On Simulating Dialects of Thought

Tim O'Connor

University of Victoria

This paper will explore Jerry Fodor's claim that there has to be a Language of Thought. It will be argued that, although this mentalese may have universal properties, each of us thinks in our own dialect of this common mental language. The syntactic structure of this language of thought is both inter-subjective and creatively diverse. Languages used for the public expression of thought are thus similar to formalized, institutionalized, or 'proper' dialects (such as Queen's English) that facilitate a stable common ground for communication. The relative nature of this syntax or grammar of thought will then be fed into the simulation theory heuristic. The simulation theory will be briefly discussed and it will be argued that the imaginative projection or mind-reading that the theory postulates is instead aimed at understanding the idiosyncratic mental grammar of others. This paper will conclude that we use simulation to understand novel syntax in mentalese.

The Language of Thought

Fodor's argument for what he calls a "Language of Thought" (LOT) presents an explanation of the form and nature of thinking itself. This theory, being both computational and representational, confounds connectionist and behaviourist theories of mind. The main thrust of Fodor's argument is to evaluate thought in the context of language and to demonstrate that, much like language, thought is made up of constituent parts, which exhibit *syntactic structure*, as well as *semantic content*. This syntactic structure allows the parts to combine in a way as to preserve the semantic content of the thought (being analogous to a mental sentence) itself. Thought then, is composed of lexicons (such as words in a language), which can be combined using general rules (such as grammar in a language) to express semantic content (or meaning).

Fodor frames his discussion around the concept of intention boxes[9], which exist in the mind of the beholder. Token *symbols* with semantic content are placed inside the box, computed, and an appropriate behaviour is consequently produced. This is the process of translating simple thought (containing one token symbol) into action or behaviour. However, when thought becomes more complex, it must display a syntactic structure in order to preserve the meaning of each token symbol to be imputed. Hence, to raise my left hand and hop on my right foot, the operator "and" is used to preserve and create semantic content. We can also see how the thought is complex and can be broken down into constituent parts, whose parts can in fact, be broken down even further. This operation is similar to the way a complex sentence, which expresses a more abstract meaning, can be reduced to atomic meanings of the individual words. When broken down to the bare ingredients, we see the combinatorial nature of each element (or pan) in that it can be used in any number of thoughts to represent the same thing without losing its semantic content. These complex syntactic

rules, which are used to manipulate the representations of the mind, can easily be described metaphorically as our *mental grammar.*

The strength of the LOT theory is that it accounts for the productivity[10]of human thought, which is for the most part unexplained by other models. That is, it shows how we can develop an infinite set of ideas with a finite set of input, or in other words, how we can create an unlimited amount of possible collages with a limited amount of pictures. The theory also explains the systematicity of thought[11]in the way that it's syntactical structure provides a semantic content that is preserved in any number of appropriate combinations (i.e. pans). Hence any native speaker of English can understand the sentence "John loves Mary" as well as the sentence "Mary loves John" simultaneously by understanding the structural rules, which govern both expressions.

As language developed from a need to communicate or otherwise express thought, it would make sense that the architecture of these languages would somehow mirror the organizational patterns of human thinking. Even the root languages of human thought such as mathematics and logic display similar characteristics and structure[12]. It is important however, to note that mentalese, the origin of all prior languages, should operate in an ultimately more advanced and complex manner than that of languages used to express these ideas. It would be all too obvious to state the inadequacies inherent to any mode of inter-subjective communication or expression of thought. Conceptual content is conveyed to others through these often clumsy and rough approximations of LOT. Here, we should resist mistaking the map of language for the terrain of mentalese[13].

Novel Syntax in Mentalese

Lexicons of thought manifest themselves through experience, whereas mental grammar, according to Fodor, is innate. Every mind, being numerically and spatio-temporily distinct, represents or senses any object differently as every mind experiences the world from a unique position. If we take into consideration the problem of the inverted spectrum, we can neither say with certainty that any two minds represent any one colour in the same way. Hence, it is reasonable to assume that to some degree, lexicons of thought (or *what* we think about) contain subjective semantic content.

Although mental grammar is innate and inherent in all minds of the species[14], it is neither objective nor identical in all minds *necessarily*. If it were, it would follow that minds that have been exposed to identical experiences (albeit impossible) would generate identical ideas. The aforementioned gap in the semantic content of the lexicons between subjects is too weak to account for the diversity of thoughts in minds of similar experience. How, too, can a universal mental grammar explain the wealth of meaningful concepts produced in a mind of severely limited experience; especially in comparison to the deficit of meaningful concepts found in a mind of wide and varied experience? It would appear plausible that the syntactical structure or grammar of mentalese differs from mind to mind either in configuration and/or intensity. As we further explore this argument, it is important to observe in what ways our mental grammar display certain common traits.

There must be comparable or similar operations of our *syntactic structure* of thought. If we had entirely contrasting mental grammar, language itself would not be possible. Language, we are told, developed by referencing similar objects or circumstances with commonly identifiable physical signals by pointing and grunting. As language grows more complex, more abstract concepts evolve. Similar ideas are cross-referenced against similar experiences and are expressed using the mechanisms of the language that have been made available. This is the way we present our ideas to each other. For example, this essay is an attempt to communicate an idea (or series thereof). Its form and the way it is written are meant to demonstrate meaning to a particular audience or set of minds. I phrase my language for practical purposes (such as efficiency) and guide my train of thought down a pre-modeled path of academic prose. The way in which I think to myself about this topic could perhaps be more accurately worded using a more idiosyncratic language, yet the process of expression itself somehow confounds the *truly*felt semantic content of my thoughts. Perhaps this is a result of the *private* semantic content of my mentalese being expressed through an inter-subjective or *public*syntactic form. If this is true, then the exact translation of any mental concept (ranging from concrete to abstract) into the public sphere may not possible to accurately decipher[15]. We have no other tools than those of signs and signals, language and behaviour, for which to represent subjective semantic content to each other.

We have chosen commonly conferred upon *icons* to which assign a range of subjectively experienced meanings. An example would be that of wine tasting. I have never experienced a wine that has had any flavour identifiable with nuts. However, I learned what a ‘nutty’ wine was by identifying a particular taste, which was present every time I was told the wine is ‘nutty’. Now, I merely associate the *qualia* previously experienced, with the word ‘nutty’, which allows me to avoid such wines. So then, for practical purposes we get by with these rough but extremely useful metaphors[16]that approximate subjective meaning accurately enough for our purposes.

This still leaves us with the question of how this *structure*of languages we use for the transmission of our thoughts (from Swahili to mathematics) came about. Clearly there are some structures of thought we share. It would be safe to say that logic is inherent in all systems, by definition. We use systems that work on the same or similar modes of logic because it is how we rationally make sense out of the world. This common trait of thought would seem likely to be innate, as well as a requisite for human thought. The *intensity* of this *universal trait*, which must mirror the structure of logic somehow, might also vary form mind to mind. For example, minds that are considered to be insane, in varying degrees, demonstrate statistical deviancies in their abilities to express logical thought patterns. As a result, they cannot be easily understood. To the insane mind, ideas may follow from others with some sort of pattern of organization and coherence of meaning. However, it could be conceived that these minds are labeled insane because of the *statistical deviancy*of their unique pattern of thought in comparison to minds with a more *common pattern*. It could be said that the insane mind lacks in *intensity*, this universal and structural trait.

So far, arguments for the subjective nature of syntactic rules in mentalese have been discussed along with arguments for their commonalities. To illustrate the argument as a whole, an analogy with genetic evolution may be appropriate. Let us suppose that within a gene there are *necessary*codes, that is, codes that determine the physical structure necessary for the organism to be human. These codes define our species. Within the same gene, there are *contingent* codes, that is, codes that provide the diversity within the species required for creative adaptation. These codes define the individual. The boundaries between what defines the species and what defines the individual blur

into each other; the species defines the individual as the individual defines the species. Although this analogy may not be entirely accurate with regards to actual genetic theory, it is a helpful illustration in order to view the purposed relationship between our common and subjective syntactic structure of thought.

The aforementioned analogy is not intended to imply the inhumanity of the mentally insane. To make this assumption would be again to mistake the map for the terrain. The point is merely, that hyper-subjective and publicly obscure thought patterns could be seen in this context as thought structure *mutations*, like physical mutations, which may or may not be beneficial in the long run. This might make evolutionary sense since, if we all acted according to the same principles of thought and our behaviour is determined by our thoughts, then we might be too predictable for our own good.

It is not the purpose of this essay, however, to discuss the extent to which this grammar of thought is innate. Only, to present a firm argument, which suggests that the syntactic structure or grammar of mentalese is inter-subjective and creatively diverse. In short, we all think in our own dialect of a common mental language.

Simulation Theory

The Simulation Theory (ST) attempts to explain how we attribute beliefs and desires to other minds. Gordon claims that we perform a kind of mind-reading or imaginative projection when developing a hypothesis about or perceiving the mental states of others.[17] In our own minds, we occupy certain beliefs and desires, which are then in turn, processed by decision-making mechanisms. The result of this process usually determines our behaviour. When predicting the behaviour of another mind, however, this process runs off-line. Simulations of foreign beliefs and desires involve a hypothetical, counterfactual, off-line accounting where one inserts what one believes are the mental states of the other person into one's own system as if they were one's own. These hypothetical mental states are then processed by the thinker's decision-making mechanisms, again off-line. The result is a simulation of how the other mind will behave given these hypothetical conditions. The outputs become images of the thinker's mind.[18] These images are of a structure presented to the mind as some kind of shape, or it could be argued, as a feeling of some particular mixture. Hence, the imagination, according to simulation theorists, allows us to switch spatio-temporal situations and is requisite for the human ability to recognize and interact with each other.

A good way to illustrate this argument is presented by Goldman in his tennis player thought-experiment[19]. Upon beginning a match, the tennis player asks himself where his opponent will aim his next shot. In order to answer this question, the tennis player must imagine where he himself would aim the next shot if he were in his opponent's position on the court with his opponent's

mental state (including beliefs about tennis skills and strategies). The tennis player then simulates having these mental states, which allows his own reasoning mechanisms to operate on this input. The off-line process allows the tennis player to adopt a feign choice in order to predict the choice of his opponent. The tennis player's accuracy in the prediction of his opponent's behaviour will clearly improve with more accurate simulation of the initial conditions in the way that we can better predict the actions of familiar people than those of strangers.

ST, as presented, seems to have more plausibility when seen running backwards from behaviour to mental states.[20] The only information available to the mind is the behaviour of others and the contextual background in which that behaviour is exhibited. From this input, a simulation of the beliefs and desires of the other individual is created using the same hypothetical reasoning. This reversal of the simulation arrow seems to conform more to experience since the behaviour of the other individual is more readily available and the goal of simulation is often to attain a prediction of another's beliefs and desires. Once this can be achieved, the prediction of future behaviour can be deduced from the simulation of what it is like to occupy the mental states of the opponent. In other words, from what one imagines *it is like* to be the other person, one can derive how that person is likely to behave.

There is a minor but noteworthy variation in the opinions of Gordon and Goldman relevant to this point. Gordon insists that the simulation is based on how an individual would behave if they were in the other's *situation*. Goldman, on the other hand, insists that the simulation is based on how the individual would behave *as if* the other person. The difficulty comes in describing what one is actually imagining. The problem is analogous to what Nagel calls an "explanatory gap" in his article entitled "What it is Like to be a Bat".[21] Imagining another person's mental processes is much akin to understanding their brain processes in the way that neither of these conceptual frameworks will accurately (if they are able at all) explain another person's qualitative conscious experience.[22] The problem is also analogous to the aforementioned indeterminacy of translation[23]between mentalese and expressive language.

The reason for this explanatory gap between actual thought processes and imaginative representations thereof, I will argue, is the same as the reason for the explanatory gap between mentalese and language. This problem of certainty is brought on because of the subjective and relative nature of our own contingent syntactic structures of thought. ST claims that the decision-making mechanism that mediates beliefs and behaviour is cross-cultural. Hence, by simulating the beliefs and desires of others, one can predict the behaviour of others — the accuracy of the prediction directly resulting form the accuracy of the simulation. ST also claims this is why we are more likely to understand the behaviour of those from a similar culture versus those from a foreign culture. It is, according to ST, merely a matter of juxtaposing beliefs.

Synthesis of Reverse Simulation and Novel Syntax Arguments

Let us now consider the *relevant adjustments* that ST proposes are required for accurate simulation. If we accept the previous argument that there exists novel syntax in mentalese, the relevant adjustments required for accurate simulation of another mind's mental states include not simply the positing of different beliefs, but the positing of different *objects* of belief and computing them through a *posited* syntactical architecture that differs in degree from our own. In other words, it is not only necessary to imagine *what* another person believes, but it is also necessary to

imagine *how* they think about these beliefs and how these beliefs and desires *fit* into the other thinker's conceptual framework. The accuracy of such a simulation would also improve with familiarity but it would not be hindered by the other mind's exposure to different experiences. Thus it can be explained why we can sometimes better predict the behaviour of a person from a foreign culture better than the behaviour of a person from our own neigbourhood. Thus it can also explain the wide discrepancy in predicted and actual behaviour. The person one imagines then, is not one's self thinking in the same way with different beliefs and desires, nor is it the case that one is imagining themselves *as if* they were the other person. Instead, it is one's self thinking differently about different things.

According to Fodor, the cause of behaviour is found in the intentional content of mental states and not in their intentional objects. In other words, behaviour is causally connected to the representational computation of constituent formulas of thought and not in the raw intentional objects themselves. Thought as mediated by the language of folk psychology results in physiological output. Again, it is not *what* we think about which determines our behaviour but *how* we think about it.

This hybrid version of the LOT and ST models conforms nicely to the evolutionary paradigm. The capacity for imaginative projection would appear not only useful in predicting *what* another mind believes and desires, but also *how* the other mind believes and desires those objects or ideas. To accurately simulate how and what another mind is thinking would be extremely practical in, not only predicting, but also in manipulating the behaviour of others. Knowing what others desire and believe, and in what way they believe and desire these things, is to know what actions they are willing to perform in order to achieve their ends. To estimate how another mind's beliefs or desires are structured, is to calculate what another mind will believe or desire. I trust the preceding argument has been successful in demonstrating the truth of the claim that we use simulation to understand novel syntax in mentalese.

Works Cited

Chomski, Noam. *New Horizons and the Study of Language and Mind*. New York, NY: Cambridge University Press. 2000.

Davies, Martin., and Tony Stone. "Introduction" *Mental Simulation*. Malden, Mass: Blackwell. 1996.

Fodor, J. "Why There Has to be a Language of Thought." *Mind and_Cognition.* Ed. William G. Lycan. Malden, Mass: Blackwell Publishing. 1986.

Goldman, A. Alvin. "In Defense of Simulation Theory." *Mental Simulation*. Malden, Mass: Blackwell. 1996.

Gordon, M. Robert. "Folk Psychology as Simulation." *Mind and Cognition*. Ed. William G. Lycan. Malden, Mass: Blackwell Publishing. 1986.

Jackson, Frank. "What Mary Didn't Know." *The Nature of Mind*. Ed. David M. Rosenthal. New York, NY: Oxford University Press. 1991.

Nagel, Thomas. "What it is Like to be a Bat." *The Philosophical Review*. Oct. 1974.

Quine, W.V.O. *Word and Object.* Cambridge, Mass: MIT Press. 1960.