

# Machines and Meaning: Wittgenstein, AI and Creativity

Nis Bojin

Simon Fraser University  
School of Interactive Arts and  
Technology

nbojin@sfu.ca

## ABSTRACT

This work is a brief examination of the manner in which the work of Ludwig Wittgenstein has been taken up in the areas of artificial intelligence and metacreation: an exploration of the way in which his work has been appropriated to advance or inspire efforts in the field of machine thought and creativity. As such, this paper commences by briefly reviewing Wittgenstein's two notoriously well-defined eras of thought, specifically addressing his response to notions of machine intelligence vis a vis the work of Alan Turing. This is followed by a summary and evaluation of thinkers and creators who have adopted Wittgenstein in their work on AI and metacreation, concluding with suggested possibilities for the application of his philosophy in the field of metacreation.

## Keywords

Metacreation, intelligence, language, Wittgenstein

## 1. INTRODUCTION

Ludwig Wittgenstein unquestionably serves as one of the most influential philosophers of the 20<sup>th</sup> century in the fields of language, logic and epistemology and his thinking has continued to permeate numerous disciplines despite the fact that he only ever managed to publish one significant philosophical text during his lifetime. However, it was the migration made by Wittgenstein between paradigms of thought during the span of his philosophical career that serves as one of the most compelling aspects of his life. This paper focuses its examination on the more recent applications of Wittgenstein's early and late work to issues of artificial intelligence and 'metacreation'; *the manner in which machines are designed to exhibit creative behaviors which resemble those creative behaviors of humans.*

why { The rationale of this work is rooted largely in earnest curiosity: an exploratory interest in the applications of Wittgenstein to a field he lived only to see the birth of. As such this work is also concerned with establishing whether Wittgenstein has been taken up in ways which remain either faithful to his vision, or in ways which have perhaps opened up new avenues of interpretation.

This paper provides a brief recounting of Wittgenstein's two eras of thought, followed by a discussion of Wittgenstein's personal reflections on machine intelligence as noted through his hitherto unpublished notes such as his *Remarks on the Philosophy of Psychology* which address

(often inadvertently) the works of Alan Turing and his thesis on artificial thinking (Shanker, 1998). I then proceed to describe and analyze four distinct works which have either taken Wittgenstein up as a practical influence or merely a theoretical accoutrement to their work in artificial intelligence/creativity to discern how and in what ways Wittgenstein has seen use in these applications. I conclude by suggesting possibilities with regard to the future relevance of Wittgenstein's philosophy in the continually burgeoning field of machine intelligence and metacreation.

## 2. THE 'EARLY' AND 'LATER' WITTGENSTEIN

Prior to engaging with Wittgenstein's applications to machine intelligence or to the design of metacreatives, it is imperative to at least briefly outline the two eras of thinking within which he operated; as both of Wittgenstein's own modes of thought have had certain influence not only on the intellectual schools which have embraced him, but also the specific manners in which he has been adopted.

### 2.1 The Early Wittgenstein

Profoundly influenced by the work of Bertrand Russell and Gottlieb Frege, Wittgenstein's initial philosophical pursuits are rooted in an ontological position which attempts to simplify language to its fundamental elements or 'simples'; an approach aligned with the 'logical-atomist' school of thought. It was this thinking which served as the underpinning for his first significant offering to philosophy, the *Tractatus Logico-Philosophicus*-a work which proposed that the world could be linguistically expressed as logical facts that served as the building blocks of larger and more complex facts. It was his aim to use language in its clearest and most precise form to state philosophical facts about the world, positing that through the use of simplified language, one could logically form propositions which stem and build upon these facts without obscurity.

For the early Wittgenstein, language in its ideal state could reflect the world as it was in a logical form: a perfect language could provide an expressible picture of the world and its facts. In this way Wittgenstein dispenses with any notion of meaning as it pertained to conventional metaphysics, given that if something could not be represented symbolically and coherently through a system such as language, that such things should remain unspeakable.

since they cannot possibly make any sense (Wittgenstein, 2000; Finch 2001).

The early Wittgenstein is deeply concerned with recognizing the limits of language and reducing 'possible' thought to the manner in which language was capable of expressing that thought--and the *Tractatus* was a concise and succinct piece which neatly reflected this vision of the world.

It is commonly held that Wittgenstein felt that all of philosophy's problems were resolved with this work and he was highly lauded by the positivist community of the time, but ultimately, Wittgenstein would abandon logical atomism and shift his focus from linguistic certainty to issues of meaning and the impossibility of transcendental certainty and facts (Wittgenstein, 2000; Finch 2001). This departure would cause a rift between himself and his mentor, Bertrand Russell, but his work would remain a fundamental influence among contemporary positivist thinkers.

## 2.2 The Later Wittgenstein

Where the early Wittgenstein argued on behalf of the certainty of language and the manner in which certain essential facts could be linguistically expressed, the later Wittgenstein had some regrets about these assertions. The later Wittgenstein placed a greater emphasis on the uncertainty and 'de-essentialization' of language, arguing that the world was not simply a series of representations capable of being expressed by language, but a series of interpretations and communal understandings which take place through the playing of 'language-games' (Wittgenstein, 2000).

Wittgenstein's detailing of language-games and forms of life supports the idea that a language-game itself is not only a culmination of words and utterances, but a meaningful activity: a practice that intones a particular organic quality and which is rooted in the dynamism of those participating. It is an activity capable of changing, evolving and growing through it very conduct.

A nebulous and yet fundamental concept, 'forms of life'<sup>1</sup> to Wittgenstein is what enables language-games to function as they do: it is the fertile soil that allows the growth and development of language-games and acts as the basis from which language grows and develops. It is the underlying foundation for human understanding and meaningful exchanges within particular conditions and cultural contexts and thus for language-games themselves (Brenner, 1999; Finch, 2001). It is the "common behavior of mankind" (Wittgenstein, 1953, p. 82).

<sup>1</sup> "forms of life" is a concept which, although remains a core principle of Wittgenstein's later work, was still only mentioned five times in the entirety of the *Philosophical Investigations* (Finch, 2001).

"our language can be seen as an ancient city: a maze of little streets and squares, of old and new houses and of houses with additions from various periods; and this surrounded by a multitude of new boroughs with straight regular streets and uniform houses [...] to imagine language is to imagine a form of life"

(Wittgenstein, 1953, p. 8)

Wittgenstein aims to make it clear that the term language-game "is meant to bring into prominence that the speaking of language is part of an activity, or of a form of life", it is the "whole consisting of language and the actions into which it is woven" (1953, p. 11; 1953, p5). It is salient to note that language-games are *embedded within* forms of life: that is to say that language-games are something that occur *within* a particular form of life. Language-games are 'active' and are made comprehensible by the form of life in which they are nested (Finch, 2001).

NOT  
APPLY  
THEN  
?

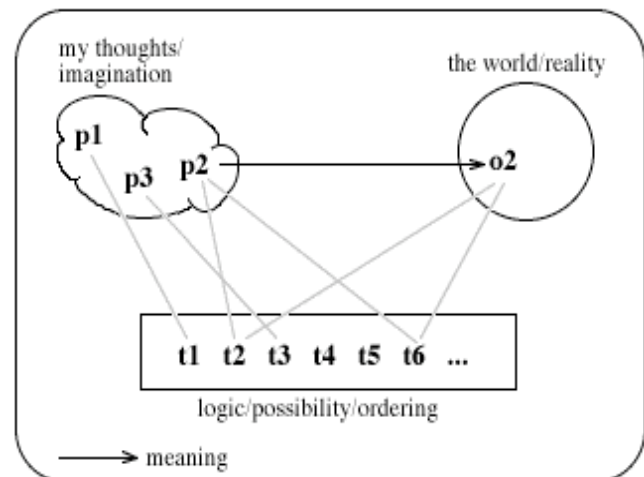


Fig 1. A visual model of the role played by logic in the establishment of reality in Wittgenstein's later philosophy (Randall, 1997)

Language-games are thus constantly protean and culturally situated phenomena rooted in action, in practice, and this notion is exemplified by Wittgenstein's allegory of the 'builder's language': a series of fictional exchanges in Wittgenstein's *Philosophical Investigations*, which depict two individuals communicating with one another in an effort to build a structure from a collection of materials. Builder scenarios feature one individual who gives instructions to the other, who through a common understanding, must retrieve the appropriate materials and supply them as asked. Wittgenstein makes use of the builders' language vignettes in order to explicitly bring the resulting praxis that stems from language exchange to the fore, but he also does so to accentuate in particular how the use of language constructs and reinforces meanings within a particular language-game. To Wittgenstein, language

moulds and massages the contextual reality shared by its users and it is here where the notion of epistemology begins to emerge in Wittgenstein's discussion of language-games.

It should be noted that Wittgenstein in his later work does not trouble himself explicitly with the question of epistemology per se, but with the question of *meanings*. If Wittgenstein's early approach invited philosophers to not speak of those issues metaphysical in nature, Wittgenstein's later approach is one that endeavored to remove the veil of metaphysics and to do away with the a priori entirely. Knowledge according to Wittgenstein is rooted in the meanings established and expressed contextually through language by those who generate and share those meanings (Wittgenstein, 1953; Finch, 2001; Brenner, 1999). There is no greater transcendental knowledge to possess. Wittgenstein believes the word "know" itself defies the shackles of the metaphysical and there can be no proof of any given 'perfect' knowledge outside of a given set of contextualized conventions because such 'superior' or transcendental truths cannot be spoken or expressed (Wittgenstein, 1953; Finch, 2001).

To Wittgenstein, issues of meaning precede issues of knowledge: we already 'know' because we must 'know'--to say we 'know' is in effect an unnecessary doubling or expression of redundancy of that which sense data has already afforded us in every day contexts. H.L. Finch writes that "the certainties of normal human life do not need to be further justified and in trusting them, we make no mistakes for they define what 'mistakes' are just as they define what 'knowing' and 'doubting' are" (2001, p. 113). Language becomes the form with which we express and describe our knowledge and which subsequently reinforces it. The meanings and descriptions that become associated with words through the use of rules consequently shape conventions: the shifting rule-based foundations of language-games. It is language and its meanings that serve as the supporting pillar of our own epistemologies and as a result, language can cement a given epistemology and paradigms within it through the adoption and use of its conventions. Wittgenstein cites the example of philosophy itself, noting that the reason we are "still occupied with the same philosophical problems as were the Greeks...[is] because our language has remained the same and keeps seducing us into asking the same questions" (1980, p. 15).

### 3. WITTGENSTEIN AND TURING

With a sense of Wittgenstein's legacy in tow, this brings us largely to the hitherto unpublished work which has unveiled a series of the previously unearthed lamentations of the Austrian-born engineer and logician as it pertains to the concept of machine intelligence.

More specifically, research into the formerly unpublished work of Wittgenstein has permitted us some insight as to the responses he had to contemporaries such as student

Alan Turing, who notably attended a series of Wittgenstein's own lectures on the foundations of mathematics (Shanker, 1998). Given the aim of this paper, it becomes crucial to have some sense as to what Wittgenstein's response was to the work of Turing and work which was founded on the premise of machine intelligence and calculation. The explication of such a premise must begin with the statement of Turing's general thesis of machine intelligence.

#### 3.1 Turing's Thesis

Turing's computational theory of the mind, espoused in his work, *On Computational Numbers*, states that we can unequivocally compare a human being in the process of computing a real number to a machine which has at its disposal, only a finite set of conditions or configurations (Shanker, 1998). A given computation then is dependant upon a symbol being observed (and subsequently recognized) along with the state of mind which is used to compute that symbol—in other words, knowing what to do with that symbol given a particular state (Shanker, 1998). To Turing, this ability could qualify machines as the computational equivalent of humans which are capable of the same calculation, with an emphasis on the similarity of output, and with a reduced importance placed on the minute details of the 'processing' of such calculations. For Turing, such details are essentially irrelevant as they simply obscure the empirically established inputs and outputs of the computational process at its most fundamental level. The mind for Turing was a result of its operations.

It should be noted of course that Turing's thesis was just that: a posited concept. There was never a physical prototype or working example of Turing's machine, his thoughts on the matter were always expressed mathematically in the form of principles, which he believed would hold should they ever be realized. Turing's machine was typically visualized as a strip of tape containing data that would be read by a machine head, which had its responses to that data determined by a rule-based 'state machine' (fig. 1). For Turing, it only mattered that it could be conceptualized that the same operational outputs of a machine could resemble those of a human-*how* that actually occurred was of significantly less importance for his 'thought experiment'.

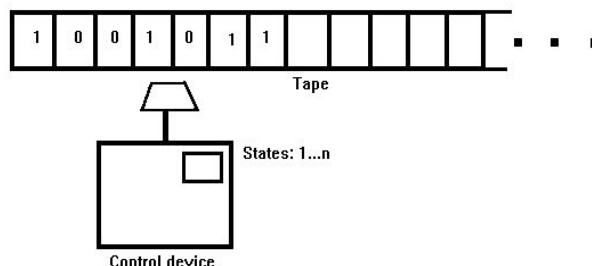


Fig 2: Model of the Turing Machine: A Tape Strip Being Read by a Head, which is in turn connected to a State Machine

One might argue, already being aware of the later Wittgenstein's thinking, that Wittgenstein himself would dismiss such a thesis. So it is perhaps unsurprising that Wittgenstein's *Remarks on the Foundations of Mathematics* contains mentions of Turing's machine, and with it, a lengthy, but non-directed refutation of Turing's notion of calculation :

"Imagine that calculating machines occurred in nature, but that people could not pierce their cases [...] think of the mechanism whose movement we saw as a geometrical proof: clearly it would not normally be said of someone turning the wheel that he was proving something. Isn't it the same with someone who makes and changes arrangements of signs as an experiment; even when what he produces could be seen as a proof?"

(Wittgenstein RMF in Shanker, 1998)

The lack of understanding, the submission to a normative practice, does not necessarily prove the ability to calculate, but simply the ability to execute (not follow) particular pre-determined rules—nor can we ever distinguish between an entity which may reach the correct result through experimentation rather than acknowledged process. Following rules to Wittgenstein is something that generally requires a justification, otherwise the following of such rules would be indistinguishable from serendipity or blind execution of a process.

Execution of a process is different from calculation in the sense that the latter requires a determination as the former requires only a pre-determination. This blind pre-determination is what Wittgenstein might liken to his notion of 'bedrock'; the point at which rules and practices in a language-game are simply followed because this is simply what must be done (Wittgenstein, 1952). Of course reaching bedrock without having to have previously exhausted any justifications is the equivalent a move in a language-game which has already been played before that move was ever made. Turing's machines could execute rules, but could not produce anymore than the outputs which were predetermined by the machines creators, the rule's creators, those whom Wittgenstein would label the true calculators of a machine's output. In the words of Shanker, "a mechanical device for sign manipulation is no more a 'rule-following beast' than an abacus" (1998, p. 32). Or in Wittgenstein's words, Turing machines are more aptly describable as 'humans that calculate'. (Shanker, 1998).

However, Turing's thesis was directed at describing what occurs at the neural level and not necessary a level of consciousness required for making formal calculations—he wished to get at the conceptual basis of those calculations which made thought possible.

### 3.2 'Thinking' and Family Resemblance

With the above rebuttal in hand, Wittgenstein would proceed to address Turing's question regarding the possibility of machine thought. Could machines think? Is there a mechanical explanation for thought? These two questions ask two different things, but for Wittgenstein, addressing the notion of the concept of 'thinking' would have to precede grappling with either of them.

Wittgenstein argues that any process of necessity (such as the general calculation of instructions by a human) comes with it some pretext of what that process departs from: what being 'incorrect' would mean, why 'incorrect' is truly incorrect and not merely recognizing 'incorrect' as a pre-assigned value. This attacks the lack of what might be considered requisite consciousness for 'thinking', but Wittgenstein takes a pause from this stance, directing his attention to the assumptions Turing makes in his thesis.

Turing's thesis is one which sidesteps the ambiguity of cognition, as his larger concern is with the manner in which the mind functions for purposes of computing instructions, not contemplating them. Focusing on brute force computation with an emphasis on the manner in which such computation leads to learning, Turing wished to emulate the core essence of the mind's ability to calculate with the context of such calculations remaining irrelevant to the ground-level process of the calculation itself. That said, Turing's question of whether machines could actually 'think' (less learn) is a prominent point of contention for Wittgenstein, as it is a question which presents a potentially violent collision between philosophy and empiricism. The question instead for Wittgenstein then is not 'can machines think?', but 'what is thinking'? How can one ontologically discern what constitutes 'thought' (a thorny endeavor on its own) and merge this notion cleanly with what one's sense experience is of the process we call thinking? Further to that, how can we conceive of what the concept of thinking actually is: how can we conceive of concepts at all?

The concept of "thought" is something which in and of itself was necessarily tenuous for Wittgenstein since it involved a thorough analysis of the manner in which that concept has been derived through the exchange of language, communal acceptance and the assessment of various 'family resemblances': resemblances which exist between activities that we tend to consistently associate with thought and thinking (Wittgenstein, 1952). It also requires that we have a sense of how concepts are formed to begin with: an evaluation of the concept of a concept (Wittgenstein, 1952; 1978; Shanker, 1998).



Wittgenstein previously argued that the elements of games, such as rules all fell short of defining what games actually are. He consequently asserted that the concept 'game' could not be contained by any single definition, but that games must be considered as the product of numerous fluid definitions that share a "family resemblance" to one another (Wittgenstein, 1952). As such, to discuss the idea of a machine that could think would first require us to agree on those things which were held in common between activities we called 'thinking' or even 'intelligence' (Wittgenstein, 1952; Finch, 2001). But such concepts are conceived of through comparison, perceived similarity, context etc.. Not to mention that the derivation of family resemblances never stem from an unimpeachable level of objectivity, but an assumed context, a paradigm of inquiry, an epistemological foundation. If we were to generate a concept of what constituted 'intelligence' we would not need to go far before we realized that definitions of this concept would range between varying disciplines largely because of the value assigned to any one of the multiplicity of resemblances between any conceivable instance of 'intelligence'. And although 'thinking' or 'intelligence' might appear to represent the most complex example of an evaluation of the concept of a concept, Wittgenstein himself uses much simpler examples to make his point:

"How did we learn to understand the word 'plant', then? Perhaps we learnt a definition of the concept, say in botany, but I leave out that of account since it only has a role in botany. Apart from that, it is clear that we learnt the meaning of the word by example; and if we disregard hypothetical dispositions, these examples stand only for themselves. Hypotheses about learning and using language and causal connections don't interest us [...] if the examples should have an effect [...] the causal connection between the examples and this picture does not concern us, and for us they are merely coincidental."

(Wittgenstein, 1978, p. 117)

This passage (and others like it in the *Philosophical Investigations*) is critical simply insofar as it establishes how concepts are created--not how concepts are concretely fashioned--but instead how they are 'played' with in a language game: there is no sense determinacy, and as such, the game play is a constantly shifting and undulating process which wavers ever closer to determination, but never reaches it. This very process defies rigidity and instead reflects more of the Heideggerian notion of mindfulness: a steady cognizance of the shifting elements and contexts from which a given concept has emerged in order to come to a greater understanding of that concept, only ever coming into fleeting contact with any transcendental knowledge associated with it (1976).

Returning then to Turing's question of thinking, Wittgenstein then finds the question of whether machines

think somewhat moot. Instead he would first question what all things which we might associate with thinking had in common and ask us to consider those resemblances and the contexts from which they were born, ultimately concluding that seeking determinacy here would be an ill-advised pursuit.

In then considering whether there was a mechanical explanation for thinking, the question would instead be 'how is it that we can describe machine computation as thinking' in an effort to unearth any and all similarities. Undercutting Turing's assumptions regarding 'thought' as a static ontological concept, Wittgenstein disarms the question of thought at its base rather than the possibilities that it implies. As such, Wittgenstein never denies the possibility of machine thought or creativity, but simply makes the argument that we are several steps back of being able to even approach the notion.

This is not to say that Wittgenstein himself did not have a concept of 'thought'. His later philosophy actually discusses the concept of 'thought' at length, describing thought and propositions as being represented by mental pictures, which are subservient to the requisite language-use and context needed to express their meaning (1953). However this concept once again addresses of a mind which encompasses more than the ability to calculate but to 'envision', which is where Turing would suggest was not an area of consideration with his own work.

Regardless, Wittgenstein rebuts Turing on two fronts:

a)- claiming that a machine cannot naturally emulate the human mind, but merely execute a process a which is already pre-determined to comprise a given 'emulation'.

and

b)- that contemplating machine 'thought' requires us to first come to grips with what such a concept means and how we generate such concepts using language. Until such time that we can answer those particular questions, querying the possibility of machine thought becomes ultimately ineffectual.

#### 4. WITTGENSTEIN, AI AND METACREATION

Despite any refutation of Turing's thesis by Wittgenstein, a number of thinkers have adopted Wittgenstein's thinking in an attempt to bolster particular theories regarding AI, establish new theories, or re-conceptualize old ones. I now look at three different appropriations of Wittgenstein and assess the manner in which Wittgenstein has either been taken up accurately, innovatively or spuriously. Given that Wittgenstein is not altogether an opponent or proponent of the possibility of artificial intelligence or machine creativity it remains to be seen whether anyone who has managed to use his work has done so faithfully.

## 4.1 Wittgenstein, Semantics and Connectionism

The paper entitled, *Wittgenstein, Semantics and Connectionism* by Goldstein and Slater offers the suggestion that Wittgenstein (the later) presents several arguments which attempt to deflate common notions of symbolic representation as being the ideal process through which human beings are modeled for purposes of programming and designing artificial intelligence. More acutely, it is the thesis of these authors that ‘connectionist’ theory is more resilient against the criticisms of Wittgenstein, thus making it a more adequate model for believable AI.

The ‘symbolic paradigm’ as it is referred to throughout the paper, is that dominant mode of thinking which has remained stalwart in the computer sciences as it pertains to machine intelligence. However, as Goldstein and Slater note, such a model generally reduces the human mind to that of a processor of representations: a computer which sterilely applies sets of rules to sentences and stimuli without the consideration of context unless such a context is represented beforehand in a pre-coded fashion (1998). Connectionism<sup>2</sup>, on the other hand, although still considered a computational theory, generally looks at ‘mental’ phenomena as being generated through networks of smaller, atomic components, which work in concert with one another to generate a particular truth or answer.

Goldstein and Slater posit that Wittgenstein’s importance on communal agreement as well as the generation of truths within particular paradigms of thought support the idea of connectionism:

“There is nothing syntactic in the brain which already has a semantic relation with the world, since that semantics is not given prior to any consensus arrived at communally. As Wittgenstein stressed, community of response, agreement in judgements is the final court of appeal [...] the locus of truth is at the semantic level[...].”

(Goldstein and Slater, 1998, p. 298)

It is here where Goldstein and Slater draw on Wittgenstein’s emphasis on the use of words as the source of their meaning, assessing that computational symbols in AI are not actually used in any semantic sense beyond that of basic recognition and execution. They suggest that for artificial intelligence, and its subsequent applications, meaning cannot be derived from one-dimensional recognition and execution of a command: particularly

because that execution is predicated on a single, pre-determined value or value system. In such systems, it is not meaning that is being exchanged between particular agents, but instead pre-defined language-games whose possible moves are simply acted out, rather than played. This particular stance echoes Wittgenstein’s own rebuttal to Turing’s thesis, but Goldstein and Slater take their comparative analysis of symbolic representation and connectionism one step further, probing the nature of thought and thinking.

Although Wittgenstein had originally spent much of his rebuttal discussing the epistemologically fluctuating nature of thinking and thought, Goldstein and Slater move to take up some of Wittgenstein’s reflection on the role of language in thinking—not to discern what thinking actually is, but instead to reinforce the importance of language as it pertains to the expression of those things which are considered as distinct moves in a language-game. As such, language usage reveals particular weaknesses of a strictly symbolic approach to computation, namely that the acquisition of a propositional attitude cannot be solely “construed as gaining access to, and adopting an attitude towards such symbolic representations” (Goldstein and Slater, 1998, p. 308). There cannot be half a ‘thought’ which leads a representational system to only half think that, for example, someone must be very hungry. However, connectionist structures such as neural networks are capable of training, being receptive to weighting, stimulation, environmental exposure and other factors, which and produce a much more analog determination of how hungry someone might be (again, if the system was being trained to look for such a feature). Such structures, as Goldstein and Slater would argue, are more effective than symbolic structures because of their repeated engagement with the world which generates a consensus based on the various layers of a neural network.

But what are Goldstein and Slater actually committing to with regard to the application of Wittgenstein here?

The argument is made in this paper that connectionism is surely supported to a greater extent by Wittgenstein than would be symbolic representation, but the manner in which Wittgenstein’s philosophy extends to a rationale for the relative success of neural networks in relation to symbolic representation is questionable.

On the one hand, Goldstein and Slater concede that calculation is merely one facet of the human brain: suggesting that the symbolic representation model of computation inaccurately reduces the human brain to a processor or calculating device. Goldstein and Slater instead choose to refer to the human mind as a ‘sensory-motor processor’, which sees the value of the mind in its engagement with the social world and its ability to adjust and learn. With neural networks being painted as a more ‘sympathetic’ version of the human mind this way, one

<sup>2</sup> Alan Turing is typically given credit for the anticipation of the connectionist approach, and more specifically, neural networks, which raises the irony of Wittgenstein be called on to defend connectionist thought given his general disagreement with Turing’s approach to conceptualizing machine intelligence.

cannot help but fathom that Wittgenstein has still been utilized here to actually reduce the human mind, not to a symbol processing machine, but to a slightly more advanced structure which responds and learns from its environment in spite of its shortcomings with regard to rationality and logic (1998).

Wittgenstein himself (particularly the late Wittgenstein, adopted by the authors) was a staunch antagonist of essentialist thought, but his work has been somewhat distorted to conform to essentialist thinking here. And although his work has at least been used to ‘de-essentialize’ current essentialist thought embodied by the symbolic representation approach, the conclusions of Goldstein and Slater still situate the human mind as something only slightly less reduced than the model of computation against which they claim to be rallying.

The conception of essence was anathema to Wittgenstein, and much of his later work did much to try and dispel essentialist thinking—much of which is rooted in either metaphysical or reductionist thought. And despite that Wittgenstein has been used here quite faithfully in terms of the recounting of his actual arguments, the application of these arguments is rudimentary at best, failing to adequately capture much of what it was that Wittgenstein had (by most accounts) intended to say. Of course, this suggests that perhaps Goldstein and Slater have simply appropriated Wittgenstein in such a way which corresponds to a particular interpretation of his work, but unfortunately this reasoning does not exonerate the authors here given that neural networks is still a model of computation which Wittgenstein’s theory would not subscribe to for the exact same reasons that he did not subscribe to Turing’s machine thesis. In other words, neural networks do not differ from symbol processing programs in the sense that humans are still providing the calculations required for neural networks to perform their function. The performance, or execution of the human-provided direction is the only thing neural networks manage to accomplish: Wittgenstein’s human calculating machine, but this time reflecting more complex approaches to execution of directives.

Goldstein and Slater adopt Wittgenstein in what I would refer to as a relatively romantic manner: marrying the world of non-essentialist philosophy with an approach to AI computation which defies the severity of symbolic processing, but which doesn’t manage to entirely shrug off its own essentialist aspects. Although connectionism may be significantly more intimate with Wittgenstein’s later thinking, it certainly is not an ideal fit.

## 4.2 Using Family Resemblances to Learn Exemplars

Vadera et al’s work on machine learning and Bayesian networks have turned to Wittgenstein not to adopt his general philosophy en masse, but to demonstrate the

application of his concept of ‘family resemblances’ to a machine learning context (2007).

The short work presented by Vadera et al outlines the way in which ‘family resemblances’ can be used to learn exemplars as Bayesian networks represent and use them in some capacity (2007). Exemplars themselves are the prime examples of a given thing: the stereotype of a phenomenon. Using a propagation algorithm, Vadera et al establish a series of potential exemplars in a given category and express them visually in a Bayesian network. Presented with the challenge of determining, and subsequently predicting the ideal exemplar among things of a particular type, Vadera et al turn to Wittgenstein’s family resemblances.

The rationale behind Wittgenstein’s usage here is as follows:

“Wittgenstein postulated that words were characterized by categories of example uses and that the categories could be based on the principle of family resemblance. [...] this kind of paradigm shift is necessary if we are to go beyond the capabilities of current systems and develop socially sensitive systems.”

(Vadera, 2007, p. 69)

Given that several authors cited by Vadera et al have also adopted family resemblances in their work, they re-appropriate some of the categories of validation which other researchers have used before them; namely ‘focality’ and ‘peripherality’ (2007).

Focality refers to an exemplar which has a high family resemblance with those things it represents while peripherality is indicative of that same ‘thing’ exhibiting low resemblance to those things outside of the family (Vadera et al, 2007).

Using Bayesian networks to represent learned exemplars and using family resemblances to determine the ideal exemplars among them, the authors assert that scenarios can be predicted with significant accuracy. In this case of this particular work, the manner in which political party members would vote and the identification of zoo animals based on their features were the scenarios of choice. The percentage of successful predictions was relatively high for each of these scenarios. For example, Republican voting predictions (predicting which candidates Republicans were most likely to vote for based on particular criteria) using Vadera et al’s family resemblance approach were reported as being 96 percent accurate (84 percent for Democrats) (2007). Family resemblance was interpreted as:

“the probability of an exemplar representing a point, which in turn can be used to compute focality and peripherality. The difference between the focality and the peripherality can then be used as a measure

of the prototypicality of the exemplar.”  
(Vadera et al, 2007)

By the authors’ own admission, their family resemblance approach to machine learning of exemplars ends up resulting in an ‘overtraining’ outcome similar to that of neural networks or decision tree learning, but is otherwise considered a successful approach which they believe to demonstrate promising results.

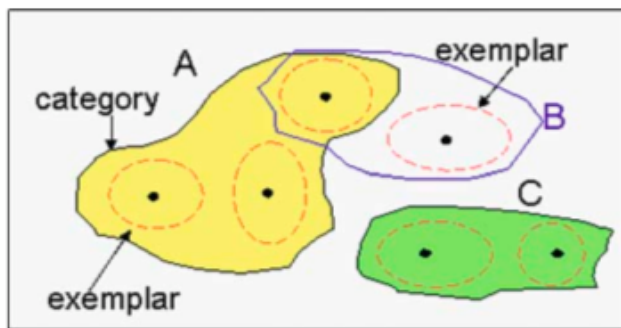


Fig 3: Visual Representation of Bayesian Grouped Classes and their Highlighted Exemplars (Vadera et al, 2007)

The research presented here is relatively straightforward, even if severely lacking in detail. However, the one polemical issue is clearly that of Wittgenstein’s misappropriation. There are a number of deficits both in the rationale behind, and consequently in the use of Wittgenstein’s concept of family resemblances that is never addressed.

Firstly, the rationale behind Wittgenstein’s use is admirable *prima facie*, but in the end, entirely unnecessary. The reasons behind the use of family resemblances are rooted in the authors perception of value in the ability to identify the ideal example of something because of how well that example resembles other things in its ‘class’. Unfortunately, the vague similarity of this particular concept to Wittgenstein’s notion of family resemblances does not make them the same, nor does it lend the former any particular credence. The authors’ rationale simply does not draw on what makes Wittgenstein’s family resemblances a unique concept.

Secondly, and stemming directly from the previous point, is the invariable conclusion that Wittgenstein’s concept of family resemblances is entirely misrepresented by the authors. Wittgenstein’s concept of family resemblances does not speak to the application of unveiling a particular example of something. On the contrary, family resemblances are what allow us to compare between phenomena because we *cannot* determine the exemplar of that phenomena. Family resemblance is a concept used in order to come to recognition of the lack of an ideal of a given thing, because the ideal of a given thing would

permit us to define it, and the very use of family resemblances is founded in our inability to define a great many things (Wittgenstein, 1952).

Family resemblances permit us to establish meaning without reference to a particular specific exemplar or perfect ideal--privileging the meaning derived from a lack of specificity and understanding the meaning behind the imprecision.

Thirdly, as previously mentioned, Wittgenstein does not subscribe to any method which seeks to reduce phenomena to various atomized criteria. The very concept of a Bayesian ‘ontology’ is representative of what Wittgenstein attempted to migrate away from, believing that mutability of meaning through use and exchange prevents us from coming to a concrete determination as to what the primary constituents of *anything* would be. One participant’s perception of what ‘patriotism’ meant in their determination of what sort of candidate they were more likely to vote for, covers a gamut of various interpretations and uses in language-games. There can be no truly determinable category of ‘patriotism’ less an ideal determination of who the most likely candidate would be based on what participants said about the role of patriotism in their selection of a candidate. That said, the authors provide very little information as to the elements which made up their Bayesian networks, and as such, I cannot comment further on the specifics of their criteria.

Regardless, Vareda et al, adopt family resemblances as a method for ascertaining what they perceive to be an attainable exemplar from their data: the frequent unlikelihood which makes family resemblances a necessary concept in Wittgenstein’s philosophy to begin with. In other words, a complete and unfortunate misunderstanding of Wittgenstein’s work makes his incorporation into this work relatively ineffectual.

### 4.3 Post-Human Creativity

In this work, Rolf Hughes explores the idea of current metacreative efforts as those which exist at the liminal boundary between chaos and order--embracing both order as well as a necessary element of surprise. This invariably re-invokes the classic paradox: the manner in which chaos that comes as an intentional result of orderly processes, is both chaotic and orderly. Hughes essentially asks “how do we move past this notion and broaden our conceptions of metacreation moving forward?” (2005, p.1).

This particular works approaches AI and metacreation from quite a different angle. Whereas Goldstein and Slater attempted to use Wittgenstein to validate one approach to computation over another, and Vareda et al adopted Wittgenstein to lend credence to an algorithmic approach which only vaguely resembled Wittgenstein’s actual thought, Hughes references Wittgenstein at the purely metaphorical level.



There is no advancement on Hughes's part to literalize any aspect of Wittgenstein's work into an algorithm, no endeavor to use Wittgenstein to theoretically bolster some aspect of machine learning. Instead, Wittgenstein is drawn into the argument to highlight the importance of the use of simile and discovering connections outside the realms of pre-ordained computational canon in conceiving of further metacreative efforts.

Hughes argues that in considering creativity in the AI community, we need to challenge notions of what creativity is as such notions are firmly rooted in particular discourses, legitimizing institutions, economic paradigms and other institutions of ownership which favor particular approaches to creativity and design (Hughes, 2005). These boundaries are tightly defined by the confines of our language use and rhetoric and as such, the manner in which we model through metaphor and through language, determines heavily the way in which we formulate our conceptions of the creative. Subsequent creations become 'performances' of those very models and become reflective of our conceptual combinatory abilities, allowing metacreative discovery through a change in perspective—not necessarily a change in technology.

Hughes's vector of thought utilizes Wittgenstein in such a way, which makes the primary locus of discussion metacreation. With the two previous works discussing issues of artificial intelligence and machine learning, this work directly addresses metacreation and does so without any explicit interrogation of computational or algorithmic minutia.

As such, Wittgenstein's use is likely the most apt in this piece, given the general focus on epistemology and rhetoric: those things which influence our given paradigms and shape the manners in which we feel we can operate as author, creator and metacreator, citing that the most truly metacreative projects have emerged from outside of the formalized boundaries of standard computational paradigms and technological grammars, lauding efforts in artificial life as some of the most unique, but more generally any endeavors which continually push and bend at the membranous boundaries of the 'hard' disciplines which tend to keep metacreative endeavors held back.

"All boundaries are at risk. Since there are as a result a growing number of problems without a discipline, this skill in *seeing connections* – a skill that fuses creative and critical modes of inquiry (or curiosity) – will become increasingly important. Philosophers such as Wittgenstein – whose work stages or frames an artistic *performance* of philosophical problems – sought "just that understanding which consists in 'seeing connections'". He described Freud's work admiringly as comprising "all excellent similes." and

said something similar of his own work in philosophy: "What I invent are new *similes*." [...] juxtaposing two or more concepts together to create a new conceptual whole with unforeseen, emergent properties [...]. The philosophical dialogue can be said to operate in a similar manner."

(Hughes, 2005, p.11)

Hughes cites several examples of work which pushes at the metaphorical boundaries of what is possible, despite whether such efforts are successful or not. One such example, MEART (referred to as a semi-living artist) consists of neurons from an embryonic rat cortex which has been grown over a multi-network array (Ben-Ary, 2007). These 'cultured' nerve cells communicate through software with a remote drawing arm which makes scrawls on a piece of paper in another part of the world. Rather than reproducing a program modeled on neural networks, such work attempts to turn to actual neural constructs to see what we can learn from them as the constructs learn on their own. The MEART also has a camera which serves as its eyes so that there may be some sense of its environment, at whatever level that may be.

MEART serves as an example of a metacreative endeavor which aims to create the artist rather than re-create the outputs of an artist—an approach in direct contrast to that of Turing who privileged only the capacity of a machine to produce results that matched those of a human (Ben-Ary, 2007; Shanker, 1998)



Fig 4: MEART's robotic arm in action with its artwork displayed on the wall in the background (Ben-Ary, 2007)

Citing examples such as MEART, Hughes draws on Wittgenstein as a prime example of a thinker who continually advocated on behalf of epistemological

dislodging<sup>3</sup> for the advancement of new thought, but more importantly, the discarding of old thought which serves to do little more than keep us firmly planted in the bedrock of complacency (1952). And this is Hughes's central thesis: avoiding complacency and embracing the agency of the author to play with metaphor, to shatter dominant rhetorics which keep us entrenched in particular computational paradigms, but also hinder our metacreative pursuits (Hughes, 2005). However, it is distinctively the 'seeing of connections' which although a seemingly vague reference, is consistently echoed throughout the work of Wittgenstein. As Wittgenstein himself might advocate, it's not about generating 'new' forms of knowledge, it's about changing the way we look at what's already there (Wittgenstein, 1953).

#### 4.4 The Aesthetic Robot

There yet remains the aforementioned issue of requisite consciousness or sentience which underlies the incapability of machines to comprehend language in all of its nuance: in context and with affect. Although this point is not relevant to Turing's thesis due to the strict limitations on what Turing intended to address, it remains a relevant component of Wittgenstein's general refutation of the much of the work being conducted in machine intelligence.

In his paper, *Wittgenstein and the Aesthetic Robot*, Julian Friedman picks up on Wittgenstein's mode of thought in relation to what he refers to as 'the aesthetic robot': a machine which has the ability to appear as though it is thoughtful, creative and intelligent, but which lacks the ability to provide more than aesthetic imitation of these things.

To Friedman, what a supposedly creative or intelligent machine lacks is an affective life of its own: the ability to use context to judge the difference between various scenarios despite the similarity of the actions which might be taking place between them. For example, a machine cannot judge the difference between someone who is pacing back and forth because they're stressed and someone who is performing the same activity because they are being histrionic or playful. A creative machine can only exhibit pre-determined behaviors—it cannot be inspired or affected by its environment (2005).

Friedman raises the idea of machines that can determine what types of music we would like best based on previous examples, positing that some might claim that we are affected by particular melodic patterns, chord types and instrument timbres which are composed in such a way to make our ear and brain desire to hear a specific arrangement of notes (Friedman, 2005). Friedman

continues, noting that "musicians often speak as if this were the case, maintaining that good stirring melodies simply play with our auditory desires by first stimulating them and then artfully satisfying them at crucial moments in a way that results in emotive arousal" (2005, p. 190).

As Wittgenstein ponders, "if you feel the seriousness of a tune—what are you perceiving?—Nothing that could be conveyed by reproducing what you heard" (1953, p. 210). A creative or intelligent machine cannot distinguish between descriptive and expressive instances without any pre-determination of specific values or weights to 'tend' the machine in the direction closest to the most accurate interpretation it is capable of. Even then, such an approach would simply be programming for what are arguably affective stereotypes and not authentic reflection. Although Friedman concedes that perhaps any feelings a human experiences about a particular piece of art or music is perhaps traceable back to subjective mental states, he asserts that certain dispositions remain unaccounted for. One observing something as a particular color requires a certain disposition as does one acknowledging that color and it's particular hue, intensity, saturation etc. as one's favorite. How does one acquire new favorites? How does a creative machine which can paint, choose to change its style based on new personal preferences that are garnered from experience or exposure to new contexts without it simply being pre-programmed to acknowledge traits of newly encountered aesthetic styles into its repertoire? In other words, to have a preference of something (style, technique etc.) does not mean that this preference satisfies a particular desire to have that thing be preferred. Desires are often rooted in unquantifiable, dynamic and evolutionary ways. Desires are frequently organic and undeterminable and creative machines have yet to achieve this state.

"Hence, coming to understand, say, Gustav Klimt's tri-partite visual depiction of Beethoven's Ninth Symphony may have a persistent effect on one's emotional and aesthetic awareness in a way that goes well beyond future encounters with that particular composer's work. Our cognitive habits in this way undergo continual spontaneous alteration. This truth paradoxically renders the Turing test all the more perspicacious [...]"

(Friedman, 2005)

Thus for Friedman, if we could successfully say that a computer was imitating us as we actually lived, rather than a mimicry of outputs, we would see machine intelligence and creativity become a remarkable reality. This is clearly easier said than done, but Wittgenstein's privileging of sentience in AI re-emerges here in Friedman's work, and one of these sentient properties is that 'truths' and 'desires' be generated as self-motivated on the part of a given artificial agent. That those things held as reality are

<sup>3</sup> The attempt to break out of a dominant framework of 'knowing' through the mindfulness of language and its use in language-games.

determined by the machine and not by us; by community, not in solitude. As Turing's thesis largely implies a machine's mastery of intentional nuance, Friedman, vis a vis Wittgenstein, is looking at the independent generation of nuance as essential to metacreation.

## CONCLUSIONS

What this paper has endeavored to accomplish is three-fold:

- 1- to briefly summarize Wittgenstein's early and later eras of thought
- 2- to briefly encapsulate the arguments Wittgenstein brings to bear against Turing's thesis of machine intelligence
- 3- to examine a series of works which have attempted to adopt Wittgenstein in their practical or theoretical approaches to artificial intelligence and creativity and evaluate the appropriateness of their use of his philosophy.

Wittgenstein makes is rather clear, as this paper has hopefully demonstrated, that conceiving of the human mind as a machine is not conducive with either his early or his later philosophy. However, what are the potential applications of Wittgenstein to artificial intelligence and metacreation?

It might be said that much like the work of Hughes, Wittgenstein's work may never translate literally into any form of computational model, but this does not exempt the work of Wittgenstein from being used as the metaphorical or conceptual flagstones for new forms of thinking in computation.

What would a non-reductionist approach to computation look like? How do we use family resemblances as a notion which does not just become errantly translated into an approach to algorithmic machine learning, but which becomes translated into a larger, high-level perspective on how to reconceive of artificial intelligence and metacreativity on the whole? How can the notion of language-games be used in ways which do more than simply invalidate the dominant approach of symbolic representation, but instead help us re-conceive artificial intelligence and machine creativity entirely?

In response to Turing's thesis, Wittgenstein acknowledged that too many steps, too many assumptions had been made in generating a thesis which presupposed intelligence as the output of a computational process designed by humans. But what Wittgenstein himself might ask is how we can take a new look at things that have always been the--how we can take a few steps back and re-evaluate the directions we have taken and re-consider what it truly is that all thinking, all acts of creativity have in common. It is difficult to conceive of even a form of computation which

is not modeled on the current paradigm of binary computation, on the standards which have been set for thinking about machine intelligence and creativity. How do we back up? How do we see new connections?

Wittgenstein once mused that his transition from logical-atomism, to the more epistemologically centered thinking of the *Philosophical Investigations* was akin to sitting in an empty room on a chair, only to realize that he'd yet to turn around in his seat to see the door leading out of the room (1953).

As such it may be that Wittgenstein's greatest offering to either artificial intelligence or metacreation is the notion of the 'epistemologically dislodging' act: the movement away from old conventions of thought which come about through a refreshing of one's perspective on the contents of a given paradigm and thus a reconsideration of the language-games used to uphold that paradigm. It was mentioned at the onset of this paper that Wittgenstein's movement from one pole of thought to another was one of the things which made his work most compelling. But how do we break away from current modes of thinking about machine intelligence? How do we reconceptualize what it means for a machine to be 'creative'? We are so tightly bound by particular conventions that an epistemological shift seems nearly impossible.

For Wittgenstein, we accept particular conventions as normative, but we do not agree to the meaning of these conventions because we necessarily understand them. "Meaning is not a matter of our assent. Without conventions, we could not understand in the first place" (Smit, 1991, p. 36). We tend to understand something because we become exposed to a series of contexts, activities and circumstances in which we establish the range of permissible meanings in and among a consensus (Wittgenstein, 1953; Smit, 1991; Finch 2001). So how do we expose ourselves to new contexts? What is keeping us rooted in our current paradigms?

A language-game approach would offer that we persuade one another--through agreement and concession, participants in a language-game establish, change and support rules as they traverse the "shifting landscape" that is the play of the language-game (Wittgenstein, 1953; Finch, 2001; Smit, 1999). We may either thoughtfully or capriciously consent to acknowledge or ignore, but we do so in the context of the various forms of life that pervade our daily activities whether it be the socio-political context of our daily living or the corporate environs of a given workplace. As Hughes explicated, we are bound by our discourses, our socio-economic context, the influences of legitimizing institutions and so forth (2005).

Thus, although we may be fully literate in the semiotic domains that we construct, whether it's the domain of metacreation or otherwise, we have internalized judgments

inseparable from a mastery of our distinctive discursive competences, and in this way, without any explicit argumentation, justification or challenge, we have become embedded within our praxis. It is not until we can engage in a communal reflexivity that conscious change can be truly made, and Wittgenstein recognizes the potential for conscious change both through mindfulness as well as powerful, controlling ideological agendas which can take hold through disinterest and apathy (1953).

With this in mind, it becomes evident that Wittgenstein's use in furthering work in artificial intelligence and creativity is not entrenched in any literal conversion of his philosophy into some algorithmic form, but in reconceptualizing those forms and attempting to remain mindful of our use of them and the language we use to speak about them (1953; Heidegger, 1976).

For Wittgenstein, it would seem that an epistemologically dislodging act is possible, but such an act cannot be committed privately: it requires a willing constituency who are willing to play a new language-game entirely or at very least, play the same game with different rules. This exercise of course requires one to become reflexively aware of the language/praxis one has already committed to; turning around in our own chairs and seeing the door behind us.

## 5. REFERENCES

- [1] Ben-Ary, Guy (2007). MEART: The Semi-Living Artist. Available at: <http://www.fishandchips.uwa.edu.au/project.html> (Last accessed: April 2nd, 2008).
- [2] Finch, H.L. (2001) *The Vision of Wittgenstein*. London: Vega
- [3] Friedland, J. (2005). Wittgenstein and the Aesthetic Robot's Handicap. *Philosophical Investigations*. 28 (2).
- [4] Goldstein, L & Slater, H. (1998). Wittgenstein, Semantics and Connectionism. *Philosophical Investigations*. 21 (4).
- [5] Heidegger, M. (1976). *What is Called Thinking?* London: Harper.
- [6] Hughes, R. (2005). *Orderly Disorder: Post-Human Creativity*. Proceedings of the Linköping Electronic Conference. University Electronic Press; Linköpings universitet
- [7] Randall, A. (1997). *Logic, Idealism and Materialism in Early and Late Wittgenstein*. Available at: <http://www.elea.org/Wittgenstein/> (Last access, April 3<sup>rd</sup>, 2008)
- [8] Shanker, S. (1998). *Wittgenstein's Remarks on the Foundations of AI*. New York: Routledge.
- [9] Smit, D. W. (1991) *The Rhetorical Method of Ludwig Wittgenstein*. *Rhetorical Review*. Vol. 10 No 1. Pp. 31-51.
- [10] Vadera, S., Rodriguez, A., Succar, E. & Wu, J. (2008). Using Wittgenstein's Family Resemblance Principle to Learn Exemplars. *Found Sci*. 13. Pp. 67-74.
- [11] Wittgenstein, L. (1980) *Culture and Value*. Edited by G.H. von Wright, translated by Peter Winch. Chicago, University of Chicago Press, 1980.
- [12] Wittgenstein, L. (1978) *Philosophical Grammar*. Berkeley: University of California Press.
- [13] Wittgenstein, L. (1953) *Philosophical Investigations*. Oxford: Basil Blackford.
- [14] Wittgenstein, L. (2000a) *Tractatus Logico-Philosophicus*. New York: Routledge.
- [15] Wittgenstein, L. (2000b) *Wittgenstein's Nachlass: Bergen Electronic Edition*. Oxford: Oxford University Press.