Artificial Agency and the Game of Semantic Extension

Fabio Fossa, Politecnico di Milano

Abstract: Artificial agents are commonly described by using words that traditionally belong to the semantic field of organisms, particularly of animal and human life. I call this phenomenon the *game of semantic extension*. However, the semantic extension of words as crucial as "autonomous", "intelligent", "creative", "moral", and so on, is often perceived as unsatisfactory, which is signalled with the extensive use of inverted commas or other syntactical cues. Such practice, in turn, has provoked harsh criticism that usually refers back to the literal meaning of the words to show their inappropriateness in describing artificial agents. Hence the question: how can we choose our words appropriately and wisely while making sense of artificial agents? After a brief introduction (§1), in §2 I present the starting point of my argument, which consists in the assumption that the dimensions of technology and language are deeply entangled. In §3 I discuss how this assumption impinges on the issue of choosing the right words to talk about artificial agents. §4 is an exposition of the main features of the game of semantic extension, while §5 reviews the related opportunities and risks. Finally, §6 elaborates some practical suggestions on how to play the game well.

Keywords: artificial agency; language game; philosophy of technology; homology; analogy.

1. A thorny linguistic challenge

More and more technological artefacts are able to interact with us in increasingly flexible and engaging ways. In the scientific debate, such technologies are commonly addressed as artificial agents [AAs]. When talking about them, it might initially seem appropriate to resort to the words we have been using so far to speak of regular instruments. However, these words have started to sound weird and leave us dissatisfied, as if something we wished to say remained unsaid and something else was said we did not mean to. When we say that self-driving cars, conversational agents, or recommendation algorithms are nothing but tools – as hammers and dishwashers – it is evident that something is left unspoken.

Since tool-like descriptions of AAs are often deemed reductive (Ihde 1990; Prescott 2017; Gunkel 2018), it only remains to talk about these new technologies using words that we commonly use to talk about the agents they substitute and actions they reproduce. The car drives itself, as if it incorporated a driver. The conversational agent chats with us, answers our queries, listens to our voice. The algorithm knows what we like and what we want.

It is indeed common practice to make sense of AAs by using words that traditionally apply to living things – particularly, even though not exclusively, to animals and human beings. Examples of a similar use of language are so ordinary that there is little need to provide them. Think, for example, to the very word "agency", but also to other adjectives through which machine functioning is frequently framed, such as "trustworthy", "moral", "creative", "intelligent", "autonomous", and so on. Arguably, the tendency of describing AAs in biomorphic terms feels so natural that is often taken for granted and goes unnoticed.

This does not solve the problem. The biomorphic words we use to talk about new technologies bring along a whole baggage of meanings, expectations, and action patterns that only partially fits the new usage. At the same time, new usages feedback onto the semantic field of the words we resort to, thus giving them new connotations. In turn, these new connotations reflect themselves also on the objects or phenomena the word was normally associated with before the new usage. I call this linguistic phenomenon *the game of semantic extension* [GSE]. As a result, in this case as well we risk saying too much or something else than what we wished to, while the meanings of our words slightly shift and silently impinge on their more established usages. Even though AAs assimilate so smoothly in the fabric of human existence, thus, they still are enigmas we lack the words to crack. A gap seems to open between the ways we experience them and the words we choose to make sense of them.

Even though GSE often passes undetected, some discomfort has been expressed and is usually signalled through the extensive use of syntactic cues such as inverted commas, words written in capital letters, and similar strategies[1]. Such practice has provoked harsh criticism that usually refers back to the literal meaning of the words used to show their inappropriateness in describing AAs (Dreyfus 1965; Searle 1980). Notwithstanding the criticism, the syntactic solution is still popular, which highlights an epistemological need that awaits adequate inquiry. Hence the question: how to choose our words wisely while making sense of AAs?

The aim of this paper is to shed light on the main features of GSE in order to propose some practical suggestions to its mindful and scientific play. The argument is structured as follows. In §2 I present the starting point of my research, which consists in the assumption that the dimensions of technology and language are deeply entangled. In §3 I discuss how this assumption impinges on the issue of choosing the right words to talk about AAs. §4 is an exposition of the main features of GSE, while §5 reviews the

[1] Many examples can be found, for instance, in the debate on trust and digital technologies: "e-trust" in (Taddeo and Floridi 2011), "TRUST" in (Grodzinsky, Miller, and Wolf 2011), "robotrust" in (Pagallo 2010).

related opportunities and risks. Finally, §6 elaborates some practical suggestions on how to play GSE well.

2. Technology, language, and meaning

The basic assumption of the argument set out in this paper is that the meaning we attach to technological artefacts is the product of the interplay between our own intentionality, their material nature, and their sociolinguistic representations – put more simply, between what we want to do with them, how they are made and how they are talked about. Even though an active role must be acknowledged to human initiative, in order to understand how technologies assume meaning it is necessary to take into account also what is already given, that is, the material and linguistic contexts within which artefacts emerge as objects of experience and use. The meaning of our artefacts is not fully instituted by the intentional activity of some absolute subjects such as designers, producers, users, and so on. Rather, meaning gets co-constituted in the recursive dialectics between the human, material and linguistic poles.

The entanglement of language and technology has been widely explored in (Coeckelbergh 2017a), on which I draw to introduce the background against which my remarks are to be read. The discussion of Coeckelbergh's research also serves the purpose of specifying what this paper strives to accomplish.

Coeckelbergh's argument starts from the acknowledgement that language and technology exhibit more than a superficial similarity. Words and artefacts, being both involved in the experience of use, can be understood primarily as instruments, or tools. Therefore, clarifying what happens when we use words may shed light on what happens when we use artefacts, and vice versa. It is in the performance of use, in fact, that words and artefacts get their meaning and that through words and artefacts the world and our selves also become meaningful.

Ultimately, the whole endeavour depends on providing the most adequate interpretation of what the concepts of "use" and "instrument/tool" consist in. This is necessary because the two notions are often characterised in too linear a fashion. Sure enough, at first sight "to use" stands for "to handle instruments as means in order to accomplish goals", where "goals" are objectives set by the same individuals who use tools. However, the relation between individuals, tools and the world is much more complex. The simplicity of the linear model is unable to accommodate for the active role played by the many elements that constitute the socio-technical context within which use takes place. Human intentionality is not the only factor here. Actually, the way in which words and artefacts become meaningful cannot be properly understood unless we start from the presupposition that subjects, words, artefacts, and worldly objects co-shape each other through use. Every element in its peculiarity is always embedded in a wider context and the interplay among these relations, which is set in motion in the experience of use, co-shapes meaning. Drawing on Wittgenstein's terminology, Coeckelbergh suggests that meaning emerges as the result of language games and technology games performed against the background of a whole form of life: a wider context where cultures, social bonds and relations, politics, narratives, characters and bodies all play their part.

This hermeneutical shift brings to light a similar pattern in both contemporary philosophy of language and technology. On the linguistic side, it has been realized that language in use is not a transparent medium. Rather, it concurs to structuring and shaping meaning and, therefore, human experience and agency. On the technological side, the material nature of artefacts has similarly been pointed at as an element that, in use, co-shapes meaning along with human initiative. Even though several differences distinguish the two approaches, a common tenet surfaces: there are no such things as neutral mediations, uses, or instruments. In the experience of use, users and instruments partake in a performative event where meaning is co-shaped.

Language and technology, therefore, concur to structure human experience and agency. In a sense, they represent the actual conditions, the applied grammar of the human experience of meaning. With this transcendental move, Coeckelbergh goes beyond the opposition and the reduction of one dimension to the other. Weaving the two threads together, he elaborates a coherent model to analyse the role technologies and language play in the process through which meaning matches with words, artefacts, worldly objects and our selves. At the heart of the model lies the experience of use, a kaleidoscope looking into which it is possible to appreciate all the contributions and influences as diverse as they are in type, origin and strength that result in the emergence of meaning.

3. Artificial agents, language, and meaning

The process through which AAs intertwine with meaning must be counted among the many phenomena that can be fruitfully analysed from this perspective. This process displays an eminently linguistic nature: various meanings become initially associated to these new technologies through the ways in which we (consciously or unconsciously) talk about them – or, more precisely, through the words we choose to describe their functions, their roles, the places they occupy in our worldview (Johnson and Miller 2008; Calo 2016). Following Coeckelbergh's suggestions, it becomes important to enlarge the focus in order to include, in addition to the intentions of those speaking and the material characteristics of the artefact, also the semantic field to which the used words belong. This field, in fact, positively concurs to shaping the meaning of new technologies and the ways they are used.

In order to "choose and value words", as Wittgenstein (1958, 218) suggested, we consider "the familiar physiognomy of a word, the feeling that it has taken up its meaning into itself" or "the field of force of a word" (219). When we pick a word, hence, its meaning is all we have on which to base our considered choice. The meaning of words, however, is not monolithic. Rather, it fluctuates within a range depending on the contexts in which the words are used and, accordingly, the objects they are supposed to be linked to. What should we do, then, to choose our words wisely?

This is the situation researchers in artificial agency constantly face. AAs are enigmas we try to crack, initially at least, *by looking for the best choice of words*. Language and technology are difficult to separate: a new technology is always discovered through the words used to conceive it and the discourses that accompany its functioning and use. What are the right choices of words? How can we establish whether some words are adequate or not? Who is entitled to do such an evaluation?

Perhaps these questions are too hard to be answered. In the meantime, negotiations have already begun. Just think of the many debates surrounding the adequacy of describing AAs by using words belonging to the semantics of personhood (Bryson, Diamantis, and Grant 2017), trustworthiness (Coeckelbergh 2012; AI HLEG 2019), creativity (Coeckelbergh 2017b), companionship (Coeckelbergh 2011; Johnson and Verdicchio 2019a), slavery (Bryson 2010; Gunkel 2018, 117–130), animal life (Johnson and Verdicchio 2018), or even autonomy (Johnson and Verdicchio 2017) and agency (Franklin and Graesser 1996; Laukyte 2017; Johnson and Verdicchio 2019b).

Besides, as some notice, this language game seems to follow its own inner rules. As a consequence, it might be pointless – if not detrimental – to oppose its logic and try to control its movements. From this perspective, so it might seem, all we have to do is to observe the language game unveiling itself and ponder over its results. A similar claim seems to be implied, at least partially, in Coeckelbergh's approach to the matter[2] and also in other authors who share his relational perspective (Gunkel 2018, 159–183).

In a way, this tendency recalls an insightful and well-known suggestion by Turing (Turing 1950) – who was well aware of the pivotal role played by language use in relation to the way new technologies are understood. Inquiring into the possibility of attributing "thinking" to machines, he writes:

> "The original question, 'Can machines think?', I believe to be too meaningless to deserve discussion. Nevertheless I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted" (442).

If this is the case, little space seems to remain to critique. We just have to play the game and see where it leads us. It would be useless indeed, almost quixotic, to resist its current. Accordingly, if a particular humanoid robot is widely experienced by its users as a companion, a friend or a trustworthy member of the family, it seems to make little sense to take a step back and ask to what extent this use is reasonable and adequate or deceitful and dangerous (and thus, perhaps, should be avoided). If the game justifies itself, there is no room for critique: it is black box that outputs unquestionable meanings.

The weakness of this perspective is that we run the risk of hypostatising the experience of use and, perhaps, also of universalizing experiences of use that, however, are only partial and particular. Oddly, the difficulty was already underlined by Turing at

[2] See, for example, (Coeckelbergh 2017b, 296): "if more people were to speak about what machines do in terms of 'artistic creations' and 'works of art', than would we really have an objective basis for saying that they are wrong? Even if today we might be opposed to the very idea of machine art, in the course of time, our language might change and let the machines in through the backdoor".

the beginning of the very same paper, thus leaving to the reader the daunting task of finding a way out of the maze:

> “I propose to consider the question, ‘Can machines think?’ This should begin with definitions of the meanings of the terms “machine” and “think”. The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words ‘machine’ and ‘think’ are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question ‘Can machines think?’ is to be sought in a statistical survey such as a Gallup poll. But this is absurd.” (433)

At this point, questions similar to those already emerged are unavoidable: Whose language use is the right one? Why? And how can we establish which language use is the most accurate? It seems we are back to the beginning, and the beginning is a dead end.

Luckily enough, our task aims at a slightly different set of issues. From a philosophical point of view, in fact, the most important matter is not to dictate how to use language, but to develop a critical attitude to word usage. This stance is not exclusively passive: it is supported by a normative intention and strives to channel the debate into a form that makes it amenable to rational discussion, considered judgment and meaningful disagreement. To use the terminology of hermeneutics, we need to ask: how can we get in the circle of understanding in the right way (Heidegger 1996, 143)? What does it take (or: what does it mean) to *play the language game well*?

Only if the game does not justify itself, but can be played better or worse, then there is room for critical debate concerning the quality of our word choices – i.e., concerning their reasonableness or unreasonableness, accuracy or inaccuracy, meaningfulness or riskiness, appropriateness or inappropriateness, and so on. I take the semantic negotiation I hinted at as a proof that this possibility is already presupposed: it makes sense to discuss about our word choices, to argue in favour of some of them and against others, and to provide reasons for them. Sure enough, words can be used cunningly or lightly. This is sometimes the case with AAs. Nonetheless, it is not just a matter of taste, rhetoric boldness, or academic pedantry. The role philosophy might play is to raise awareness on word usage; show that words co-constitute our understanding of what we are talking about (and, thus, our experiences and actions); and, finally, help develop awareness of the strengths and weaknesses associated with every word choice. This might possibly be the only assurance we could ever get that we are those actually playing the game, and not those being played.

In what remains I will focus the attention on the language game that, in my opinion, has characterised the discourse on AAs since its very beginning – I call it *the game of semantic extension* – and I will try and offer some suggestions to play it well.

4. The game of semantic extension

In this section I present a description of how the game is played when players find themselves in the tight spot of making an effort to "find the 'right' word" (Wittgenstein 1958, 218) to talk about AAs. As anticipated, I propose to interpret this linguistic situation as a game of semantic extension.[3]

While facing the hard task of finding words to talk about these enigmatic technologies, we are commonly led by what could be called the *dominant character of similarity*. The novelty of the artefacts is approached with words the meaning of which

[3] In order to do so, I extensively draw on the book I just quoted that has already proven extremely useful in this inquiry, i.e., Wittgenstein's *Philosophical Investigations* (Wittgenstein 1958). Other important observations on what it means to play a game and what this entails in relation to meaning can be read in Hans-Georg Gadamer's *Truth and Method* (Gadamer 2004, 102–130).

is somewhat similar to what requires to be expressed. The new technology is wrapped in words mostly on the ground that “we see a complicated network of similarities overlapping and criss-crossing” (32) and we pick our words following this intuition. Similarity, then, is the main logic that drives the linguistic process through which the semantic field of a word is extended in order to apply to AAs.

As seen in §1, AAs exhibit a connection to at least two main semantic fields. The first is obviously that of artefacts as instruments. As anticipated, however, even though tool-related words readily offer themselves as possible solutions, they also seem to fall short of what asks to be expressed here. This often leads to experimenting with the much more enticing semantics of life. Indeed, it feels all too natural to use biomorphic words to make sense of AAs. Consider the words that usually raise suspicion – agency, autonomy, trust, intelligence, consciousness, creativity: they all belong to the semantic field of life. Perhaps the most evident effect of GSE, then, is that the meaning of words that are commonly used to talk about organisms is extended and brought to bear on AAs as well, on the basis of similarities or “family resemblances” (32) exhibited by the two phenomena. The logic of similarity nudges our choice of words away from the semantic field of instruments and on to that of life.

In line with the remarks presented in §2, it seems reasonable to state that this peculiar semantic extension is not to be studied exclusively on an intentional level, but also with an eye to the material and linguistic contexts. This is to say that some reasons why the extension occurs are not merely linked to the intended practicality of the word usage, but are possibly connected to some material features of AAs themselves, which make the usage functional in the first place, and to previous language uses that steer our preferences towards biomorphic word choices.

Let us consider the material level first.

In my opinion, semantic extension is supported by a material connection between AAs and organisms – a connection that is rooted in and mediated through design. In the case of AAs, the similarity with organisms is a condition of design and, therefore, is instantiated in the technology itself. As Norbert Wiener put it, it is certainly true that AAs are not the pictorial image of organism. Still, they exhibit a deeper connection to living things, one that revolves around the reproduction of the functions they execute:

> "Thus, besides pictorial images, we might have operative images. These operative images, which perform the functions of their original, may or may not bear a pictorial likeness to it. Whether they do or do not, they may replace the original in its action, and this is a much deeper similarity." (Wiener 1964, 31)

Even in cases where design is not strictly bio-inspired, a *family resemblance* to organisms is constitutive here. AAs are technological products that carry out functions independently from human intervention and supervision while adapting to changes in the environment and automatically tweaking their functioning on the basis of previous work sessions in order to maximise results. Now, organisms are the only existing entities which carry out functions in a similar way. The bulk of designing AAs is the reproduction of the ability of autonomous functioning, which is first of all a distinctive ability of organisms. Although not necessarily in a specific fashion, then, artificial agency is generally modelled on organic agency. It could hardly be otherwise, since organic agency is the only available form of agency and, therefore, the most immediate model for such reproduction. This offers a strong material support to GSE and fuel the tendency to apply a biomorphic vocabulary.

Moreover, linguistic support to the same tendency is to be found in cultural heritage. In western culture, for example, AAs have been objects of imaginative, mythological, and magical thinking long before becoming objects of scientific and technological inquiry (Mayor 2018). Since the ancient time of Homer's *Ilyad* human imagination has entertained the dream of artificial products capable of behaving like organisms or human beings. We have been applying the semantics of life to artefacts ever since. In the last century sci-fi literature, movies, TV series, and video games have taken up the baton long carried by mythology and magical speculation, embedding the habit of framing AAs in biomorphic words even deeper into our cultural mindset. Hence, the semantic extension of organic vocabulary to AAs seems firmly supported both from a material and cultural points of view.

As a result of these conditions, the GSE inclines us to interpret AAs as duplicates of organisms, suggesting that no relevant differences distinguish the two. From an epistemological perspective, this means assuming a framework of *homology*, where archetypes and copies can entirely be discussed by referring to the very same concepts and words. However, the connection between AAs and organisms might not be one of duplication, but of imitation. In imitation, archetypes and copies show not just similarities, but also relevant differences – so that both aspects must be taken into account. The epistemological framework is now one of *analogy*, where elements clarify one another precisely through their similar dissimilarity or, that is the same, dissimilar similarity. As I show in the next section, opportunities and risks connected to GSE depend on whether it leads to frame the relation between AAs and organisms as a case of duplication/homology or imitation/analogy.

5. Impacts, opportunities, and risks

The persistence of the linguistic issue in the literature on artificial agency suggests that the dominance of similarity underlying GSE is a tendency that meets deeply entrenched needs. As seen, the discussion puts a tremendous pressure on ordinary language, which is expected to offer appropriate and comprehensible words to talk about AAs. In this situation, the convenience of fully exploiting the logic of similarity is hard to resist. Communication by similarity is intuitive and effective: it allows quick and powerful descriptions, providing easy and imaginative access to complicated technologies which, however, are increasingly part of our daily experience. Through the logic of similarity conveyed by GSE, in a sense, the familiarity of interactions is met with the familiarity of ordinary language use, which allows AAs to be smoothly integrated in our worldview. By addressing complex technologies as “autonomous”, “agents”, “creative”, “intelligent”, “trustworthy”, and so forth, a rich, powerful and familiar characterization is made available for everyone to apply in order to make sense of a rather obscure and puzzling technology. The shortcuts thus provided help us to manage the complexity of the technology and to push communication forward.

Moreover, it could also be argued that there is no other equally effective alternative than playing the game according to its logic. Any attempt to syntactically signal the difference between copies and archetypes sounds factitious, convoluted, and rather inconsistent with the smoothness of ordinary language. Likewise, counting on the formal language of computer science or devising an entirely new lexicon for AAs are also solutions that are either too technical or tortuous to truly compete with the naturalness of a similarity-based logic. Besides, even these two ‘formal’ strategies must rely on ordinary language to some extent. So, these sure are possible ways to play the game, but bounded to be outclassed by the other team’s style of playing. In addition, it might also be noticed that biomorphic language is actually the only language we have to start thinking about automating functions normally executed by organisms, so that it originally and genuinely belongs to the effort of designing AAs and understanding them. How could we imagine self-driving cars without any reference to human driving, or recommendation algorithms without any reference to the human experience of recommending contents, or assistive healthcare robotics without any reference to caretaking tasks usually executed by medical personnel?

It seems that, to have a chance in the game, players must be open to embrace its underlying logic. Once the game is started, it appears almost unavoidable to humour its inner impetus. However, acquiescing in the dominance of similarity evidently does not come void of any risk. The logic of similarity is addressed as dominant since it overwrites a recessive competing logic, which is of course that of difference. This means that GSE, by its own momentum, conveys a tendency to highlight what is similar at the expense of what is different. What is similar, as said, is most evidently the connection between AAs and organisms. In so doing, GSE promotes homology over analogy, duplication over imitation, and suggests that the application of biomorphic words to AAs is to be taken literally. Using the same words to describe both the organisms and AAs implicitly conveys that no significant difference separates the two objects[4].

What is worse, it is often difficult to realize that the logic of similarity is actually at work. As Wittgenstein noted, new usages hide themselves behind "the uniform appearance of words when we hear them spoken or we see them written" (1958, 6) – if not even when we speak and write them ourselves. This implies a patent peril: if differences are worth noticing, the dominant character of similarity may get in the way and cover them. Due to what has been said, GSE seems to automatically promote homology and contribute to blurring the line between organic and artificial agency. Extra, conscious efforts are required not to loose track of what differentiates the two kinds of agency.

[4] For an example, see (Fossa 2017).

This may result in two undesired situations.

First, GSE may engender misguided mental models and, subsequently, irrational expectations towards the technology itself. The use of a biomorphic vocabulary might nudge users into projecting onto technologies qualities that are commonly associated with living things but extraneous to AAs – like, for instance, the possibility of experiencing love, of caring, or of being trustworthy. A similar preoccupation arises in the debate over the dangers of anthropomorphising AAs, effectively summarized by Bryson as "misassignations of responsibilities and misappropriations of resources" (Bryson 2010, 63). This is no surprise. Insofar as its inner logic may promote anthropomorphic word choices, GSE must be counted among the factors that cause users to humanize AAs. Moreover, the tendency of opting for biomorphic words may also have an epistemological effect, in that it will foster the subsumption of AAs under categories to which they belong only partially – or, better, analogically. This might be detrimental in the long run, since it makes it harder and harder to pinpoint what is peculiar of artificial agency vis-à-vis other forms of agency. Getting this right, however, is crucial to develop adequate knowledge concerning the ways in which AAs is to be conceptually framed and socially operationalized.

Secondly, semantic extension does not work only one way. When the semantic field of a word is extended to accommodate for a new application, the new usage cannot but feedback onto the usual usages of the word. To quote, again, an image by Wittgenstein, we might say that the semantic "atmosphere" (Wittgenstein 1958, 48) of a word gets slightly, almost imperceptibly altered by any new element added to its composition – and the alterations spread throughout its overall extension. The extended, techno-related meaning of the words used feedbacks onto their original meaning. This may induce to frame biological and human activities by reference to technological criteria, which are usually easier to measure and control, thus forcing phenomena into linguistic and conceptual schemes to which they do not belong[5]. When this happens, we risk moving from imitation to duplication – or from analogy to homology – without realizing it and, thus, impoverishing both the semantic richness of our words, the accuracy of our concepts and the rationality of our agency. So, if the semantic feedback is upheld not as just a new layer of meaning, but as the true or scientifically most accurate meaning of the word, the risk arises of reducing the richness and mobility of the semantic field and, therefore, the degree of adequateness it can provide[6].

6. Playing the game well

Now that opportunities and risks connected to GSE have been sketched, it is possible to search for an answer to the question: how can the game be played well? In fact, more proactive styles of playing could also be imagined, where the risks associated to the logic of similarity are contained by reflectively taming its overabundant power rather than blindly complying with it. So, let us ask: in order to play the game *well*, how should players behave in relation to the underlying impetus towards highlighting the similar?

To be true, the previous analysis might also raise a different question, which asks whether it is really a good idea to get involved in the game at all. The risks to which the game exposes us might seem too dangerous to take it as granted that play we must. Consequently, stepping out of the game might appear as a potential workaround

[5] See, for plenty of examples, the use of proxies for measuring human virtues such as loyalty or dependability and the connected epistemological and ethical issues brought up by (O'Neill 2016). From the point of view of biology, see (Boldt 2018). This issue is also extremely visible in Wiener's theoretical writings on cybernetics.

[6] For an example, see (Fossa 2018).

to the difficulties we experience when we face the challenge of making sense of AAs. Unfortunately, this move seems not to be applicable in our case, since there is no court to leave. If we follow through the analogy, we find soon enough that the court where GSE is played is common language. There is no other way to make sense of new technologies in a sharable and scientific way if not through language. Language, however, is a dimension, a horizon we cannot step out of as we step out of a court. Once new technologies enter the domain of human experience, the game is on – and there is no way out of it.

Acknowledging the impossibility of getting out of GSE, which makes it more of a medium rather than a well-delimited practical domain, clears the field of some misconceptions or simplifications concerning this phenomenon. First, it would be simplistic to reduce GSE to a cunning marketing strategy targeted at generating hype and attract attention, as if its establishment would depend exclusively on the intentions of the players. We are not in the position of starting the game, nor of stopping it, but just of choosing how to play. Sure enough, GSE can be played in a way that exploits the dominance of similarity to make bold claims and take advantage of the fascination for sci-fi possibilities. However, GSE cannot be reduced to this game style. Indeed, it is the condition for this game style to express itself. Neither would it be reasonable to disqualify the whole linguistic phenomenon as an irredeemable source of confusion and misunderstanding. This would be a rather paradoxical move, since there is no alternative that might yield better results. The fact that GSE is inevitable does not exclude by principle the possibility of it being useful and legitimate, provided that it is played with due care. Rather, this fact only excludes the possibility of setting it aside.

Also, it is clear what follows if we play the game poorly: we will engender misunderstandings concerning AAs while threatening the semantic power and adequacy of notions surrounding life in the process. This outcome seems to follow from complying too much to the most apparent logic of the game, the dominant character of similarity. Riding the inner impetus of the game will leave the possibility open for the logic of similarity to affirm itself indeterminately. This, however, we know to be problematic. Too passive a game attitude will lead players not to play the game, but to be played by it. Through language use, new technologies would be categorised only according to what they are similar to, missing out their specificities. What is left to determine, then, is a matter of *style*. It is crucial to learn how to play the game well.

In light of what has been said so far, it appears that playing well entails playing *actively* and *reflectively*. The aim is not to let the logic of similarity impose itself on other aspects that equally matter in the complex and compound environment where meaning-giving experiences take place. As discussed in §2, these aspects are, among others, the personal intentions that involved us in the game, the material features of the technology, its socio-technical contexts, its linguistic profile, and so on. However, the imposition of similarity is precisely the outcome the inner movement of the game tends to by its own momentum. This is why a conscious, reflective, and critical effort is needed. Since the logic of similarity must be countered, active methodological care is required.

Accordingly, user experience and social practice cannot be turned into oracles of meaning-giving, although they surely remain dimensions to be considered. What emerges from the observation of how language is used when new technologies are talked about in real life situations can certainly provide philosophical inquiry with interesting data to evaluate. However, no conceptual understanding is offered by simply taking for good all the ways ordinary language is actually used. To claim that thinking should passively stick to ordinary language use would mean to take as the end of philosophical analysis what is actually only its beginning.

The philosophical side of the problem, in fact, revolves around the necessity of shedding as much light as possible on the hidden processes that influence word choosing by reflecting on the many criteria that might help us to be aware of, comfortable, and satisfied with the words we use. To use a recurring image, philosophical analysis strives to open the black box of common language uses concerning AAs. This, of course, entails submitting criticism when terms are chosen uncritically or with scarce awareness of their semantic implications. However, the primary objective is not to regulate the use of language – a rather pointless effort – but to foster a scientific attitude in the discourse over AAs and to make the same discourse amenable to rational (dis)agreement, in the hope that this will contain the spread of language misuses and deceitful representations on a wider social scale. To cultivate a reflective attitude towards language use and not to fall prey of the game of mirrors set up by the dominance of similarity is definitely a task hard enough for philosophy to tackle.

As it is immediately clear, the best way to curb the power of similarity is to keep an eye on *difference*. Contrary to the logic of similarity, which thrives against ill-defined, opaque, and ambiguous backgrounds, paying heed to differences will help specify arguments and discourse. In fact, it will necessarily lead to clarify the conditions under which a particular choice of word applies, the extent to which it does, and the purposes that are supposed to be accomplished through it. Playing well means taming the dominance of similarity, resisting the lure of duplication and homology, and thinking analogically with the aid of the notion of imitation. To do so, it is necessary to nurture a critical attitude towards language use.

Attaining a critical attitude towards the use of language in GSE implies first of all to duly acknowledge the advantages and limitations for every word choice. To use a metaphor, every word casts a different light on the phenomenon we wish to talk about. Depending on their meanings, words illuminate it from a specific angle, letting some aspects of it shine while, at the same thing, casting shadows on others. As a consequence, every word has its merits and demerits in the effort of framing new technologies such as AAs. Developing a critical relation to language in this respect entails to choose words reflectively, taking time to clarify which aspects they allow us to grasp and to determine which aspects remain covered instead. Moreover, due to the recursive structure of GSE discussed in §5, it also requires to consider how a particular word choice might feedback onto the semantic context of the chosen word and to take action were the feedback to cause confusion or illegitimate projection of meaning patterns from one specific domain of application (e.g., artificial agency) to other related, more general contexts (e.g., human agency).

Reflections on the adequacy of a word choice, thus, must strike a balance between two poles. The first pole is the factual evolution of language, according to which the meaning of words cannot be abstractly fixed but is significantly regulated by their usage. However, as clarified, usage is not enough: it is just one element of the game. The second pole is the semantic field of words, which may very well be flexible and adapt to different applications, but still exhibits a determinate nature which founds the possibility of distinguishing between unproblematic and problematic word choices. This second dimension, moreover, is intimately connected to the other given elements that compose the material context in which the performance of use takes place. All these elements concur to shape the semantic field of the word and work as levees that channel the mobility of usage into a form that can lend itself to rational inquiry and discussion.

Even if the metaphorical evolution of language is fully acknowledged, then, still there is room for distinguishing between unproblematic and risky word usages.

Critical awareness of how words change in light of technological advancement helps see through the opaqueness of ordinary language and to exploit the effectiveness of its flexibility without falling prey of its illusions. Taking time to ponder over word choices and looking for dissimilarities will help to keep discussion as transparent as possible and, thus, amenable to rational examination and meaningful criticism without overformalizing common language use or pretending to step outside of it. Since this methodology demands to explicitly disclose the assumptions behind any word choice and to trace the boundaries of its application, it provides sharable criteria to foster rational scrutiny, meaningful disagreement, and linguistic accuracy. When GSE is played well, it definitely makes sense to argue for or against a word choice in relation to its reasonableness, appropriateness, and riskiness.[7]

To sum up, making an effort to funnel the power of similarity by acknowledging the determining role of differences allows to avoid messy generalisations, inaccurate game of mirrors, and simplistic reductions of copies to archetypes and vice versa. By assuming a reflective, critical and difference-based attitude towards language use, the inner logic of similarity is orderly channelled, thus avoiding dangerous misconceptions and the spreading of unreasonable expectations. Following the hint of difference helps to see through what is already given – the semantic, practical, socio-political, material contexts – and the players' intentions, thus keeping our lexicon both accurate and significant. Playing the game well demands to resist the pull of homology and duplication while familiarising more and more with the delicate methodological

[7] I tried to do this in (Fossa 2019).

framework of analogy and imitation, where the validity and meaningfulness of claims always depend on the assumptions embraced and the epistemic domain they establish.

7. Conclusion

Since GSE represents the condition of possibility of any argument or discourse on new technologies, such as AAs, it is pointless to try and eradicate it from the realm of science. On the contrary, it is necessary to explicitly assume the difficulties it implies and learn how to cope with them, while at the same time taking advantage of what it has to offer. GSE is played well when players engage in reflective thinking on the conditions and limits of word choices. This task can be accomplished only by incorporating differences to the game style. By doing so, the risk will be contained of choosing crucial words instinctively, unattentively, or on the basis of reckless, opaque generalisations.

In the case of AAs, GSE allows powerful, imaginative, and easy communication throughout the scientific community and the general public. If played with style, it represents a great opportunity to create knowledge and spread awareness concerning a key technology that is changing how we experience the world and how we act in it. At the same time, however, GSE risks hiding the difference between imitation and duplication and pushing us to think homologically where we should think analogically. In so doing, words that are now in inverted commas might lose them and be used in their literal sense—with illusory outcomes, dangerous misconceptions, and harmful consequences.

If the dominant character of similarity were left to impose itself uncontested, we would also be exposed to a worrisome epistemological risk. In fact, we would reduce the novelty of the technology to the features of its closest relatives, missing out the opportunity of addressing the *specific* sense in which AAs may be legitimately said to be intelligent, creative, autonomous, and so on—that is, to determine the most proper characteristics of artificial agency, without reducing them either to its organic counterpart or to simple tool use. From this philosophical endeavour depends our social understanding of the technology and, therefore, the organisation of social action surrounding it.

In conclusion, I believe that much attention should be paid to the way in which words are used here. A naïve or enthusiastic word usage may generate false expectations, illusions, and ultimately deception (if not even self-deception) in both researchers and members of the public. However, too strict an approach to the use of language may lead to communication failures, obscure jargon, and confusion. GSE must neither be embraced as the deciding factor in the extension of life-like qualities to AAs nor be discredited as hopelessly misleading, though problematic it is. Finding a middle way between these two extremes will make it possible to address the precise domain of AAs, as differently similar as they are to tools and organisms.

References

AI HLEG. 2019. *Ethics Guidelines for Trustworthy AI*. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

Boldt, Joachim. 2018. "Machine Metaphors and Ethics in Synthetic Biology." *Life Sciences, Society and Policy* 14 (1). Life Sciences, Society and Policy. doi:10.1186/s40504-018-0077-y.

Bryson, Joanna J. 2010. "Robots Should Be Slaves." In *Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issue*, edited by Yorick Wilks, 63–74. Amsterdam: John Benjamins. doi:10.1075/nlp.8.11bry.

Bryson, Joanna J., Mihailis E. Diamantis, and Thomas D. Grant. 2017. "Of, for, and by the People: The Legal Lacuna of Synthetic Persons." *Artificial Intelligence and Law* 25 (3). Springer Netherlands: 273–291. doi:10.1007/s10506-017-9214-9.

Calo, Ryan. 2016. “Robots as Legal Metaphors.” *Harvard Journal of Law & Technology* 30 (1): 209–237.

Coeckelbergh, Mark. 2011. “You, Robot: On the Linguistic Construction of Artificial Others.” *AI and Society* 26 (1): 61–69. doi:10.1007/s00146-010-0289-z.

Coeckelbergh, Mark. 2012. “Can We Trust Robots?” *Ethics and Information Technology* 14 (1): 53–60. doi:10.1007/s10676-011-9279-1.

Coeckelbergh, Mark. 2017a. *Using Words and Things. Language and Philosophy of Technology*. New York: Routledge.

Coeckelbergh, Mark. 2017b. “Can Machines Create Art?” *Philosophy and Technology* 30 (3). Philosophy & Technology: 285–303. doi:10.1007/s13347-016-0231-5.

Dreyfus, Hubert L. 1965. *Alchemy and Artificial Intelligence*. https://www.rand.org/pubs/papers/P3244.html.

Fossa, Fabio. 2017. “Creativity and the Machine. How Technology Reshapes Language.” *Odradek* 3 (1–2): 177–213.

Fossa, Fabio. 2018. “Artificial Moral Agents: Moral Mentors or Sensible Tools?” *Ethics and Information Technology* 20 (2). Springer Netherlands: 115–126. doi:10.1007/s10676-018-9451-y.

Fossa, Fabio. 2019. “«I Don’t Trust You, You Faker!» On Trust, Reliance, and Artificial Agency.” *Teoria* 39 (1): 63–80. doi:10.4454/teoria.v39i1.57.

Franklin, Stan, and Art Graesser. 1996. “Is It an Agent, or Just a Program?: A Taxonomy for Autonomous Agents.” In *Third International Workshop on Agent Theories, Architectures, and Languages*. Springer-Verlag.

Gadamer, Hans-Georg. 2004. *Truth and Method*. Edited by Joel Weinsheimer and Donald G Marshall. London; New York: continuum.

Grodzinsky, Frances S., Keith W. Miller, and Marty J. Wolf. 2011. “Developing Artificial Agents Worthy of Trust: ‘Would You Buy a Used Car from This Artificial Agent?’” *Ethics and Information Technology* 13 (1): 17–27. doi:10.1007/s10676-010-9255-1.

Gunkel, David J. 2018. *Robot Rights*. Cambridge: MIT Press.

Heidegger, Martin. 1996. *Being and Time*. Edited by Joan Stambaugh. Albany: SUNY Press.

Ihde, Don. 1990. *Technology and the Lifeworld*. Bloomington/Minneapolis: Indiana University Press.

Johnson, Deborah G., and Keith W. Miller. 2008. “Un-Making Artificial Moral Agents.” *Ethics and Information Technology* 10 (2–3): 123–133. doi:10.1007/s10676-008-9174-6.

Johnson, Deborah G., and Mario Verdicchio. 2017. “Reframing AI Discourse.” *Minds and Machines* 27 (4). Springer Netherlands: 575–590. doi:10.1007/s11023-017-9417-6.

Johnson, Deborah G., and Mario Verdicchio. 2018. “Why Robots Should Not Be Treated like Animals.” *Ethics and Information Technology* 20 (4). Springer Netherlands: 291–301. doi:10.1007/s10676-018-9481-5.

Johnson, Deborah G., and Mario Verdicchio. 2019a. “Constructing the Meaning of Humanoid Sex Robots.” *International Journal of Social Robotics*, September. Springer Netherlands. doi:10.1007/s12369-019-00586-z.

Johnson, Deborah G., and Mario Verdicchio. 2019b. “AI, Agency and Responsibility: The VW Fraud Case and Beyond.” *AI & SOCIETY* 34 (3). Springer London: 639–647. doi:10.1007/s00146-017-0781-9.

Laukyte, Migle. 2017. “Artificial Agents among Us: Should We Recognize Them as Agents Proper?” *Ethics and Information Technology* 19 (1). Springer Netherlands. doi:10.1007/s10676-016-9411-3.

Mayor, Adrienne. 2018. *Gods and Robots: Myths, Machines, and Ancient Dreams of Technology*. Princeton; Woodstock: Princeton University Press.

O’Neill, Cathy. 2016. *Weapons of Math Destruction*. Edited by Penguin Random House.

Pagallo, Ugo. 2010. “Robotrust and Legal Responsibility.” *Knowledge, Technology & Policy* 23 (3–4): 367–379. doi:10.1007/s12130-010-9120-x.

Prescott, Tony J. 2017. “Robots Are Not Just Tools.” *Connection Science* 29 (2): 142–149. doi:10.1080/09540091.2017.1279125.

Searle, John R. 1980. “Minds, Brains, and Programs.” *The Behavioral and Brain Science* 3: 417–457.

Taddeo, Mariarosaria, and Luciano Floridi. 2011. “The Case for E-Trust.” *Ethics and Information Technology* 13 (1): 1–3. doi:10.1007/s10676-010-9263-1.

Turing, Alan M. 1950. “Computing Machinery and Intelligence.” *Mind* LIX (236): 433–460.

Wiener, Norbert. 1964. *God and Golem, Inc. A Comment on Certain Points Where Cybernetics Impinges on Religion*. Cambridge: MIT Press.

Wittgenstein, Ludwig. 1958. *Philosophical Investigations*. *Philosophical Investigations*. Oxford: Basil Blackwell.