

Artificial Mental Representation and Creative Pursuit

Kuil Schoneveld

BUILDING artificial intelligence capable of creativity requires that the system hold, manipulate, and integrate mental representations. For this to be possible, we must include a notion of intentionality in the artificial intelligence. The 19th century psychologist Franz Brentano makes an early distinction between the physical realm and the object of the mental: this sets the groundwork for how to interpret the abstract activity of the mind. The cognitive scientist and philosopher Daniel Dennett progresses this conception to describe the *intentional stance* with which we can interpret a system's behavior. I will first situate my discussion of creativity by referring to Paul Thagard's philosophical account of it. In comparison, Margaret Boden's more psychological account lays out a hierarchy of techniques ranging from merely combining mental concepts together in new ways to manipulating the conceptual space itself. With Dennett's conception of mental structures (such as beliefs and desires) as entities that exist in the world, we can better understand Boden's description of creative endeavors. Thus, I intend to argue that mental representation integrations are possible within artificial systems and that this strongly implies human degrees of creativity along Boden's dimensions. I will briefly discuss the English mathematician Alan Turing's vision for machine cognition and respond to Lovelace's objection to it. Following this, I present a vari-

ety of recent works which describe the bridge between neural structures and mental structures¹. I will integrate this work with Boden's ideas of exploration within a conceptual space, as well as explicate how this type of creativity can be done artificially. I describe multiple examples of apparent artificial creativity and offer a modern version of Lovelace's objection and an appropriate rebuttal. I conclude by overviewing the progression from intentional mindedness to its implementation in algorithms, and argue that similar creative activities conducted by humans are capable in machines.

To begin discussing creativity, we need a type of mental entity to manipulate and interpret. To accomplish this, Franz Brentano gives a seminal account of mental representation: he postulates that mental phenomena contrast the physical, because mental phenomena contain *intentional objects* toward which they are directed (Brentano, 481). He refers to this characteristic of mental content being directed at objects as *intentional inexistence* (Brentano, 481). Thus, it follows that our mental concepts and ideas are acquired through, or based upon, sensory perception. However, it is worth noting that the objects of the mental are not required to be existent in the world for them to exist as mental objects. For example, simply conceiving of Pegasus does not mean it exists in the world. Though there is

¹Much of the work I overview is that of Benedek Papp, David Plaut, and Paul Smolensky (a biotechnologist, computational psychologist, and computational neuroscientist respectively).

some controversy surrounding the interpretation of Brentano², I will avoid delving into his metaphysical arguments and instead focus on the relevant characteristic of minds containing mental content. On this matter, Brentano outrightly states the importance of mental representations when he says, “[the] act of presentation forms the foundation of every mental act” (Brentano, 480). Thus, to interpret and discuss mental activity, he is stating that we must consider the mental entities being held by the mind-havers. This applies to any mental activity more complex than an externalized black-box analysis; if we want to understand the mind, Brentano would argue that we must assume the mind-haver is experiencing mental representations. Further, if we realize that since intentional mental content is directed at something in the world (though its physical target is not contingently necessary), we then know that mental content is *about* something. More specifically, the mind can be described as an intentional system. If we understand the mind in this way, we begin to develop an account of mental conception that allows for the manipulation of mental entities. This leads us on a path toward understanding creativity and provides insight on how to replicate it artificially.

If we accept Brentano’s argument that the mind creates representations, we can now decide how to interpret them. Dennett describes a sensible means of interpreting these conceptions through what he calls the *intentional stance*. Firstly, Dennett argues that beliefs and desires are things that exist in the world. Secondly, he argues that to determine beliefs we must assume a kind of strategy that makes predictions based on the beliefs: we can then confirm their existence based on whether the predictions were accurate (Dennett, 556). Rather than rely upon any external criteria to determine whether we hold some belief, he proposes that we could determine beliefs with a sufficiently nuanced understanding of the brain states that consti-

tute them (Dennett, 556). Thus, to argue that mental representations like beliefs are existent phenomena in the world, Dennett suggests that the answer lies within the complex patterns of the brain state that realizes it. To determine whether some system is a *believer*, we must now move to the predictive strategy. If the strategy adequately predicts a system’s behavior, that system should be considered a believer.

Following Brentano’s discussion of intentionality, Dennett says that we can better understand the behavior of certain systems using the *intentional stance*. This can be done relatively straightforwardly. To begin, we “treat the [system] as a rational agent; then figure out what beliefs that agent ought to have, given its place in the world and its purpose” (Dennett, 558). In this way, he abstracts out much of the complexity that adds little to our understanding of the overall system. In other words, it is unnecessarily difficult to try learning what food someone prefers by mapping the interactions and positions of the particles in the person’s brain. Rather, we intuitively assume the intentional stance and verbally ask the person to determine their beliefs and desires. Dennett argues that this is especially helpful when considering what it means to have a mind, the role of the mind in agency, and what mental phenomena really are (Dennett, 561). Thus, if we are to understand the apparent patterns in this higher level of abstraction, it is sensible to assume that the system being investigated has beliefs, desires, and other mental conceptions. A similar kind of detail abstraction resulting in pattern emergence is present in artificial systems striving to imitate thought. It is therefore appropriate to attribute to these systems mental conceptions.

These mental conceptions can then be manipulated in creative ways: broadly speaking, when a system integrates and alters mental conceptions to create products that are original³

²Ex. Tim Crane suggests that Brentano did not think any objects of thought exist, while Gabriel Segal instead claims Brentano posited ‘intentional objects’ to sidestep the problem of thinking about non-existent objects.

³This is deliberately skipping over Boden’s discussion of Psychological vs Historical creativity; all discussion of originality is in the more objective historical sense.

and valuable to society, it is exhibiting creativity. Humans can quickly create new sentences and lines of inquiry seemingly of our own volition. This ability is one component of an artificial superintelligence; a theoretical agent whose intelligence greatly surpasses that of any human. For a system to truly *seem* intelligent, it should not just answer our questions with the knowledge it has been given, rather it should be able to create answers that were not hard-coded and had no direct programmer input. An intelligent and nuanced system might answer our open-ended questions with better questions. However, rather than ask whether we can create this hypothetical system, this essay will focus on whether artificial systems can exhibit creativity as separate from their general superintelligence. With this context in mind, we can see that creative potential is necessary for this superintelligence, though it is not sufficient.

Before we judge the creative potential of a machine, we first need an account of creativity within ourselves. We must determine whether creativity is possible and what specifically constitutes it before any attempts at artificial implementation can be made. Firstly, is ‘pure’ creativity possible? Our logical intuition would say that no effect can come without a cause, and thus no creative venture can be entirely unprompted. This outlines the foundation of Alan Turing’s rebuttal to the Lady Lovelace Objection which will be discussed later in the specific context of artificial intelligence. First, I will focus on Paul Thagard’s overview of creativity to situate it and motivate its artificial implementation, then I will discuss Margaret Boden’s theory of creativity as a bridge toward understanding it as a component of artificially intelligent systems.

Thagard says that creative thinking relies upon images, which he defines as “mental representations based on sensory modalities” (Thagard, 313). He asserts that visual images in particular are responsible for the advancements made in many academic fields. If images are condensed into an understanding of what each means, they can be juxtaposed and integrated

together to build new images. Concepts are Thagard’s idea of either more abstract, non-verbal images or images that have been combined to create a small interacting system. It follows that different concepts can be combined to create new ones which induce societal change, similarly to images but in a less verbally-specific manner. Finally, Thagard offers a description of rules which are more creatively powerful than concepts; he argues that rules are the highest form of human creativity possible. He describes rules as showing the relationship between multiple concepts. Rules can portray a complex interaction between concepts that simply is not possible within a single concept, let alone a single image (Thagard, 318). It is within rules that Thagard says we find the most socially-altering creative endeavours. Concepts were the first step away from the limitations of sensory input of images, and now rules show us the relations between these abstracted concepts. Thagard contrasts the descriptive power of rules with the normative power of what he calls, methods for procedural creativity. These methods are usually a type of rule that attempts to accomplish some goal by prescribing a series of steps. One example he gives is an interpretation of Newton’s calculus; if you want to describe the notion of ‘change’ mathematically, use the calculus (Thagard, 322). We have seen how Thagard describes creative endeavors as both new ways of describing the world, and new ways of accomplishing goals within it. This characteristic of newness motivates why we should attempt to automate creative processes. To make advancements and solve rising problems, we need to think in new ways and generate new rules.

Margaret Boden offers an account of creative behavior that has similarities to that of Thagard but is more clearly defined as a hierarchy of categories. Boden relies upon a similar notion of concepts paralleling Thagard’s. Boden’s concepts are more abstract versions of images that allow them to be understood, manipulated, and integrated together. Similarly, she is relying upon the metrics of newness, value to society, and surprise. Thus, she describes all creative

endeavor as being a matter of degree; all ideas may be creative, but that is not to say they are equally creative (Boden, 2). Two different creative ideas may be of completely different form as well as degree. For example, take the advent of the light bulb and compare it to cubism in painting. It could be said that the light bulb was a reimagined candlestick, but this loses much of the value of the idea of the light bulb. It is the combination of multiple concepts - like the wire, a carbonized filament to produce light, and surrounding glass - which resulted in its invention. Comparatively, the cubism movement was a general rejection of the traditional portrayals of perspective: a common theme in cubism is an attempt to emphasize the two dimensional nature of the canvas being used. Thus, there is a significant difference between these two examples. The first combines previous concepts together to create a new product, while the other comments on the very space within which it is situated. Both types of creativity are analogous to two of Boden's categories of creative thought. The three categories in her hierarchy are combinational, exploratory, and transformational creativity (Boden, 3). Each category has a quaint analogue to action within a child's sandbox, which is a common example of a creative space.

We see that, like Thagard, Boden describes concepts as the entities to be combined in creative processes. However, Boden also describes how these concepts exist within a creative conceptual space. She asserts that a conceptual space is essentially a "structured style of thought...[or] any disciplined way of thinking that is familiar to (and valued by) a certain social group" (Boden, 4). These spaces can differ across metrics like their creative limit or the relation of the space to other spaces. For example, compare the games of tic tac toe to chess. Every possible move of tic tac toe has been played because the potential for creativity is severely limited; there can only be nine moves in a game at most. While chess has many standard openings that have been played countless times, its creative cap is far higher than that of tic tac toe,

because after roughly thirty moves, most games reach a state of originality. We can see some differences among conceptual spaces and their analogues in board games. Generally, we are more interested in designing conceptual spaces with high potentials for creativity. Their results are broader in scope, more difficult to predict, and often more valuable as learning experiences. It is within these richer conceptual spaces that we can more clearly differentiate between Boden's three categories of creativity.

First, combinational creativity involves the synthesis of new concepts from old ones to create something of value (Boden, 3). This is arguably the simplest and least effortful form, since the agent is only required to find a novel way of interlocking two concepts with which it is familiar. For our sandbox example, combinational creativity might include stacking two differently-sized bucketful's on top of one another to create a two-tiered sandcastle tower. One could make additions to the sandcastle to further combine concepts and alter the final product.

Boden defines the second category as exploratory creativity. She describes this form such that it appears more fundamental and impactful in its role in the conceptual space because it involves the discovery of new concepts themselves. Here, a concept has been created that is significantly different from all others in that conceptual space (Boden, 4). Exploratory creativity can therefore allow for more intricate combinational creativity. Thus, exploratory creativity is the causal antecedent because it has drastic effects on the combinations that are possible, where the reverse is not necessarily true. The new concepts may have always been possible, but they had gone unnoticed until the agent assumed an exploratory state of mind (Boden, 5). These explorations may not be very adventurous, and thus their resultant concepts will not be significantly different from the ones already obtained. If we continue with the sandbox analogy, we can imagine a child who digs in the sand using their hands to create a moat for their castle. Instead of just building, the child has

now begun to dig as a means of creation. Similarly, they may no longer restrict themselves to using sand and instead make a drawbridge out of twigs. Thus, they have come up with entirely new ideas to add to their toolbox of possible concepts to employ.

Finally, transformational creativity involves acknowledging the limitations of the conceptual space itself (Boden, 5). Once acknowledged, these limitations can be changed, avoided, or surpassed such that the individual can think in ways that were impossible while within the previous conceptual space (Boden, 6). When the assumptions and limitations of a conceptual space are realized, they themselves can become elements to manipulate for the betterment of the creative product. Transforming the bounds of one’s conceptual space allows for radical changes to be made within it: for example, the new mental conceptions of the transformed space might have been literally impossible to achieve in the old one because of its limitations. As a result, if one expands their space, the range of creative possibilities broadens accordingly. Hence why transformational creativity is the most fundamental category of the three. Like the relationship between exploratory and combinational, transformational creativity exerts a new degree of influence upon its dependent categories. It involves the manipulation of the conceptual space itself, which influences the exploratory category. With a changeable space, the concepts that can be explored and created are also subject to change. And when there are many new concepts to discover, the potential combinations change as a direct result. This chain of unidirectional dependencies defines the hierarchy of Boden’s three categories. Returning to our sandbox analogue, transformational creativity would involve transcending the bounds of the sandbox itself: imagine the child removes one of the four walls of the sandbox to reconceptualize the entire space as a desert for toy soldiers. In this case, the previously assumed boundaries of

the sandbox not only limit, but also blind the agent to any alternative possibilities. Expanding the sandbox implicitly acknowledges its previous limitations and allows the new concepts within to be of a completely different nature.

We have now seen two accounts of creativity. Thagard’s creativity has acted as a situating tool which allows us to understand the value of creativity as a component of our daily practice and lifelong pursuits. We realize that major creative junctions form the backbone of progress throughout history and we have seen his account of how creativity is achieved to create social change. Similarly, Boden describes three categorical tiers of creative thought, each more fundamental and revolutionary than the last. As we move from Thagard’s description to Boden’s more detailed conception, we ask ourselves the question: “how does this idea of creativity lead us closer to its artificial implementation?”. For this, we need even more detail and explicit notation than what Boden offers.

Though Boden argues that her conceptions of creativity are possible in machine minds, Geraint Wiggins describes these conceptions with significantly more detail. He formalizes her categories as mathematical functions to explicate their potential for application across domains. Specifically, Wiggins aims to make a coherent distinction between exploratory and transformational creativity as they relate to formalized conceptual spaces and the objects within them (Wiggins, 1). To discuss his formalization in detail would be inappropriate for the scope of this paper. However, it is worth mentioning that his attempt to give formal precision to Boden’s categories helps to translate them into artificial systems⁴. Wiggins acknowledges that his work is only the beginning, but his progress toward such a specific characterisation sets a strong precedent for the artificial and systematic formalization of Boden’s categories. Furthermore, it spawns a range of new questions to further advance the

⁴Though Wiggins argues that transformational creativity is a form of exploratory, I continue to use Boden’s *three* categories rather than explain the technical distinction since this would require a longer and unnecessary tangential discussion.

project of artificially replicating creativity. Wiggins also mentions some new problems that his work has prompted. For example, he describes the problems of moving toward even more abstract meta-levels of the hierarchy than transformational creativity, or how the interaction of agents might influence their learning (Wiggins, 8). These questions can only benefit the progression toward artificially-conceived creativity.

To implement these notions of creativity in artificial systems, we must first describe how mental representations can be realised computationally. Traditionally, computers algorithmically operate upon data structures of information to accomplish tasks. In this case, the data structures act as the mental representations which are processed by their task-specific algorithms. In contrast, more modern approaches to artificially intelligent systems involve interconnected layers of artificial neurons that mirror the functionality of those in the human brain. In what follows, I'll discuss how initial attempts at machine cognition were hypothesized by Alan Turing using the data structure and algorithm method. Then, I'll show how advancements toward deeply layered neural networks offer more potential for the artificial creativity endeavour. Furthermore, I'll argue that mental representations in artificial neural nets can be responsible for the resultant creativity.

It is important that an artificial system is not simply performing some randomizing function on the input to create what appears to be a new output. What is desired is for the machine to be capable of outputting deliberately-made creations rather than brute-forcing a thousand products for a human to only find a select few which are worthwhile. As such, we hope that the artificial system uses a methodology that mirrors a human's in this important way; by producing a limited number, but highly creative pieces. The consideration for the machine's method is first recognized in the Lady Lovelace objection to

Alan Turing's conception of a thinking machine.

Turing believed that for a machine to be considered intelligent, it must be able to convince a human interpreter that it is also a human. In his conception of an Imitation Game, a computer would engage in faceless conversation with two humans (Turing, 434). One adjudicator attempts to distinguish the human participant from the machine only through text-based messages. Turing's Test implicitly asserts two things; firstly that thought can be reduced to computation and secondly that the act of convincing a human is composed of multiple elements sufficient for affirming the existence of thought (Turing, 435). If through adequately elaborate computation the machine can identify and perform its role in this Imitation Game, we must admit that it is exhibiting some remarkable traits. For instance, the computer must be able to react to certain nuances of language like humour, sarcasm, or insinuation. If the computer manages to imitate a human well enough that it cannot be distinguished from one, Turing says that we must consider it of comparable intelligence.

Some disagree with this claim because they think the Turing Test is not truly gauging the system's intelligence. When considering creativity, Lady Lovelace challenges the possibility for creativity when she says that "a machine can 'never do anything really new'," (Turing, 450). This directly challenges the notion of originality required for creative ideas. This idea of newness also permeates Boden's three categories, though she only mentions it briefly⁵. Lovelace, on the other hand, is arguing that a machine is not capable of creating something original that deviates from simply manipulating its inputs. The essence of her challenge can be quickly parried in two ways. First, Turing argues that it would be difficult to prove a human is capable of original thought that is entirely independent of inspiration (Turing, 450). Thus, Turing argues that

⁵Boden says newness is required for creative thought, but she makes no explicit reference to the degree of *how* new something must be for it to be a significant. We can piece together a potential answer from the rest of her writing, but it is worth mentioning that she seems to address newness as a precise binary (and creates H- and P-creativity).

humans cannot uphold the degree of originality that Lovelace is demanding of the machine, and hence he implies that it is unfair to hold the artificial system to such a standard. Second, it should be noted that advancements in machine learning also address this objection in a more literal sense. By their very nature, machine learning algorithms are designed to deduce rules and patterns that were not explicitly coded (Langley and Simon, 1). These algorithms do often make errors in their initial pattern deductions. However, as the algorithm iterates, the initial error helps to increase the probability that the algorithm ends up identifying a correct pattern (Langley and Simon, 2). This method of propagating the error backward through the network is one such means of supervised learning, where programmers compare generated outputs with ones deemed correct (Plaut, 2). Such learning methods can be applied to many pattern-recognition problems where the artificial systems outperform humans and thus surpass the standards Lovelace might have imagined.

Lady Lovelace's challenge was directed toward a system vastly different than those used today. Modern computers are orders of magnitude faster and more complex than Turing machines, but this fact might not discourage her point about pure originality. If neither of the above answers seem adequate in responding to Lovelace's objection, we might return to Boden's three categories of creativity and examine whether they can be replicated. First, I'll argue that mental representation is possible in artificial systems. More specifically, I'll show that low-level implementations of neural structures can lead to mental structures like those Dennett describes. Second, I'll refer to Boden to argue that these processes of creativity are replicable in artificial systems. Furthermore, I'll describe a modern reconstruction of Lovelace's argument and offer a more precise counterargument to illuminate the future of artificial intelligence development in this domain.

As mentioned previously, where Turing machines may fail, more advanced forms of arti-

ficial intelligence may flourish. This is especially the case in neural networks, where David Plaut shows that representations can exist as distributed entities; encoded by patterns of activity among many neuron units instead of encoded within a single unit (Plaut, 1). Furthermore, he argues that compared to simply scaling up a Turing Machine, these systems have new and more accurate ways of modelling the relationship between brain function and their associated cognitive processes (Plaut, 3). The neurobiological realism of these neural networks means that the mental representations they create are more analogous to the mental representations we humans experience. Moreover, to emphasize the analogous nature of the duality between the artificial and biological systems, Benedek Papp et al. argue that if the structure of an artificial neuron system mirrors that of a biological one, the cognitive capabilities are also mirrored (Papp et al., 7). They explicate how mental information is stored within the weightings of the connections between neurons (Papp et al., 8). This aligns with Plaut's description of distributed representation and reinforces the similarity between artificial and biological networks. Thus, we can see that biologically-inspired artificial neural networks mimic the mental representations with which we are familiar. Next, we'll briefly see a clearer picture of how neural activity manifests into the concepts, beliefs, and knowledge that guide our action.

Dennett argues that the *intentional stance* is far more sensible for predicting an individual's beliefs and actions than, say, an astrological explanation (Dennett, 557). He asserts that intentional beliefs are real patterns of action that are describable only from the intentional stance (Dennett, 562). However, Paul Smolensky offers a detailed account of how we can speak meaningfully as we bridge the gap between neural activity and mental structures. He posits a paradigm of abstraction where neural structures support emergent symbolic layers which conclude with the mental structures of folk psychology, such as beliefs and desires (Smolensky, 97). He defines this as the symbolic paradigm

but argues that this linear type of abstraction is inaccurate. Instead, he describes the more complex sub-symbolic paradigm, which attempts to formalize the processing which occurs in the nervous system (Smolensky, 99). Where the symbolic paradigm supported each layer directly upwards from the neural to the mental, the sub-symbolic has an intermediate connectionist system that precedes the approximate higher-level descriptions. To describe the details of Smolensky’s work would be inappropriate for this paper, but it is worth summarizing by saying that the sub-symbolic level is intended to highlight the dynamic nature of the fundamental neural system. In contrast, the symbolic description is only the resultant *approximation* of the sub-symbolic connectionist system that underlies it. This inaccuracy is due to the attempt to move directly from the neural to the symbolic level; because “connectionist systems are much closer to neural systems than are symbolic systems” (Smolensky, 99). Thus, the addition of the sub-symbolic level enables a higher degree of subtlety and complexity in understanding mental structures. It is worth noting that Smolensky defines connectionist systems similarly to how Plaut and Papp et al. describe neural networks; knowledge is encoded in the patterns of weighting strengths between neurons rather than in symbolic structures (Smolensky, 95). Thus, we’ve described the neural-cognitive gap that we must bridge and seen a few accounts of how to begin this process. Smolensky explicitly acknowledges the need for continued work in this area when he says that “the precise relationship between the neural and sub-symbolic levels is still an open research question” (Smolensky, 99). Nevertheless, we now have a clearer conception of how mental representations can be accurately recreated in artificial systems.

If we recall Boden’s three categories of creativity, we can see how our artificial mental representations can begin to constitute artificial creativity⁶. Boden nicely summarizes how

combinational creativity could be achieved using the two general forms of artificial representation mentioned earlier; data structures and neural network/connectionist methods (Boden, 7). It is important to mention that a system roughly mashing two representations together does not necessarily constitute creativity. Instead, the system must form and evaluate intentional links between representations that are intelligible. To make these links thought-provoking or humorous, the artificial system would require “a database with a richness comparable to ours and methods of link-making comparable in subtlety to ours” (Boden, 8). Though this is a difficult programming task, it is important to notice that Boden is stating that the issue is one of ability, not possibility. Thus, we can see that it is possible to use analogous mental representations in artificial combinational creativity. Furthermore, artificial systems can explore a given creative space by following a certain style and designing new concepts, such as musical compositions or architectural blueprints, which fit within certain parameters that bound the style (Boden, 9). This can also be extended such that a program alters its parametrical rules to allow itself more opportunity for novel creation. One example of this parametrical rule alteration occurs in genetic algorithms, which generate new rules based on previously successful ones (Langley and Simon, 2). An algorithm that successfully reaches its goal is the foundation of the rules of future generations. (Boden, 9). Returning to neural networks, Ahmed Elgammal et al. describe how two networks can be arranged such that they generate and discriminate creative works of art (Elgammal et al., 1). In their work, a variant of generative adversarial networks has been tailored toward exploratory creative pursuit. Their network generates images that controllably deviate from stylistic norms to increase their stylistic ambiguity while remaining artful enough for on-lookers to enjoy as much as human-generated art (Elgammal et al., 5 & 16). Thus, we see that both Boden’s description of genetic algorithms

⁶If we recall and implement Wiggins’ formalizations of Boden’s categories, we can discover more precise examples of Boden’s implementation of creativity both in human endeavor and in artificial systems.

and Elgammal et al.'s neural networks can be used to produce creative results.

Another attempt at artificially-conceived creativity was met with a modern reconstruction of Lovelace's argument. The program designed for literary creativity, called BRUTUS, can create grammatically correct English stories that contain intrigue and mystery (de Sousa, 642). However creative its writing may seem, Ronald de Sousa refutes what he calls the Direct Argument (put forth by Bringsjord and Ferrucci) that challenges BRUTUS' capability for novelty in its stories. The argument says that machines are things that implement algorithms and that nothing can count as creative if it cannot produce anything new (de Sousa, 645). The argument dubiously asserts that algorithms produce nothing new and thus algorithm implementation is not creativity; by extension, no machine can be considered creative (de Sousa, 646). This directly corresponds with Lovelace's objection to the notion of a machine being able to *originate* anything. De Sousa responds to this expanded form of the argument by asking why it is assumed that humans are not implementing (albeit complex) algorithms in our creative endeavors. Furthermore, De Sousa explains that any attempt to give an example of purely inspiration-free creativity undermines this argument because the examples require only a change of perspective

to recognize "that the 'absurd' consequences entailed no contradiction" (De Sousa, 646). Thus, since it seems inherently impossible to identify the facet of human creativity that cannot be formalized, we must accept that our artificial approximations of creativity will eventually be indistinguishable from the human process⁷.

As we tread the line between imitating creativity and truly implementing it, we realize the boundary begins to blur as the imitation becomes more convincing⁸. In this paper, we've seen how mental representations were initially described and how they can be seen as patterns of behaviour. Dennett gives an account of mental conceptions that can be rendered and understood with the intentional stance. Boden discusses how these concepts can be integrated together to constitute creative pursuits of three kinds. These three categories can be formalized and realized in different types of artificial systems. Papp et al., Plaut, and Smolensky have described how to understand the transition from low-level neuronal structures to more abstract mental structures, though this remains an active research area. In brief, creative pursuit is a feat that can be accomplished by artificial systems. Furthermore, the obstacles we face in the advancement of these systems are of an epistemological nature rather than a conceptual one.

AUTHOR BIOGRAPHY

My name is Kuil. I'll soon be completing my undergraduate degree in Knowledge Integration and Philosophy at the University of Waterloo. I've focused my time on studying the mind through the lenses of linguistics, artificial intelligence, and psychology while relying largely upon the toolset of philosophy. This article began as a term paper for my courses on Intelligence and Philosophy of Mind and investigates something central to the human experience: the creation of thought. I am also interested in our attempts to model the mind, both in its cognitive and neural functioning, so AI offers a keystone that bridges between many of my conceptual interests and grounds them in more technical realizations. I find myself motivated to publish in JIRR because of its means of analyzing topics from interdisciplinary perspectives. I hope that in the future I am able to make

⁷As attempts at formalization (like those we saw from Wiggins' earlier) strengthen, we see creative results that match those of a human. See Elgammal et al.'s four experiments in Sections 4.3 to 4.4.

⁸See Bryan Lawson, Cliff Kuang, and Tim Rettig for further discussion of imitation and creativity. Refer to Yasuo Kuniyoshi Chapter 9 for a more detailed look at imitative learning and its emergent cognitive effects or see Section 2.4 of Elgammal et al.'s paper for more details on navigating and overcoming the boundary I only mention here.

some small contribution to this frontier of human knowledge.

WORKS CITED

- Boden, M. (2004). The Creative Mind – Myths and Mechanisms. *Psychological Medicine*, 21(03). doi: 10.1017/s003329170002256x
- Chalmers, David John. *Philosophy of Mind Classical and Contemporary Readings*. Oxford University Press, 2002. (used for Brentano & Dennett)
- Bringsjord, S., & Ferrucci, D. (2013). *Artificial Intelligence and Literary Creativity: Inside the Mind of BRUTUS, a Storytelling Machine*. New York: Psychology Press. Reviewed by Ronald de Sousa, University of Toronto.
- Elgammal, A., Liu, B., Elhoseiny, M., & Mazzone, M. (2017). CAN: Creative Adversarial Networks Generating “Art” by Learning About Styles and Deviating from Style Norms. *International Conference on Computational Creativity (ICCC)*, 1(1), 1-22. doi: 1706.07068v1
- Kuang, C. (2016). Is Imitation the Hidden Key to Creativity?. Retrieved from <https://www.fastcompany.com/1457113/imitation-hidden-key-creativity>
- Kuniyoshi, Y. (2015). Learning from Examples: Imitation Learning and Emerging Cognition. In: Cheng G PhD, editor. *Humanoid Robotics and Neuroscience: Science, Engineering and Society*. Boca Raton (FL): CRC Press/Taylor & Francis; 2015. Chapter 9. Retrieved from <https://www.ncbi.nlm.nih.gov/books/NBK299030/>
- Langley, P., & Simon, H. (1995). Applications of machine learning and rule induction. *Communications of the ACM*, 38(11), pp. 1-17. doi: 10.1145/219717.219768
- Lawson, B. (1999). ‘Fake’ and ‘Real’ Creativity using Computer Aided Design: some lessons from Herman Hertzberger. *Conference Paper*. doi: 10.1145/317561.317591
- Papp, B., Junger, F., Beres, K., & Stanyo, G. (2004). Mental Representation by Electric Neural Nets, 1-12. Retrieved from <http://conf.uni-obuda.hu/HUCI2002/PETRES.PDF>
- Plaut, D. (1999). Connectionist Modeling. Kasdin, A. (Ed.) *Encyclopedia of Psychology*. American Psychological Association. Retrieved from <http://www.cncb.cmu.edu/plaut/papers/pdf/Plautt00chap.conn.pdf>
- Rettig, T. (2017). Why imitation will always be the key to all learning and creativity. Retrieved from <https://medium.com/@rettigtim/why-imitation-will-always-be-the-key-to-all-learning-and-creativity-5ef323625566>
- Smolensky, P. (1987). Connectionist AI, symbolic AI, and the brain. *Artificial Intelligence Review*, 1(2), 95-109. doi: 10.1007/bf00130011
- Thagard, P. (2016). *Brain-mind: From Neurons to Consciousness and Creativity*, Chapter 11, draft 4, University of Waterloo

Wiggins, G. (2001). Towards a more precise characterisation of creativity in AI, *1*, pp. 1-8.
Retrieved from <https://pdfs.semanticscholar.org/f266/a962520f89b77e9e4023a76a3a76eede4687.pdf>