, or in other words, they are not in control of the production of some of the thoughts they are experiencing. In his article, Campbell proposes that this lack of agency in some thoughts is due to a malfunction in a feedback system involved in cognition. “…a breakdown in the mechanism of efferent copy and comparator would result in the breakdown in the sense of agency that is characteristic of schizophrenia”

The ability to check for the authenticity of thoughts is probably not one of the functions that the human brain evolved to have. That is to say, it is unlikely that during a period in our evolution humans faced a danger that foreign thoughts were being inserted into our minds and that we evolved a specialized portion of the brain, a “Thought-agency-authenticator.”, to detect whether or not we are the agent responsible for thinking the thoughts we experience. Given that it is highly unlikely that we evolved such a specialized mechanism, it is probably not the case that schizophrenics experience thought insertion because of a malfunction or absence of a “Thought-agency-authenticator.”. Far more likely is that schizophrenics suffer thought insertion because of a damaging or malfunction of some other area of the brain, the effect of which is the experience of thought insertion. As stated in the quote above, Campbell believes that a breakdown in a cognitive feedback system would produce the effect of thought insertion, rather than a malfunction of a specific area or module responsible for authenticating thoughts as ones own. As he puts it, “a breakdown in the mechanism of efferent copy and comparator”, implies a malfunction in a
feedback mechanism that operates in a similar way to the feedback involved in motor function. The feedback mechanism in motor function utilizes the efferent copy and comparator to synchronize the current physical action, (i.e. the location and speed of the arm or hand in relation to the environment) with the mental impetus behind the action. For example, when you visualize throwing the basketball into the hoop and then you attempt to throw the basketball into the hoop, the comparator compares the actual physical location of your arms with the efferent copy of what you want to happen. The feedback mechanism of efferent copy and comparator in motor function has been verified by empirical evidence.

“For someone wearing laterally displacing prisms, everything looks to be somewhat to the left of where it really is. However, after wearing them for a while, people unreflectively reach to the correct place to grasp something they see. Held (1961) suggested that copies of the motor instruction are sent to a comparator, stored there, and compared to the proprioceptive or visual--"reafferent"--information about what movement was actually made. Whether there was a match between the motor instruction sent and the effect achieved determined which motor instructions were sent on later occasions.”. xxxiii The feedback mechanism of efferent copy and comparator explains the self-regulation of the motor functions despite the distortion imposed by the prisms. In addition to playing a role in the operations of motor function, feedback is also implicated in the development of motor function. “Held proposed that this model also explained why developing normal motor skills also depends on the subject making self-generated movements rather than moving passively.”xxxiv In other words, the baby needs to crawl on its own if it is going to learn to crawl. Grabbing a hold of its arms and legs and making a crawling movement for it will do no good in terms of development. This is
because movement that is not self-generated will not have a goal or impetus to be judged against.

The role of feedback in motor function is uncontroversial and has much to recommend it. The question is, does feedback play as important a role in cognition as it does in motor function? In other words, “Is the process of thinking similar to motor functions?”

Campbell thinks that the process thinking is very similar to motor function. Although this paper uses feedback to answer questions about syntactic ambiguity, and Campbell’s article uses feedback to answer questions about thought insertion (“delusion can again be explained by viewing thinking as a motor process, and supposing that for the schizophrenic, something has gone wrong with the monitoring of efferent copies of instructions to think thoughts…”) the feedback system described in both is essentially the same. Although Campbell does not mention heuristic rules as being involved in the feedback mechanism, he does suggest that the process of thinking involves feedback, “…one input to the formation of an occurrent thought may be the earlier occurrent thoughts that you had, as when you are following through some train of thought. On this picture, the earlier occurrent thoughts will be yet more inputs.” The similarity in the operations of the feedback systems proposed by this paper and Campbell’s article, despite the difference in intended function and specific operation, is promising. Thought insertion would lead to precisely the kind of incoherence that is the topic of this thesis, so it is plausible that the mechanisms whose breakdown is responsible for that kind of incoherence are also responsible, when functioning normally, for coherence. Therefore, the empirical literature cited by Campbell saves the hypothesis of
conscious feedback from being pure speculation, and adds some empirical weight as well.

There is further evidence for conscious feedback playing a role in cognition. In addition to being used to explain thought insertion, cognitive feedback is also being explored as a possible explanation for the effects of hypnosis\textsuperscript{xxxvi}. The effectiveness of hypnosis has long been known, with hypnosis being used in entertainment, psychology and in India it was even used to manage pain during surgery up until anesthetic was introduced. While it was known that hypnosis was effective, there has been no commonly accepted explanation about how it actually worked. New information about the structure of the brain involving the flow of information shows feedback is highly important in cognitive functioning, and is a possible explanation for the effectiveness of hypnosis. The commonly accepted view of the flow of information in the brain is that the relationship between the senses and the sensory information processing centers was essentially a one-way street. Information went from the senses to the higher centers of cognition where it was processed, and very little, or nothing went back to the senses. Recent studies of the brain suggest that this conception of the flow of information is actually backwards. It appears that the amount of information from the higher processing centers that goes to the senses is approximately ten times the amount that comes from the senses.\textsuperscript{xxxvii} While this has little to do directly with thought to thought causation, it does supply evidence that feedback is involved in many aspects of cognition and brain function.
Chapter Eight. Objections and Replies.

An immediate objection that presents itself upon contemplation of the consciousness feedback model is that it appears to appeal to a kind of supervenience between the semantic content and the syntactic properties of thoughts, while at the same time breaking the rules of supervenience:xxxviii namely, there appears to be some changing of the semantic (mental) properties without a resulting change in the syntactic (physical) properties. This is the result of having abandoned, or argued against the connection between syntax and semantics that first made the language of thought hypothesis so inviting. At first blush, the language of thought hypothesis offers a promising solution for integrating the content of thoughts into the psychological mechanics of the mind, using the connection between syntax and semantics. The problem with that approach is that natural language is syntactically ambiguous. If thoughts occur in a natural language, then the connection between syntax and semantics is not tight enough to allow a formal system operating only on the syntactic properties of thoughts to ensure the coherence of semantic content in chains of thoughts. The syntactic properties of a thought may be ambiguous, but the actual instancing of a thought in the brain is unambiguous. For example,

Here is a sequence of three thoughts, with their syntactic and semantic properties tightly connected.

Sem1 – Sem2 – Sem3
T1 - T2 - T3
Syn1 - Syn2 - Syn3*
Sem1-2-3 = Semantic content exhibiting coherence with one another.
T1-2-3 = Thought 1, 2, and 3
Syn1-2-3 = Syntactic property 1, 2, and 3

As long as the connection between syntax and semantics is very tight, then the syntax can account for the semantics. Unfortunately, the connection between syntax and semantics can be ambiguous, and so syntax cannot guarantee content coherence. The same example, this time taking into account syntactic ambiguity would look like this.

Sem1 – Sem2, Sem9 – SemZ*
T1    -    T2    -    T3
Syn1 -  Syn2^        -  Syn3
Sem2, Sem9 = The possible semantic interpretations of Syn2^
SemZ = The semantic content of T3, the content is not coherent with the previous or following thoughts in the chain.
Syn2^ = A syntactic property of T2, with multiple semantic interpretations (ambiguous).

Under the scenario above, it appears that Sem2 and Sem9 cannot both supervene on Syn2^ without breaking the rules of supervenience. To avoid denying the supervenience between the semantic and the syntactic properties, this example includes the instancing of the syntactic properties.
Sem1 – Sem2 – Sem3
T1    -    T2    -    T3
Syn.i.1- Syn.i.2 – Syn.i.3*
Syn.i.1,2,3 = The syntactic properties of the instance of a thought in the brain.

While the syntactic property, Syn2^ may be ambiguous, in each instance that it occurs (Syn.i.2) it has a distinct and singular semantic interpretation. The ambiguous syntactic property Syn2^ can be interpreted as either Sem2 or Sem9. The supervenience of the semantic content on syntactic property means that the semantic content is in some sense dependent on the syntactic property of the thought. While the syntactic property of the thought may be ambiguous with several semantic interpretations, this does not break the rules of supervenience as the ambiguity only exists so long as the thought is not actually instanced. For example, while an ambiguous sentence may be able to be interpreted several different ways, to read it is to interpret it in one particular way. You can read the same sentence several times, and interpret it in several different ways, but not all at the same time. It is the same with syntactically ambiguous thoughts, although there may be several semantic interpretations of a particular thought, each instance of the thought occurring in the brain correlates with only one semantic interpretation. As is evidenced here, “Since any event can be an instance of two or more distinct properties, it may well be the case that one and the same event is both an instance of the property, being a mental property, and an instance of another, physical property, say, being a brain event where being an instance of the former just is being an instance of the latter. That is to say, the different properties are co-instantiated; they, in this particular event, “share” the same instance. Their difference at the property-level is
ensured by the two properties not always being co-instantiated."xxxix

Therefore, there are no mental differences without physical differences. The real worry for syntactic ambiguity is not that you have two or more semantic (mental) events occurring for only one syntactic (physical) event: instead, the worry is that ambiguity will lead to incoherence in chains of thought.

Another benefit of the consciousness feedback system is that it gives us an answer to questions about why our inner monologue thoughts occur in natural language. One of the arguments raised by opponents of the machine language version of the language of thought hypothesis centers around the fact that the entirety of our inner monologue thoughts occur in our natural language. The essence of the argument is that a system where our thoughts occur in a natural language that is then compiled into a biological machine language by the brain is overly complicated. The reply to this challenge by the machine language fans is presented in the following paragraph, and is followed by the reply from consciousness feedback theory.

The machine language of thought reply to this argument is that infants who are pre-linguistic are still able to think, despite arguably not having access to natural language. If the language of thought resembled a natural language rather than a machine language, then infants without access to natural language would be unable to think. This reply is essentially the objection from animal cognition wrapped up in slightly different clothes. The pre-linguistic infant reply suffers from the same flaw as the animal cognition argument, and that is that it was never claimed that animals or
infants possess the same cognitive abilities as adult humans. The difference between the animal and infant cognition arguments is that human infants will eventually become human adults, and the intuition is that infant cognition is not radically different from adult cognition. Since our infant mind eventually becomes our adult mind, we would expect infant cognition to perhaps be different in complexity or power to adult cognition, rather than different in nature or operation. In other words, a human infant will eventually come to think like an adult human, while a dog or ape will presumably never entertain the cognitive ability of a human adult. However, the argument from animal or infant cognition misses the point. While it may be the case that infants and animals do not think in natural language, that does not mean that adult humans don’t think in natural language. In addition, although infants cannot form words or sentences with their mouths and vocal chords, that in no way proves that they do not think in a form of natural language.

The fact that the vast majority of our abstract thought occurs in an inner monologue of natural language is not a problem for the consciousness feedback theory. The semantic content of an inner monologue thought is consciously experienced, and it is that experience which acts as the feedback receptors in the cognitive system. Natural language does not suffer from semantic poverty, and is able to account for the representational nature of thoughts. Therefore, it would be odd if our experience of the semantic content of our thoughts did not occur in natural language.

An objection from the machine language of thought to natural language of thought, and by association consciousness feedback, is that the language of thought cannot be a natural language, since natural language suffers from syntactic ambiguity. As we have seen, this would be a troubling
problem, if we were proposing a system that operated without consciousness feedback. Syntactic ambiguity is only a problem if the formal system is only sensitive to syntactic operations. A system that operates with consciousness feedback deals with syntactic ambiguity in a manner that closely resembles our experience. The semantic content of surrounding dialogue provides hints and clues about the proper meaning of ambiguous words or phrases, and a system with consciousness feedback and memory has the advantage of having access to context. The conscious experience of a thought, as well as access to recently experienced thoughts in memory is sufficient to establish context. Context is not only useful for providing the meaning for unknown words and phrases; it is also the only way we currently have to determine relevance. At this time, there are no theories that explore or attempt to explain relevance without mentioning context. Relevance is vital to content coherence between thoughts, as two thoughts that are not relevant to one another at all would not display any content coherence. Explaining content coherence without appealing to either context or relevance is very difficult, and is currently a problem faced by machine language of thought theories. Access to context and relevance gives consciousness feedback and natural language of thought theories a more convincing answer to the problem of content coherence.

It appears that in general, materialists intentionally ignore the topic of consciousness. Opting instead to define “a theory of mind” as actually meaning “A theory of cognition.”. The reason for the imposed limitation on the scope of a materialist theory of mind is explained by Cummins in *Meaning and Mental Representation*. “It seems plausible to suppose that a
creature could have a mind without having emotions, as is supposed to be
t the case with Star Trek’s Mr. Spock. Descartes held that the essence of mind
is thought, Locke that it is the capacity for thought. A system that could do
nothing but think might be a rather colorless mind by human standards, but
there seems to be something to the traditional idea that such a system would
nevertheless be a mind. On the other hand, a system that could not think but
could feel, have emotions, and so on does not seem to qualify as a mind. If
this is right…then it is just the idea that thinking (and / or the capacity for
thought) is the essence of mind and can be studied independently of other
mental phenomena.”xl. Cummins takes this to mean that a theory of mind
does not necessarily need to address emotions or feelings, as they can be
separated from cognition. Cummins expands on the argument above, to
show that accepting this conceptual distinction between cognition and
feelings and emotion allows for a theory of mind to ignore consciousness. “It
is possible for the cognitive scientist to ignore (provisionally, at least) such
mental phenomena as moods, emotions, sensations and –most important-
consciousness.”xli The reason most often given for ignoring consciousness in
materialist theories of mind is simplicity. A number of tricky questions and
issues can be avoided by leaving consciousness out of the picture. As
Cummins’ says in the quotes above, it is at least conceptually possible to
strip the mind down to purely cognitive elements. It is certainly possible to
imagine a mind that was incapable of feeling emotion, and only produced
thoughts with no emotional characteristics. If ones goal is to explain
cognition, which is one of the central goals of cognitive science, then it
would make sense to utilize a model of the mind which focused on cognitive
properties rather than emotional properties. However, it is also possible to
imagine a mind that was purely cognitive and conscious as well.
The removal of emotions and feeling from a model of the mind is not necessarily the same thing as the removal of consciousness. There is a sense in which a mind that was not conscious could not feel emotions, but this does not mean that by including consciousness into a theory of mind, you must also include emotion. When one experiences an emotion, say one feels sad, the sadness is the conscious experience of being sad. Because an emotion relies so heavily on conscious experience, it would be strange to think that a mind could not be conscious and still experience emotions. On the other hand, it is not so strange to think of a mind that is conscious and feels no emotions.

Aside from theoretical simplicity, the other reason materialists typically exclude consciousness from their theories of mind is that it has been the source of several objections to materialism. There are strong intuitions that conscious experience is essentially subjective, and is therefore unexplainable by objective science. “We may call this the subjective character of experience. It is not captured by any of the familiar, recently devised reductive analyses of the mental, for all of them are logically compatible with its absence. It is not analyzable in terms of any explanatory system of functional states, or intentional states, since these could be ascribed to robots or automata that behaved like people though they experienced nothing.” These intuitions form objections that normally focus on phenomenal characteristics, such as qualia. These objections are made more troublesome by the fact that there is really no generally agreed upon idea of what consciousness actually is. The objections and lack of consensus makes the inclusion of consciousness in any kind of explanatory capacity in a materialist theory of mind rare.
The argument against using consciousness in any kind of explanatory capacity is a strong one. The lack of knowledge about what consciousness actually is makes it a potentially dangerous addition to a theory. In some ways, it resembles the Parallelists’ practice of including God as an explanatory force in a theory. Given the central role that consciousness plays in the feedback model of cognition, it may end up being too much of a *deus ex machina*. It may be problematic to rely on something whose nature and origin threaten to remain unknown as the intended solution to the problems faced by the proposed theory. While there may be questions about the exact interaction between consciousness and the heuristic rules, and whether or not that interaction is problematic for materialism, those questions must remain unanswered for now. This is not to say that the questions can never be answered, the fact of the matter is that more information about the nature of consciousness is needed before the questions are able to be answered.

On the other hand, several theorists have viewed the exclusion of consciousness from cognitive science theories of the mind as a problem for some time now. The answer to the question of whether or not the inclusion of consciousness is appropriate or not is decided by what function you want a theory of mind to serve. As a theory of mind that intends to function as a blueprint for constructing artificial intelligence, the inclusion of consciousness may be too much, too soon. On the other hand, if one is looking for a philosophical theory of the more abstract workings of the mind, then the exclusion of consciousness is arguably more troubling than its inclusion. It is the latter of the two theories that this paper provides, a philosophical theory of mind that hopes to solve more problems than it raises.
One danger that the consciousness feedback system allows us to sidestep is the possibility of the syntactic zombie\textsuperscript{xliii}. A syntactic zombie is a machine built to look like a person just like you or I, only their thoughts are not causally influenced by their semantic content. The syntactic zombie operates on an automated formal system that operates on syntax alone. The disturbing thing about the zombie is that it functions normally from outside observation, but the psychological mechanisms of the zombie’s mind operate with no concern about its conscious mental life. There would be no way to test for the existence of syntactic zombies, since they could pass the Turing test. The syntactic zombie is based on essentially a solipsistic worry (that there is no way to know if anyone else is conscious) with a slight variation. The fear of the syntactic zombie is not so much that other people might not be conscious, but rather that consciousness plays no part in their mental life. In other words, they could be fully conscious, but they would have no conscious control over the operations of their own mind. Perhaps the most worrying thing about the syntactic zombie is that if the syntactic, formal system responsible for thinking were good enough, it would be very difficult to tell if you were a syntactic zombie or a human. The psychological mechanisms of the mind of the syntactic zombie are purely formal, with no sensitivity to semantic content. Consciousness feedback negates the possibility of the syntactic zombie, since consciousness is necessary to fulfill the role of receptor in the feedback system.

When compared to a purely syntactic (machine language) or semantic (natural language) approach to the process of thinking, the theory of consciousness feedback has more to recommend it. Consciousness feedback
accomplishes the same goals, but without suffering from the limitations that accompany a purely natural or machine based version of the language of thought. In addition to avoiding arguments and objections from semantic poverty and syntactic ambiguity, the consciousness feedback theory is far more compatible with our own experience than purely machine or natural language theories.
Bibliography.


Campbell, John. “Schizophrenia, the space of reasons, and thinking as a motor process.” Monist; Oct99, Vol. 82 Issue 4, p609, 17p

Carruthers, Peter “The cognitive functions of language.”, Behavioral Brain Sciences, 2002, #25


ii Socrates is speaking with Glaucon.

[Socrates:] And now, I said, let me show in a figure how far our nature is enlightened or unenlightened: --Behold! human beings living in a underground den, which has a mouth open towards the light and reaching all along the den; here they have been from their childhood, and have their legs and necks chained so that they cannot move, and can only see before them, being prevented by the chains from turning round their heads. Above and behind them a fire is blazing at a distance, and between the fire and the prisoners there is a raised way; and you will see, if you look, a low wall built along the way, like the screen which marionette players have in front of them, over which they show the puppets.

[Glaucon:] I see.

And do you see, I said, men passing along the wall carrying all sorts of vessels, and statues and figures of animals made of wood and stone and various materials, which appear over the wall? Some of them are talking, others silent.

You have shown me a strange image, and they are strange prisoners.

Like ourselves, I replied; and they see only their own shadows, or the shadows of one another, which the fire throws on the opposite wall of the cave?

True, he said; how could they see anything but the shadows if they were never allowed to move their heads?

And of the objects which are being carried in like manner they would only see the shadows?

Yes, he said.

And if they were able to converse with one another, would they not suppose that they were naming what was actually before them?

Very true.

And suppose further that the prison had an echo which came from the other side, would they not be sure to fancy when one of the passers-by spoke that the voice which they heard came from the passing shadow?
No question, he replied.
To them, I said, the truth would be literally nothing but the shadows of the images.

That is certain.
And now look again, and see what will naturally follow if the prisoners are released and disabused of their error. At first, when any of them is liberated and compelled suddenly to stand up and turn his neck round and walk and look towards the light, he will suffer sharp pains; the glare will distress him, and he will be unable to see the realities of which in his former state he had seen the shadows; and then conceive some one saying to him, that what he saw before was an illusion, but that now, when he is approaching nearer to being and his eye is turned towards more real existence, he has a clearer vision, -what will be his reply? And you may further imagine that his instructor is pointing to the objects as they pass and requiring him to name them, -- will he not be perplexed? Will he not fancy that the shadows which he formerly saw are truer than the objects which are now shown to him?

iii Descartes, Rene. Meditations. Meditation II
iv Descartes, Rene. Meditations. Meditation II
v Descartes, Rene. Meditations. Meditation II
vi “But [as to myself, what can I now say that I am], since I suppose there exists an extremely powerful, and, if I may so speak, malignant being, whose whole endeavors are directed toward deceiving me? Can I affirm that I possess any one of all those attributes of which I have lately spoken as belonging to the nature of body?” – Descartes, Meditation II
viii Spurrett and Papineau “The completeness of ‘Physics’” Analysis, 1999
ix Andrew Pyle “Malebranche”, , Routledge, 2003
x Andrew Pyle “Malebranche”, , Routledge, 2003
xi “One is obliged to admit that perception and what depends upon it is inexplicable on mechanical principles, that is, by figures and motions.” Leibniz, G.W “Monadology” (1714)

xiii Fodor, Psychosemantics. Pg 134.
xiv Fodor, Psychosemantics, pg23
 xv Haugeland, mind design, pg5
 xvi Haugeland, mind design, pg6
 xvii Haugeland, mind design, pg8
 xviii Haugeland, mind design, pg10
 xix Haugeland, mind design, pg11
 xx SEP, LOT
xxi Fodor, Methodological Solipsism, pg.63
xxii Stanford Encyclopedia of Philosophy, Language of Thought Article.
The author uses ‘Command Language’ as a term referring to a language constructed entirely of internal commands for a machine, rather than referring to a higher level programming language used to run commands that would otherwise be run at a prompt.

David Cole, “Hearing yourself Think.”

Carruthers, Peter “The cognitive functions of language.”, BBS 2002,#25


Fodor, Psychosemantics. pg. 140

Fodor, Psychosemantics page 140

Fodor, Psychosemantics. pg. 140

Fodor, Psychosemantics. pg. 140 (Italics added by author.)

Campbell, John. “Schizophrenia, the space of reasons, and thinking as a motor process.” Monist; Oct99, Vol. 82 Issue 4, p609

Campbell, John. “Schizophrenia, the space of reasons, and thinking as a motor process.”

Campbell, John. “Schizophrenia, the space of reasons, and thinking as a motor process.

Campbell, John. “Schizophrenia, the space of reasons, and thinking as a motor process.

Campbell, John. “Schizophrenia, the space of reasons, and thinking as a motor process.


“The fact that supervening properties need not be identical to their subvening properties is the source of the great appeal of supervenience to contemporary philosophers of mind who have come to think that the mental cannot be identical to the physical (largely due to considerations of multiple realizability) yet want to be physicalists and thus hold on to the notion that the mental is nonetheless determined by the physical. Thus they subscribe to the thesis of psychophysical supervenience, AKA, the supervenience thesis.

The supervenience thesis states that mental properties and facts supervene on physical properties and facts. The supervenience thesis can be further unpacked as the following three theses about objects and their properties. (Please note that I use the term "object" as a place holder for the unwieldy "object, event, or state of affairs" and the term "properties" to denote extrinsic as well as intrinsic properties.)

(i) No two objects can differ in their mental properties without differing in their physical properties.
(ii) A single object cannot change its mental properties without changing its physical properties.

(iii) If, at a given time t, a single object has two different subsets of mental properties, it must have two different subsets of physical properties.

The above three corollaries to the supervenience thesis are each captured by the slogan "no mental differences without physical differences." Dictionary of Philosophy of Mind.


xl Cummins, Meaning and mental representation. Pg. 19

xli Cummins, Meaning and mental representation. Pg. 19

xlii Nagel, Thomas. “What is it like to be a bat?”. Pg 436

xliii The syntactic zombie is not like other philosophical zombies, in that the syntactic zombie can also be conscious. The access to consciousness makes the syntactic zombie immune to arguments against philosophical zombies in general. Daniel Dennett argues that philosophical or p. zombies are preposterous and cannot function as they are imagined to function because, “they understand what they say (or, not to beg any questions, they understand what they say), they believe what we believe, right down to having beliefsz that perfectly mirror all our beliefs about inverted spectra, "qualia," and every other possible topic of human reflection and conversation.” Daniel Dennett, Brainchildren, essays on designing minds.

The syntactic zombie, unlike the p. zombie, is able to actually believe what we believe, understand what they say, and have qualia and inverted spectra just as we are able to. The difference between a syntactic zombie and a human being is that the syntactic zombie has no real agency from thought to thought, thinking is a causal chain of events which has nothing to do with the will of the syntactic zombie.