AI & SOCIETY (2024) 39:775–787
https://doi.org/10.1007/s00146-022-01491-2

OPEN FORUM

Do submarines swim? Methodological dualism and anthropomorphizing AlphaGo

Vincent J. Carchidi[1]

Received: 3 January 2022 / Accepted: 28 March 2022 / Published online: 4 June 2022

Abstract

The victories of the Go-playing artificial intelligence (AI) "AlphaGo" against professional player Lee Sedol in 2016 had a profound impact on public and academic perceptions of AI. This event shocked observers, as the ability of a machine to defeat a world champion human in a highly complex game seemed to indicate that a machine had achieved human-like—or more than human—intelligence. But why was AlphaGo so readily anthropomorphized by academic and non-academic audiences alike? Drawing from existing analyses of reactions to and arguments concerning AlphaGo and AI generally, this paper argues that "generative" cognitive science—a school of thought exemplified by the linguistic work of Noam Chomsky—offers two novel contributions to this subject. First, generativism sheds light on an irrational double standard in the study of the human mind in contrast to the study of non-cognitive systems—"methodological dualism"—which, I argue, has been transferred to evaluations of AlphaGo and other AI. Second, by exposing this irrational double standard in perceptions of AI, I employ generativism's more well-known arguments concerning the nature of human intelligence and its scientific study to the evaluation of AI, exposing deficient interpretations widely used in the case of AlphaGo and AI generally.

Keywords Methodological dualism · AlphaGo · Generative cognitive science · Willingness to be puzzled · Artificial intelligence · Creative aspect of language use

1 Introduction

Artificial intelligence (AI) is routinely evaluated against human intellectual abilities. Indeed, demonstrations of AI including IBM's Watson, Apple's SIRI, and Microsoft's Cortana (to name a few) have provided evidence for the success of machine learning, excelling in characteristically human domains such as natural language processing and voice responsiveness (De Spiegeleir et al. 2017, p. 37). Perhaps the most striking success of an AI exceeding human abilities in a specific domain is DeepMind's "AlphaGo" in its defeat of professional Go player Lee Sedol in 2016. AlphaGo's victories against Lee shocked the Go-playing community and inspired soul searching on the relationship between AI-enabled machines and human beings (Dong 2016).

One phenomenon sticks out in the discourse following this event: AlphaGo was subject to a kind of anthropomorphizing in which human qualities were associated with its ability to excel at playing Go. Curran, Sun, and Hong explore this issue directly, observing that AlphaGo's victories "resulted not just in a reassessment of AI, but also prompted introspection about what it means to be human" (Curran et al. 2020, p. 727). These authors explore the question of how the case of AlphaGo elicited discourse in the Chinese and American press on the meanings of "human" and "machine" through a content analysis, identifying the presence of the frames "human" and "threat" in reference to this event (Curran et al. 2020, p. 729).

This paper seeks to address the anthropomorphizing of AlphaGo, and AI generally, from the perspective of "generative" cognitive science (or "generativism"), a school of thought exemplified by the linguistic work of Noam Chomsky. Generativists like Chomsky (1994) have observed for decades that human beings routinely employ an irrational double standard in the study of their cognitive capacities (e.g., language, moral judgment) in contrast to the study of non-cognitive capacities (e.g., the circulatory system). This

✉ Vincent J. Carchidi
carchidi.vince@gmail.com; vcarchi1@villanova.edu

[1] Political Science Department, Villanova University, Villanova, PA, USA

pervasive and irrational bias typically operates implicitly in assumptions regarding human intelligence, mandating that the study of the human mind be subject to essentially arbitrary standards, thereby diminishing individuals' abilities to grasp the distinctiveness of the human mind. This "methodological dualism," I argue, has been implicitly transferred to evaluations of AI systems, with AlphaGo being a case-in-point, negatively impacting conceptions of AI generally. To put it bluntly: the existentialism that results from an implicit transfer of methodologically dualist assumptions about human intelligence to evaluations of artificial intelligence is irrational. Existential *conclusions* about AI are *inferences* made from these assumptions (i.e., the premises).

An important caveat is that, while adopted, the famous "poverty of the stimulus" argument—which holds that humans possess innate, sophisticated cognitive faculties relative to the information they are exposed to during development—is *not* the primary argument used here, focusing instead on *prior* steps in reasoning.

I thus pose the following research questions, whose names I will use to reference them throughout the essay:

1. "Anthropomorphizing AI": What are the assumptions regarding human intelligence that allow individuals of diverse backgrounds to anthropomorphize AlphaGo?
2. "Science of Intelligence": What are the assumptions which ought to be adopted in studying both human and artificial intelligence?

The first question is *descriptive*, aiming to characterize the reasoning about human intelligence that individuals frequently employ to anthropomorphize AI. The second is *prescriptive*, seeking to characterize rigorous means by which human and artificial intelligence are conceptualized. These questions are logically related, with the second dependent upon the first. They are each ripe for engagement by generativism.

To the skeptical reader, I want to be clear about the "prescriptive" nature of the second, "Science of Intelligence" question: generative linguistics, particularly the work of Chomsky himself, has experienced many twists and turns throughout the years. But there is a permanence to some of the most fundamental insights provided by generativism which prove powerful in grappling with the study of intelligence. The case made here is not that one, specific model of the mind is the final word on the matter, but rather that these fundamental insights allow scholars to identify distinctive properties of human intelligence that stand in stark contrast to existing AI systems. Particular attention is paid to the "creative aspect of language use"—a property of human intelligence which cannot be *explained* in any meaningful sense, carrying powerful implications for the study of natural and artificial intelligences alike.

Ultimately, my hope is to inject a new way of thinking about AI and its relationship to human intelligence with old ideas that hold significance across diverse schools of thought. Although his work and terminology are cited throughout, I am less concerned that the articulation of these ideas follows the exact path taken by Chomsky and more interested in helping the field of AI reach out into a new direction. If the reader disagrees with, say, the poverty of the stimulus argument, this should not detract from their interest in the creative aspect of language use and the lessons it carries for the evaluation of AI systems.

I begin with an introduction to the case of AlphaGo and the public and scholarly commentary that surrounds it, proceeding to describe the methods involved with answering our two questions. Then, I delve into the first question of "Anthropomorphizing AI" by explicating methodological dualism and connecting it to perceptions of AI by characterizing the steps in reasoning which lead individuals from assumptions about human intelligence to evaluations of AlphaGo. This naturally leads to the second question concerning a "Science of Intelligence." Because generativists view methodological dualism as irrational, the need for an *alternative* characterization, through generativism, of the steps in reasoning needed to understand intelligence are explored. I end with a discussion of the analyses' results.

2 AlphaGo and intelligence

Go is an ancient Chinese game that is typically played on a 19 × 19 grid with a total of 361 points on which black and white stones are placed. The goal of each player is to acquire enough points on the board to surround the opponent's stones, with the winner holding the most territory. The sheer number of possible moves leads to a great deal of complexity and need for strategizing to achieve a winning posture.

In March 2016, AlphaGo faced off against the professional Go player Lee Sedol (who, now retired, held 18 international titles) over the course of five games. Go experts had predicted prior to the showdown that Lee, given the practically infinite number of possible moves, would emerge victorious. Shockingly, AlphaGo defeated Lee in four out of these five games. What followed was a sense of existential crisis in the Go community not only that a program had defeated a professional human, but that a Western technology had achieved a winning posture in an Eastern game. AlphaGo, shortly thereafter, defeated 3-time world champion Ke Jie (Dong 2016).

It is difficult to overstate the impact these events had on not only the Go community, but also on public discourse surrounding the differences (and similarities) between humans and AI-enabled machines and academic thought on

AI's nature and progress. Curran, Sun, and Hong conducted a content analysis of the Chinese and American press on how AlphaGo's showdown against Lee was framed. They find that there was widespread "anthropomorphizing" of AlphaGo as well as a varied use of "threat" and "non-threat" frames applied by commentators to capture this event's impact on both the definition of "human" and "machine" and the significance of humans in domains increasingly dominated by machines. Bory (2019), furthermore, observes how AlphaGo's victories can be placed in a long-running narrative among participants in machine-human confrontations and their observers about the ability of machines to acquire human-like abilities or qualities.

Academically, one finds in *Nature* the confidently-titled "Mastering the game of Go without human knowledge"—an article written by AlphaGo's programmers on later iterations of their program, notably "AlphaGo Zero" (Silver et al. 2017). AlphaGo Zero's programmers not only claim that their program reached a "superhuman level" in the game Go, but they also say, without qualification, that their techniques could be utilized "even in the most challenging of domains" (Silver et al. 2017, p. 358). Essentially, they claim that AlphaGo Zero's programming could be applied to other domains ripe for AI applications with the same potential for superhuman intelligence.

Claims such as these have inspired counter-arguments concerning the nature of AI and the possible trajectories for AI systems designed in the lineage of AlphaGo Zero. Jebari and Lundborg (2021), for example, take aim at the claim that AIs of this sort can become 'general agents' that can use their context-specific intelligence for cross-context goals. Svensson (2021) approaches the concept of "intelligence" by arguing that it is dependent upon an organic body with features other than narrow, calculative abilities, thereby making "artificial intelligence" an oxymoron.

These and related works can be categorized according to which of our research questions they could answer. Those focused on the public commentary and discourse surrounding AlphaGo and the more general anthropomorphizing of AI are associated with the "Anthropomorphizing AI" question. Scholarly work, on the other hand, can be categorized with the "Science of Intelligence" question of which assumptions are (or could be) adequate in the study of human or artificial intelligence.

3 Methods

The "Anthropomorphizing AI" question is viewed by the generativist as a question about *perception*—it is best answered by looking beneath social-scientific phenomena and into the reasoning used to evaluate the nature of AI systems. This does *not* indicate that social-scientific elements should be dismissed. Indeed, content analyses concerning the frames and concepts used to discuss AI and AlphaGo in popular and academic imaginations are consistent with this approach. An answer to this question can either be a discovery that discourse on AI and human intelligence is so diffuse that it lacks common intellectual ground, or it can come in the form of a characterization of the steps in reasoning that lead an individual from specific assumptions to attaching human properties to AI systems. AlphaGo may be understood as a data-rich case study in that its cross-cultural and cross-disciplinary impact provides insights for the study of AI generally.

The "Science of Intelligence" question is a matter of philosophical *justification* for one's conceptions of intelligence. Indeed, this question is *directly* connected to the first in the following way: while the generativist aims to provide a stepwise characterization of the reasoning involved in anthropomorphizing AI, this characterization through methodological dualism is considered irrational. Thus, if one wishes to construct a philosophically and scientifically viable account of AI, an alternative chain of reasoning is needed. But this alternative account cannot be constructed until one knows with confidence that relevant assumptions about human intelligence have been identified.

The overall analysis is decidedly interdisciplinary, applying generative concepts, theoretical frameworks, and methodologies to AI. The notion that individuals frequently adopt an irrational methodological dualism in the study of the human mind, and that such a dualism should be exposed and eradicated in the context of AI, is an interdisciplinary effort. Because this draws from existing empirical analyses—which includes content analyses of public commentary on AlphaGo and of general AI cultural trends, media narratives of AlphaGo and other AI, and basic reporting on AlphaGo—it is an application of conceptual and theoretical frameworks to existing data to provide a renewed understanding of the relationship between AI and humanity.

The analysis pertaining to the second question is effectively an extension of the first: it remedies a methodological dualism used to perceive human-like AI by explicating the steps needed to achieve a "methodological naturalism" in the study of intelligence. This revised understanding exposes the deficiency of widely held conceptions of current AI systems and the human intelligence their design draws inspiration from.

4 Answering question 1: perceptions of AI

Turkle wrote that computers possess two natures, one being their analytical functions, and the other "as an evocative object, an object that fascinates, disturbs equanimity, and precipitates thought" (Turkle 2005, p. 19).

Curran, Sun, and Hong presciently "suggest that the AlphaGo match itself can be considered as an "evocative event" …The match prompted an explosion of concern of AI in terms not only of its material implications… but also in terms of what was often framed as an unwelcome encroachment on humanness" (Curran et al. 2020, p. 728). This is an effective way of describing the sense of mystique and existential anxiety or curiosity drawn out by AlphaGo. But why was AlphaGo's match with Lee an evocative event? The authors further suggest, citing the "dialogic process through which Twitter users interact with [chatbot Tay]," that "AlphaGo and its agency and humanness are socially constructed through its interaction with human interlocutors and through its coverage in the American and Chinese press" (Curran et al. 2020, p. 728).

The authors' subsequent content analysis found that a "non-threat frame" was used more frequently in reference to AlphaGo in the Chinese press than the American press. This may find its roots in cultural divergences concerning relative openness to AI. Either way, however, both the Chinese and American press found themselves at times grappling with the boundaries of "human" and "machine" in the discourse on AlphaGo's victory (Curran et al. 2020, pp. 731–732). It is worth noting that, while one may think China's unique cultural history with Go may seem at first to be responsible for its anthropomorphizing of AlphaGo, the surprising finding in this analysis was not that the Chinese and American press differed in covering the *topic* of the boundaries of "human" and "machine," but rather in the relative prevalence of "threat" and "non-threat" frames applied to AlphaGo alongside this broader discussion.

Natale and Ballatore take a more general approach to the anthropomorphizing of AI through a content analysis. They find that the "AI myth"—"the ensemble of beliefs about digital computer [*sic*] as thinking machines" (Natale and Ballatore 2020, p. 4)—was constructed during the 1950–1975 period in America and Britain. This belief about machines that can think has persisted in popular and academic cultures *despite* both scholarly distinctions between "weak AI" (i.e., specialized applications which comprise most examples of AI) and "strong AI" (i.e., human-like AI) *and* the "AI winters" that followed this period (Natale and Ballatore 2020, p. 7, 13). In this sense, the anthropomorphizing of AI is a technological myth.

Together, these content analyses provide partial answers to our "Anthropomorphizing AI" question concerning the assumptions entering AI anthropomorphizing. However, while each study adopts social-scientific levels of analysis, generativism offers a deeper, complementary approach to the question of why individuals engage in reasoning that so readily leads to human characteristics being associated with demonstrations of AI systems like AlphaGo. How exactly can this chain of reasoning be derived from reactions to AI demonstrations?

The first step is to recognize that "humanity will probably increasingly be forced to define itself in relation to artificial intelligence, and AlphaGo presents us with a particularly salient opportunity to consider the implications of this process of redefinition" (Curran et al. 2020, p. 728). Nuances in AI perceptions aside (e.g., whether "smarter" AI is something to be feared or hoped for), there *is* a pervasive sense of inevitability in carving out a space for humanity in the face of AI advancements. This indicates that human–machine confrontations—especially ones the AIs win—elicit a sense of existentialism. This is an *inference* that individuals make. Our first, "Anthropomorphizing AI" research question is effectively seeking to identify the *premises* from which an anthropomorphized conclusion is reached.

So, we turn to the conceptual or theoretical premises necessary for this conclusion to arise. Whichever position one takes on AlphaGo will involve adopting either implicit or explicit assumptions about the nature of intelligence. Given that any individual's foremost experience with intelligence is *human* intelligence, it is likely that such assumptions will reflect assumptions concerning the nature of human intelligence which have been transferred to AlphaGo. Generativism, through its identification of a methodological dualism, deals explicitly with these assumptions.

Consider, then, the following: if methodological dualism is operative in the minds of individuals evaluating human intelligence (i.e., the premise of the inference), then they will fail to identify the characteristics of the human mind that make its intellectual capacities distinctive. As a result, when individuals approach the topic of AI, they implicitly use their unrefined notions of human intelligence in their evaluations of AI systems. The culmination of this chain of reasoning is an anthropomorphizing of AlphaGo. The individual who adopts methodological dualism is engaged in an irrational endeavor, misunderstanding *both* natural and artificial intelligences. Put simply, generativists suggest that individuals are *not* reliable witnesses to their own intellects.

To fully characterize the reasoning associated with the anthropomorphizing of AlphaGo, methodological dualism and the generative tradition it occupies must be explicated.

4.1 Methodological dualism and the willingness to be puzzled

Generativism draws from a rich tradition in the cognitive sciences and philosophy concerning not only specific theories of the human mind's capacities (e.g., language, vision, auditory, moral judgment), but also of the assumptions and dispositions underpinning a scientific discipline that purports to study higher-order cognitive abilities. In its most fundamental form, generativism sheds light on the assumptions and standards implicitly used to scientifically study human intelligence *in contrast* to those used in the natural sciences.

The double standard Chomsky (1994) identifies concerning the study of the body and the study of cognition—"methodological dualism"—draws inspiration from the development of the natural sciences, particularly physics, as a model for how sciences of the human mind should be constructed. Methodological dualism is embedded in a larger narrative about the assumptions and dispositions that evolved over the course of centuries in fields like physics, illustrating an undeniable trend *away* from commonsense positions towards counter-intuitive notions of greater explanatory depth.

This begins with a simple observation: physics did not develop into a mature scientific discipline until figures such as Galileo and Newton did away with commonsense ideas about how the world works. One such idea was that physical reality could be explained in "mechanical" terms which comport with intuitive notions of how objects interact with one another. A centuries-long, intellectually painful process led to the development of a physics which accepts that theories are designed to *explain* phenomena like gravity and motion in an intelligible manner, but they are not meant to gain direct insight into such phenomena, leading to theories that are counter-intuitive and bizarre yet widely respected (Chomsky 2009a). "In brief, if we are biological organisms, not angels, much of what we seek to understand might lie beyond our cognitive limits" (Chomsky 2009a, p. 184).

The study of human cognitive capacities has not, according to Chomsky, reached these scientific heights. Disciplines ranging from the social sciences to the cognitive and neurosciences often implicitly embrace a methodological dualism, as opposed to a methodological *naturalism* that accepts the direct inaccessibility of the human mind and embraces counter-intuitive explanations. This dualism in the study of human beings is summed up as "the view that we must abandon scientific rationality when we study humans 'above the neck' (metaphorically speaking), becoming mystics in this unique domain, imposing arbitrary stipulations and a priori demands of a sort that would never be contemplated in the sciences…" (Chomsky 1994, p. 182).

Concretely, methodological dualism is evidenced throughout the study of human intellectual capacities in several ways. In linguistics, it may surface as an insistence that languages are what can be heard when individuals are speaking to one another (as opposed to being a property of the mind/brain); or that languages are fundamentally "learned" (rather than biologically acquired); or that words must possess direct associations with external objects. Whatever the specific instance, such examples are tied together by generativism as common forms of resistance to an assumption taken for granted when studying *other* biological life forms: that cognitive capacities are fundamentally genetic endowments specific to certain species, whose variation therein (e.g., different languages) is evidence only that shared cognitive capacities can be expressed in various, though limited, ways.

Imagine that we insisted a dog's ability to smell in a comparatively richer way than humans is learned. Perhaps, we further insist, humans could learn this ability if only we subjected them to abundant arrays of scents in controlled settings. This is an absurd suggestion. Yet, methodological dualism compels individuals to make equally absurd claims about the human mind, suggesting that all manner of things—from language, to morality, even musical taste—must be matters of learning and environmental associations, or some wishy-washy brain-stimulus interaction, or that the theory of one cognitive faculty must be perfectly consistent with the theory of another.

It is in this context that Chomsky urges scholars to adopt a 'willingness to be puzzled' about the study of human intellectual capacities as a means of moving beyond commonsense beliefs (Chomsky 2013, p. 38). In linguistics, he often invokes a Martian scientist studying human languages, devoid of methodological dualism, who "[concludes] that there is one human language with minor variants" (Chomsky 1995, p. 13)—an indication that our commonsense view of languages like Japanese and English as being wildly different is an unscientific one yet has made its way into scientific approaches to linguistics. On this view—which can be characterized as "methodological naturalism"—because language is a product of an individual's genetic endowment, English, Japanese, Mandarin, Swahili, and the like are not languages but rather potentials afforded by a discrete cognitive system. Lay observers could be forgiven for believing language is what they hear when someone speaks, but the scientist who accepts this position may have fallen prey to methodological dualism.[1]

4.2 Methodological dualism and anthropomorphizing AlphaGo

The key lesson here is that methodological dualism *diminishes* the sophistication and distinctiveness of the human mind. If one uses human intelligence as a baseline for intelligence generally, then they are likely to transfer their methodologically dualist assumptions in the study of the human

[1] To temper the tone of this argument, it is worth noting that there are more than the two options of "Chomskyan generativism" and "behaviorism" in the study of the mind. One need not be a Chomskyan linguist, for example, to study language simply because they are not a behaviorist. I urge the reader to pay careful attention to the *underlying* mindset and reasoning as the argument progresses. I also wish to thank an anonymous reviewer for pointing out the need for this moderation.

mind to the study of *artificially* intelligent systems. Thus, it is a short distance between this inability to identify human-like intelligence and an assumption that artificially intelligent systems will be similarly identifiable. A more formal illustration of this reasoning is below:

1. Unwillingness to be puzzled about human intelligence: Individuals are frequently unwilling to be puzzled about the nature of the human mind in a manner comparable to the puzzled dispositions which presuppose the formations of the natural sciences.
2. Methodological dualism: The study of the human mind is frequently subject to a different standard of inquiry than is taken for granted in the study of non-cognitive systems. This mind-specific standard is pervasive, often implicitly embedded in theories and arguments, and irrational.
3. Underappreciation of human cognition: Because the study of the human mind is subject to an irrational standard of inquiry, theories of human cognition in areas such as linguistics, moral psychology, visual cognition, and the like are often unable to identify distinctive characteristics of human intelligence and the specificity of their underlying, innate structures.
4. Inability to Identify human-like intelligence: Because human cognition is frequently underappreciated, individuals who adopt methodological dualism will lack the scientific mindset and technical vocabulary needed to identify examples of human-like intelligence.
5. Anthropomorphizing AI: Because of deficient conceptions of human cognition, individuals frequently misidentify AI systems as "intelligent" in ways comparable to human intelligence, irrationally relating the two.
6. Seeing ourselves in AlphaGo: Finally, because AI is frequently anthropomorphized in accordance with a methodological dualism, demonstrations such as AlphaGo's victories over Lee Sedol and Ke Jie are misinterpreted both as examples of human-like strategizing *and* as capacities which can be refined and developed into abilities matching or exceeding the human intellect.

To be sure, methodologically dualist assumptions may not be the *only* operative assumptions regarding AI. There are cultural factors that make AI seem threatening or non-threatening depending on the society in question. Furthermore, the commercialization of AI provides private firms with an incentive to inflate AI's current nature and trajectory by comparing AI systems to humanity, as Bory (2019) contends. These examples are not, however, inconsistent with methodological dualism. The assumptions which come with methodological dualism are *fundamental* to conceptions of human intelligence, thereby shaping the character of individuals' perceptions of AI. Perceiving a "threat" or "non-threat" from anthropomorphized AI are different ways of utilizing the same baseline assumptions about intelligence. Even marketing campaigns concerning anthropomorphized AI could only succeed if individuals are prepared to see themselves in AI demonstrations, with the "AI myth" playing off this existing perceptual bias.

There are two trends, however, which indicate that methodological dualism can be either challenged or reaffirmed. First, the criteria needed to be considered "intelligent" shift in response to machines exceeding humans' abilities in certain domains, such as Deep Blue beating Garry Kasparov in Chess (Curran et al. 2020, p. 727). This would only occur if prior conceptions of human intelligence were deficient, prompting existential anxiety (or curiosity). It indicates a willingness to cast doubt on prior conceptions, though not a guarantee that methodological dualism will be exposed. Second, some who have gone up against AIs have reevaluated in the *opposite* direction, believing that AIs are *too* human-like. Kasparov was confounded by Deep Blue's move 36 in game 2 precisely because the move felt as if it could only be made by a human (Bory 2019, p. 637). Kasparov's intuitive sense of what it is like to play Chess as a human is emblematic of a common-sense view of the human mind.

4.3 Question 1 (anthropomorphizing AI) verdict

Through the chain of reasoning above, generativism offers an answer to our "Anthropomorphizing AI" question: the assumptions concerning the nature of human intelligence which allow AI to be anthropomorphized are bound up in a broader, irrational methodological dualism in the study of the human mind in contrast to the study of non-cognitive systems which have been transferred to AlphaGo. Methodologically dualist assumptions are not guaranteed to be operative in perceptions of AI, nor are they guaranteed to go unchallenged, but they are typically implicit. Existentialism induced by observations of AI systems in action is a conclusion inferred from an irrational set of assumptions about human intelligence.

What this answer tells us is that there are widespread misperceptions regarding the nature of human or artificial intelligence. However, while generativism holds that such perceptions are grounded in irrational approaches to the study of the human mind, there *are* philosophical arguments centering on just this problem—the subject of our "Science of Intelligence" question—to which we now turn.

5 Answering question 2: justifications for perceptions of AI

The most prominent claims concerning AlphaGo come from its programmers (Silver et al. 2017). Writing on AlphaGo Zero, they made bold and often unqualified claims regarding their program's ability to acquire and create Go knowledge.

They note that AlphaGo Zero was "trained solely by self-play reinforcement learning, starting from random play, without any supervision or use of human data" (Silver et al. 2017, p. 354). Their boldest claim comes within this context: they suggest that AlphaGo Zero began "*tabula rasa*" and proceeded to "rediscover [human] Go knowledge, as well as novel strategies that provide new insights into the oldest of games" (Silver et al. 2017, p. 358). They proclaim that "it is possible to train to superhuman level, without human examples or guidance, given no knowledge of the domain beyond basic rules" (Silver et al. 2017, p. 358). In simple terms, AlphaGo Zero is said to have played games against itself, starting as a blank slate, and rapidly learned both human and novel strategies by which to win.

This hype surrounding AlphaGo Zero's "superhuman" level has elicited suspicion. Jebari and Lundborg, discussing "general" agency in humans that can direct behavior across multiple contexts, make a distinction between desires that are conducive to general agency and those that are not. Using this distinction, they argue AI-enabled machines like AlphaGo Zero cannot attain general agency because they can never escape beyond the bounds of their initial, context-specific desires (Jebari and Lundborg 2021, pp. 810–811). Svensson, furthermore, argues that AIs like AlphaGo are often perceived as "intelligent" when it comes to *quantification*, yet lack other relevant dimensions such as approaching topics without large, or complete, sets of data, as humans routinely do (Svensson 2021, pp. 4–6).

While these analyses are valuable, they adopt certain assumptions about AlphaGo that cloud fundamental issues in approaching the concept of intelligence. Jebari and Lundborg, for example, accept Silver et al.'s claim that "while the predecessor of AlphaZero (AlphaGo), had some pre-programmed beliefs, AlphaZero learned how to play chess purely through trial and error" (Jebari and Lundborg 2021, p. 811). Svensson, furthermore, comes to his conclusion about intelligence in part with the premise that most "AI success stories…often revolve around specific and narrow calculation tasks…But once the use of AI is expanded to outside of the realm of narrow rule-based contexts…it becomes more problematic" (Svensson 2021, p. 4). These authors assume that AlphaGo Zero was not preprogrammed with beliefs about Go and that a major distinction between human and artificial intelligence is "narrow" intelligence (rule-based, context-specific activities) and "general" intelligence (dynamic, cross-context behavior).

Consider, first, the assumption that AlphaGo Zero was a blank slate. The critical reader will notice, as Marcus (2018, pp. 6–9) points out, that AlphaGo Zero's programmers contradicted themselves. "Tabula rasa" indicates that AlphaGo Zero started with nothing except the most basic, neutrally designed elements of deep learning systems and then proceeded to use a "pure reinforcement learning approach" to "learn" old and new Go strategies. However, they also say that AlphaGo Zero was "given no knowledge of the domain *beyond basic rules*" (emphasis mine). What exactly are these "basic rules?" Consider a partial list:

> Rule 1: Monte Carlo tree search: A search technique that statistically tests moves and countermoves which is *commonly* programmed into computer games.
> Rule 2: Translation invariance: The *layers* of nodes (common to deep learning systems) are placed in such a precise way as to allow patterns on the board to be consistently recognized by the system.
> Rule 3: Representations and algorithms of the Go Board: *Algorithms* for recognizing "symmetries" on the board like reflections and rotations as well as *representations* of the board structure and the specific rules of the game (Marcus 2018, pp. 7–8).

In sum: "To the extent that [AlphaGo Zero] does build in innate algorithms, knowledge and representations, its constructs are *more* specific to Go and to game playing than any human might plausibly possess" (Marcus 2018, p. 9). AlphaGo systems thus lack the flexibility of human cognition and excel in a far more constrained domain in part by combining game playing and translation invariance programming. These systems are not blank slates but rather endowed with Go-specific knowledge in advance of self-play trial and error.

Marcus' analysis echoes, as Childers et al. (2021, pp. 1–3) observe generally, the intellectual debates in linguistics in the 1950s and 1960s concerning the innate structure of the human mind. Marcus (2018, p. 2) observes that the claim that AlphaGo Zero started tabula rasa is emblematic of an *empiricist* approach to the study of the mind, while his critical analysis is representative of a *rationalist* approach. Marcus has thus drawn from the generative tradition in relation to AlphaGo. Just as the human mind, according to linguists like Chomsky, requires a rich, innate, language-specific faculty to acquire any natural language, AI systems themselves require comparable innate structures to achieve human- or animal-like intelligence. The importance of innate structure as an enabler of intelligence is, however, strangely sidestepped by AlphaGo Zero's programmers who instead choose to inflate the importance of pure reinforcement learning. This resistance to promoting AlphaGo Zero's innate structure does not make much sense *unless* a methodological dualism is operative in either the programmers' minds or the minds of their anticipated readers.

Marcus' analysis has, however, drawn from the aforementioned "poverty of the stimulus" argument in linguistics in his claim that AlphaGo Zero's innate, Go-specific structure is responsible for the program's success. He does not, surprisingly, draw attention to methodological dualism's possible influence in the programmers' decision to inflate the importance of pure reinforcement learning as opposed to AlphaGo Zero's innate structure. The poverty of the stimulus argument, while relevant to AI generally, is *preceded* by a methodological naturalism that is sorely missing in this context.

Even Svensson's comments on narrow and general AI fall into this category. While this distinction is helpful in conceptualizing the ways in which humans *use* AI, it clouds a deeper issue. If we assume that narrow AI is context-specific, then its innate endowment must, similarly, be designed for operating in just such a context according to specific rules. This is, essentially, what Marcus is saying by applying generative linguistics to the case of AlphaGo—innate structure is what enables intelligence in humans *or* machines. But if this is narrow AI, then how would general AI be constructed?

Presumably, general AI would still require innate structures for particular domains of behavior—just as humans have a language faculty, visual system, auditory system, possibly even a moral faculty. But these structures would have to be *less* specific to the contexts of their use than AlphaGo Zero's endowment is to Go, or directed by some "higher" faculty, or else they would be hopelessly confined to operating in extremely rigid, rule-based scenarios. Yet, if AlphaGo Zero's innate structure is made less specific to Go, then it may lose its ability to play Go at the level of professional human players. Naturally, then, we may ask how *humans* simultaneously possess rich, innate structures for various cognitive abilities yet also possess a dynamic intelligence capable of using their cognitive systems interactively in the service of novel goals and ideas. The poverty of the stimulus argument does not help us here, as this merely tells us what internal properties or components of the mind *enable* a broad range of intellectual abilities. But it tells us nothing about how these properties are *used*.

The puzzle, then, is this: how can an AI's constituent components be made less specific to the context of its use—or subordinate to a higher component—if the aim is to make such AI general in nature? By turning to this problem, we can more adequately address our "Science of Intelligence" research question concerning the adequacy of the assumptions used to study human—and thus artificial—intelligence.

6 A willingness to be puzzled about ourselves

If one watches the match between Lee Sedol and AlphaGo, it certainly seems—particularly on move 37—that something intelligent and innovative had been done by a machine. But it also seems like the Sun revolves around the Earth. Commonsense perceptions of how the world works are extraordinarily difficult to dislodge from academic studies of the world, with physics taking centuries to overcome the urge to construct theories that comport with intuitive perceptions. The study of AI stands in need of just such a process of self-reflection.

Generativism aids this endeavor by encouraging scholars to adopt a willingness to be puzzled about the human mind. One result of this disposition is uncovering a methodological dualism operative in perceptions of both human and artificial intelligence. But what, specifically, about the human mind is distinctive and apparently lost to methodologically dualist assumptions? While the poverty of the stimulus (POS) argument is relevant to answering this question, constructing an alternative conception of the human mind means homing in something less known: the "creative aspect of language use" (CALU).

POS, as described below, is about the innate systems of the human mind which enable a broad range of intellectual abilities. But, in trying to understand how an intelligence can go from "narrow" to "general," we run into the thorny issue of just how these abilities are used and interact with one another. CALU is a feature of human language use that is, according to generativists, inexplicable through scientific means, with only descriptions of this ability available to us. Understanding CALU—and its relationship to intelligence—depends on a prior understanding of POS, to which we now turn.

6.1 Poverty of the stimulus

Every human child, absent serious developmental disabilities, can acquire any natural language to which they are exposed. By a young age, a child can generate a rich linguistic output, using a rapidly growing vocabulary to construct sentences which are novel to their personal history and appropriate to their circumstances. The examples of a given language—the data—that an infant has been exposed to are, however, highly *limited* relative to the child's rich linguistic output. The linguistic data the child has received in its development is not only finite but often *deficient* (i.e., others may use grammatically incorrect sentences to communicate). This is a well-established developmental trajectory for infants, but the puzzle is this: how does the infant readily and rapidly acquire a rich grasp on any given language, thereby allowing it to generate an infinite number of linguistic utterances, if it has only been exposed to finite data that are often faulty?

POS starts from the observation that a child can acquire any natural language to which it is exposed and use its limited data to generalize well beyond the examples it has encountered in ways that are both productive and appropriate. The stimulus—the linguistic data—is impoverished, *relative* to the child's output. Proponents of POS argue that this informational gulf between the child and their environment strongly suggests that the child must be equipped with a genetic program which gives rise to a faculty of the brain that houses principled knowledge of language. This knowledge "grows" within the child's mind during normal biological development and, critically, enables them to acquire any natural language which depends on such knowledge.

This argument pinpoints some fine-grained features of language acquisition which are not obvious. Ordinary environments for an infant on the journey of linguistic development will present series of stimuli including interactions between people, animals, objects, sights, sounds, and feelings. A generativist will ask, when confronted with this "blooming, buzzing confusion," how on Earth does the infant know which stimulus is linguistic and which is not? Put another way, how does the infant appropriately categorize words in their environment as part of one discrete ability (language) and not another? The answer provided by the generativist is that the child can identify linguistic data in a world of stimuli because its innate language faculty is encoded with the knowledge requisite for such identification (Berwick et al. 2011; Lasnik and Lidz 2017, pp. 1–5).

How could a child possess unlearned "knowledge" of language? Consider the following perspective on POS: for the past several decades, linguists have gone through painstaking efforts to uncover the principles which underlie the world's natural languages. Despite these efforts, however, they have yet to approach a full characterization of such principles. Children, in contrast, can attain just this principled knowledge without any empirical studies nor expert knowledge of natural languages and they do so on a largely unconscious level by a young age (Jackendoff 2008, p. 26). It would be strange, against this backdrop, to believe human beings do *not* develop with a rich, innate linguistic structure.

This returns us to one of generativism's fundamental insights about human cognition: studying the mind should be done with the same assumptions we use to study human physiology. Do we assume that the immune system—which is an abstraction of various bodily processes (Collins 2004, p. 508)—is the product of some kind of "learning?" Do children need vast amounts of data on immunology to acquire this system? Answering these questions in the affirmative would be absurd, yet methodological dualism pushes us to answer in just this way when human intelligence is on the line. There *is* a kind of learning going on in language acquisition, but the question is *what* is being learned (Mikhail 2011, p. 104).

Once we are willing to be puzzled about human cognition, its richness becomes apparent and its origin in human biology the most plausible explanation for socially pervasive capacities. Without complex, innate structures for language acquisition, morality acquisition, even visual and musical cognition, human intelligence would be like the body without a skeleton—complicated but useless. Humans can learn a tremendous amount about the world, but this remarkable capacity for acquiring, synthesizing, and interpreting information is made possible by the mind's innate structures, themselves serving as the foundations from which human beings think, create, discover, and innovate.

The human ability to acquire language from impoverished data stands in stark contrast to AI systems' need for massive amounts of data for even moderately successful functioning. Indeed, this point has been well established and was weaved through Marcus' (2018) article on AlphaGo and innate structure. But the most important lesson from POS is not that humans require only limited data to acquire mastery over a natural language; rather, it is that language is *not* what comes out of one's mouth or sees on this page. Indeed, if acquiring any given language depends on the principled knowledge encoded into one's brain/mind, then the implication is that *all* natural languages are merely variations of a discrete cognitive system. "Language" as an object of scientific study, then, is a property of the human mind.

How individuals *use* this cognitive system is quite another problem.

6.2 The creative aspect of language use

The POS argument begins with the *observation* that there exists a range of possible languages which humans can acquire. This observation is followed by a *description* of the human capacity for language, namely the acquisition of a rich grasp on complex grammatical structures from a young age and ability to generate new expressions without limit. Finally, this phenomenon is *explained* by postulating the existence of an innate "language faculty" which houses linguistic knowledge requisite for the acquisition of natural languages. POS is the result of a procedure which moves from observation of a part of human intelligence to an explanation for it, with subsequent empirical inquiry operating within these parameters (Mikhail 2011, pp. 21–23).

There is, however, an aspect of language use that can be observed and described but *not* explained. This aspect—known as the "creative aspect of language use"—is "the distinctively human ability to express new thoughts and to understand entirely new expressions of thought…" (Chomsky 2006, p. 6). CALU refers to ways in which language is *ordinarily* used by individuals, though its importance "is often overlooked…the simple fact is that humans deal easily and frequently with what does not exist, or what does not yet exist" (Kriedler 1998, p. 4).

Something does, however, get lost in this description, so it is important to explain CALU's uniqueness. Non-human animals routinely communicate in reference to direct stimuli in their immediate environments, such as bees locating nectar. But human language is not like this. Language use is free of any particular stimulus in one's environment, yet it is appropriate to the context and capable of being recruited infinitely.

As Kreidler says, “that is just what happens when the architect envisions a building not yet erected, the composer puts together a concerto that is still to be played, a writer devises a story about imaginary people doing imaginary things…” (Kriedler 1998, p. 4).

The ability to create new thoughts free of any specific stimulus and express them to others who understand them is remarkable, yet this ability is difficult to recognize unless we are willing to be puzzled. Even here, in this essay, one can find linguistic firsts. Many of the sentences here were thought and written by me for the first time in my personal history, and possibly for the first time in the history of the human species. Yet these sentences are immediately intelligible to those who read it, appropriate to the topic of generativism and AlphaGo, and bracketed by novel and productive expressions.

Thus, ordinary language use is “creative” in three respects:

1. Stimulus freedom: a particular linguistic expression is neither *caused* nor *determined* by an individual’s circumstances nor is it *random*;
2. Unbounded: there is no limit, in principle, to the number or kinds of sentences an individual can produce within or across contexts;
3. Appropriateness: despite thoughts or expressions being stimulus free and unbounded, they are nonetheless *appropriate* to the circumstances of their use, whether fictional or real (McGilvray 2017, p. 187).

Each component of CALU provides human beings with an intellectual potential. The unboundedness of language use sets the stage for an infinite number of productive thoughts and expressions. The appropriateness of language use allows for the productive exchange of thoughts and expressions between individuals within any scenario with which they are faced. The stimulus freedom of language use—CALU’s most powerful component—enables the peak of human ingenuity. Stimulus freedom provides humans with an ability to spontaneously create thoughts which can be linked to one another productively, but causally detached from the circumstances of their use. A problem in need of solving may elicit a thought relevant to the situation but not caused by it. Yet, CALU enables the individual to express this thought in a manner that makes sense to those around them. Even aesthetically,

> [o]ne can speak of elephants when there is nothing in the speaker’s environment that could conceivably be called a stimulus that caused the utterances. Or one could speak of Federico Lorca’s *Poet In New York* when the only conceivable stimulus in the speaker’s environment is elephants and the African landscape (Asoulin 2013, p. 230).

Still, CALU is a highly abstract concept which can be better grasped with prior conceptual distinctions. First, a generative grammar is a characterization of an individual’s linguistic *competence*; the innate knowledge of language an individual possesses unconsciously. Second, competence—while enabling the development of a linguistic capacity—is conceptually distinct from an individual’s *performance*; the concrete ways in which an individual *uses* their linguistic capacity (Mikhail 2011, p. 18).

CALU, as Chomsky (1982, pp. 429–431) understands it, falls under the category of *performance*, as it deals with the creative way language is *used*. The three components of stimulus freedom, unboundedness, and appropriateness which constitute CALU, while *enabled* by the principles embedded in linguistic competence, are perhaps the most powerful reasons to believe that human intelligence has a distinctive character.

This distinctive character, one might think, could be automated. Indeed, the idea that machines could replicate human creativity “intrigued the seventeenth-century mind as fully as it does our own” (Chomsky 2006, p. 5). Seventeenth-century characters like Galileo, Gassendi, and Hobbes, who believed (at points) that the world could be explained in “mechanical” terms, were taken to task by Descartes who instead argued that human linguistic creativity “escape[s] the methods of natural science” (McGilvray 2017, p. 187). Working within this tradition, Chomsky categorizes CALU as a “mystery” beyond the bounds of scientific inquiry because it is neither determined through stimulus–response relationships nor is it random, a conclusion he reaches by *lowering* his theoretical expectations for what a science of the human mind can realistically accomplish (Chomsky 2009a, pp. 91–93).

Human beings possess an ability which adds a remarkable dimension to their intelligence, but they are stuck with mere descriptions of it. Employing POS can allow scientists to uncover the principles which *enable* linguistic creativity, but it cannot break through the wall put up by an ability that cannot be articulated causally or probabilistically. This does not mean, however, that a description of CALU is worthless to understanding human intelligence. If language is a cornerstone of human thought, then the complexity and limits of thinking will be substantively shaped by the language faculty. CALU, then, plays a critical role in molding the character of human intelligence.

None of this is obvious. It is ironic that our ordinary use of language possesses a quality so remarkable but that so few of us are prepared to acknowledge it. Only with a willingness to be puzzled about our ordinary use of language and its relationship to the human intellect can we be prepared to admit that "nature rarely comports with commonsense intuition" (McGilvray 2017, p. 188).

6.3 Question 2 (science of intelligence) verdict

The lesson that nature and commonsense intuitions rarely align is curiously absent in discussions on automating intelligent behavior. In this vein, generativists consider the comparison of machine intelligence to human intelligence a strange one, applying human labels like "thinking" and "language processing." While Turing believed an imitation game "would bypass what he considered to be a dangerous haggling over definitions of the terms 'machines' and 'thinking'" (Fazi 2019, p. 813), generativists consider the comparison of machine intelligence to human intelligence as mistaking behavior—which an AI-enabled machine was meant to simulate—with the proper objects of scientific study, namely the internal systems of the human mind that *enable* an individual's behavior.

More than this, however, generativists argue that AlphaGo does not engage in "strategizing," no matter how decisive its victory against Lee. Human intellectual abilities like strategizing—enabled by rich, innate, yet flexible cognitive structures—*cannot* be reduced to the data which individuals manipulate to specific ends, *nor* can their uses be reduced to the mechanisms which enable them.

Could *anything* resembling CALU be said to exist in current AI systems? AlphaGo was able to play Go against Lee and take turns that were appropriate to the game and could, in principle, do so infinitely. However, this description is misleading. AlphaGo was programmed with a Go-specific endowment meaning it could not do anything *except* play Go. Moreover, it could only play Go on a specific board with specific stones with specific rules. Humans, in contrast, could easily create a makeshift Go board and play on it. Or they could readily replace standard Go stones with ones found in nature and play relatively normally. Finally, there is no comparison—not even a crude one—between AlphaGo's ability to "choose" movements on the Go board and human linguistic stimulus-freedom.

It is for these reasons that generativists resist the tendency to use terms such as "think," "learn," or "strategize" in reference to AI—these are not terms of scientific expertise which denote precise meanings. Rather, they are terms of ordinary language which lack the kind of abstraction needed to understand human intellectual capacities like CALU. Arguing over whether AlphaGo can "learn" or "strategize" is "as if we were to debate whether space shuttles fly or submarines swim…it is idle to ask whether legs take walks or brains plan vacations…".

(Chomsky 2009b, p. 104).

While terminological rankling is familiar to the study of AI, generativism's distinct perspectives on human and artificial intelligence are the product of a "methodological naturalism" on a par with the natural sciences' baseline approaches to inquiry. Generativism offers, as described above, a characterization of the *anthropomorphizing* of AI through a methodological dualism. By exposing this reasoning, generativism is then free to break off the shackles of these irrational assumptions and pursue a methodologically naturalist characterization of *human* intelligence, which serves as the baseline contrast for AI.

Below, a characterization of methodological naturalism and the study of human intelligence is constructed based on the foregoing remarks:

1. A Willingness to be puzzled about the human mind: Any scientific or philosophical approach to the nature of the human mind should begin with a willingness to be puzzled about the mind as an object of study. This should be on a par with the mindset adopted in the mature natural sciences prior to investigation.
2. Methodological naturalism: Once sufficiently cultivated, a puzzled mindset seeking to study the human mind should proceed with the same assumptions and standards of inquiry as the study of other natural phenomena.
3. Openness to counter-intuitive ideas about the mind: By adopting a methodological naturalism, individuals should be open to characterizations of human cognition that may seem counter-intuitive or bizarre yet offer explanatory depth.
4. Appreciation for the distinctiveness of human cognition: Being open to counter-intuitive ideas entails recognizing that humans possess abilities that are distinct from those of all other known species. Most notably, humans possess an inborn capacity for language which enables unbounded, stimulus free, and context-appropriate expressions of thought.
5. Embracing limits on human self-understanding: Out of this appreciation should come a recognition that certain aspects of human cognition will remain out of reach of scientific and philosophical inquiry temporarily or indefinitely, in part due to the limits of the human intellect. Features of human intelligence—such as CALU—are unlikely to be understood or explained.
6. A willingness to be puzzled about artificial intelligence: Finally, because of this reasoning about the nature of the human mind, a willingness to be puzzled about the possibility and nature of artificial intelligence is needed to pursue its study. Comparisons between human intelligence and artificial intelligence systems should be resisted in pursuing the study of the latter.

This chain of reasoning encapsulates the movement in generativist thought from a willingness to be puzzled about the human mind, to the POS argument and embracing the idea that some features of human cognition are simply off-limits to our inquiry, and finally, culminating in a recognition that AI—in virtue of both a pervasive methodological dualism and the requirements of a mature science—itself requires a willingness to be puzzled.

Our second, "Science of Intelligence" question is, therefore, answered through this chain of reasoning, providing an overview of the key assumptions and dispositions needed to study both human and artificial intelligence, and resisting methodological dualism.

7 Discussion

7.1 First research question: human-like perceptions of AI

Two research questions were explored in this paper. The first concerns the tendency to *anthropomorphize* AI systems by homing in on the prominent example of AlphaGo and the effect it had on Go players, academic observers, and public commentary. A generativist analysis, tasked with explicating the assumptions which make such anthropomorphizing possible, answered this question by characterizing the reasoning necessary for individuals to make the perceptual leap from assumptions about human intelligence and its study to assumptions about the study of AI. Drawing from the readily available facts regarding the case of AlphaGo, in addition to content analyses of commentators' reactions to AlphaGo and the broader "AI myth," this analysis looked beneath such social-scientific data to pinpoint a bundle of assumptions falling under a broader methodological dualism about the study of human intelligence which have been implicitly transferred to evaluations of AlphaGo. Interestingly, while methodologically dualist assumptions merit attention as a factor in reasoning about AI, this analysis also shed light on an "unwillingness to be puzzled" about human intelligence—a disposition or mindset rather than an assumption.

There are, however, drawbacks to this portion of the analysis. While the existence of a methodological dualism may not be highly controversial (opinions about its justification notwithstanding), this analysis can provide only a general indication that it is operative in evaluations of AI like AlphaGo. Why, for example, did move 36 by Deep Blue prompt Kasparov to see *more* humanness in this machine while critical observers felt a need to shift the criteria required for intelligence *away* from such demonstrations? Generativism cannot provide a confident answer to such a question, other than to say that methodological dualism may be pervasive yet vulnerable to reevaluation.

Future research using generativism should maintain its role as a *fundamental* driver of perception while attempting to pinpoint the circumstances under which methodological dualism is either challenged or reaffirmed in the context of AI. It should, furthermore, attempt to identify the conditions under which methodological dualism is *not* operative in AI anthropomorphizing. Finally, and perhaps most importantly, this analysis should guide not only evaluations of AI systems from outsiders' perspectives, but also of AI research itself. Should AI researchers lack a willingness to be puzzled about themselves and their artificial creations, they are likely to set unrealistic goals for its study.

7.2 Second research question: science of intelligence

The second research question concerns *justifications* for conceptions of AI. Generativism is in a unique position to answer this question. As an intellectual tradition in philosophy and the cognitive sciences, generativism offers mutually-reinforcing insights on both the nature of scientific inquiry (i.e., methodological naturalism) and the role of commonsense assumptions in perceptions (i.e., methodological dualism). In answering the first question, generativism sets itself up to fill in the blanks of how AI *could* be evaluated without irrational premises.

This answer reconstructed the interpretation of AIs like AlphaGo by formulating a separate, normative characterization of the relevant steps in reasoning. The influence of the perceptual analysis plays a clear role here in that the alternative chain of reasoning is built upon an overview of POS and CALU which require a willingness to be puzzled about human cognition, which was lacking in the first, descriptive characterization. This analysis accepted Marcus's (2018) characterization of AlphaGo Zero as benefitting substantially from its innate endowment, contrary to the claims of "tabula rasa" made by its programmers. It built on this, however, by explaining the steps needed to get to the POS argument employed by Marcus, namely a willingness to be puzzled and a methodological naturalism.

One of the most striking implications of this answer is how it exposes fundamental issues with the widely accepted "narrow" and "general" distinction in AI. The distinction between narrow and general is strange: is AlphaGo's ability to "play" Go really a "narrow" version of the human ability to strategize? The former is only "narrow" if we stop at the characterization of human cognitive faculties as independent structures. But, while studying cognitive faculties as independent is a hallmark of generativism, it is always recognized that they are sub-systems of the mind/brain. There is a certain narrowness to, say, the POS argument, but this operates in a broader context of these capacities' interactions with one another. The main lesson to take away here is that the trajectory from "narrow" to "general" AI may not be feasible with existing AI systems, and may not become feasible given the limitations of the science of the mind. Achieving this would be equivalent to achieving an explanation of CALU—possible in principle, but unlikely.

8 Conclusion

This paper addressed two research questions regarding the anthropomorphizing of AlphaGo and the lessons that can be learned by exposing a methodological dualism operative in evaluations of AI. It thus tackled a two-sided issue dealing with *descriptions* of the psychological processes involved with interpretations of AI and *prescriptions* (or, less stridently, suggestions) for how AI could be interpreted, which depends upon justified presuppositions of human intelligence.

It seems perfectly *natural* and *obvious* that AI be compared to human intelligence, and a part of the reader's mind may pull them away from this high-in-the-sky analysis, saying to themselves that *of course* machines performing characteristically human activities elicits anthropomorphizing! But the real contribution of generativism in turning this tendency into an object of study is in its ability to resist that temptation in favor of what has worked so well for the natural sciences: embracing a sense of wonderment about something as obvious as creations made in our own image.

Funding The authors did not receive support from any organization for the submitted work.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Asoulin E (2013) The creative aspect of language use and the implications for linguistic science. Biolinguistics 7:228–248

Berwick RC, Pietroski P, Yankama B, Chomsky N (2011) Poverty of the stimulus revisited. Cogn Sci 35:1207–1242. https://doi.org/10.1111/j.1551-6709.2011.01189.x

Bory P (2019) Deep new. Convergence 25:627–642. https://doi.org/10.1177/1354856519829679

Childers T, Hvorecky J, Ondrej M (2021) Empiricism in the foundations of cognition. AI Soc. https://doi.org/10.1007/s00146-021-01287-w

Chomsky N (1982) A note on the creative aspect of language use. Philos Rev 91:423–434. https://doi.org/10.2307/2184692

Chomsky N (1994) Naturalism and dualism in the study of language and mind. Int J Philos Stud. https://doi.org/10.1080/09672559408570790

Chomsky N (1995) Language and nature. Mind 104:1–61. https://doi.org/10.1093/mind/104.413.1

Chomsky N (2006) Language and mind. Cambridge University Press, New York

Chomsky N (2009a) The mysteries of nature. J Philos 106:167–200. https://doi.org/10.5840/jphil2009106416

Chomsky N (2009b) Turing on the "Imitation Game." In: Epstein R, Roberts G, Beber G (eds) Parsing the Turing Test. Springer, Dordrecht, pp 103–106

Chomsky N (2013) Problems of projection. Lingua 130:33–49. https://doi.org/10.1016/j.lingua.2012.12.003

Collins J (2004) Faculty disputes. Mind Lang 19:503–533. https://doi.org/10.1111/j.0268-1064.2004.00270.x

Curran NM, Sun J, Hong J (2020) Anthropomorphizing AlphaGo. AI Soc 35:727–735. https://doi.org/10.1007/s00146-019-00908-9

De Spiegeleire S, Maas M, Sweijs T (2017) Artificial intelligence and the future of defense. The Hague Centre for Strategic Studies, The Hague

Dong Y (2016) AlphaGo and the clash of civilizations. Foreign Policy Magazine. https://foreignpolicy.com/2016/03/18/china-go-chess-west-east-technology-artificial-intelligence-google/. Accessed 27 Nov 2021

Fazi MB (2019) Can a machine think (anything new)? AI Soc 34:813–824. https://doi.org/10.1007/s00146-018-0821-0

Jackendoff R (2008) Patterns in the mind. Basic Books, New York

Jebari K, Lundborg J (2021) Artificial superintelligence and its limits. AI Soc 36:807–815. https://doi.org/10.1007/s00146-020-01070-3

Kriedler CW (1998) Introducing English semantics. Routledge, New York

Lasnik H, Lidz J (2017) The argument from the poverty of the stimulus. In: Roberts I (ed) The Oxford handbook of universal grammar. Oxford University Press, Oxford, pp 221–248

Marcus G (2018) Innateness, AlphaZero, and artificial intelligence. ArXiv 1–18. arXiv:1801.05667.

McGilvray J (2017) Cognitive science. In: McGilvray J (ed) The Cambridge companion to Chomsky. Cambridge University Press, Cambridge, pp 106–175

Mikhail J (2011) Elements of moral cognition. Cambridge University Press, Cambridge

Natale S, Ballatore A (2020) Imagining the thinking machine. Convergence 26:3–18. https://doi.org/10.1177/2F1354856517715164

Silver D et al (2017) Mastering the game of Go without human knowledge. Nature 550:354–359. https://doi.org/10.1038/nature24270

Svensson J (2021) Artificial intelligence is an oxymoron. AI Soc. https://doi.org/10.1007/s00146-021-01311-z

Turkle S (2005) The second self, Twentieth Anniversary Edition. The MIT Press, Cambridge