Revista de Estudos Constitucionais, Hermenêutica e Teoria do Direito (RECHTD)
14(3):382-401, setembro-dezembro 2022
Unisinos - doi: 10.4013/rechtd.2022.143.06

Consciousness and liability of non-human intelligence

Consciência e responsabilidade da inteligência não-humana

Henrique Marcos[1]
Maastricht University, Netherlands
h.jbmarcos@maastrichtuniversity.nl

Abstract

The legal liability of non-human intelligence is a complicated matter that has concerned philosophers for a long time. Nevertheless, with the current advances in research and technology, the problem grows ever closer. This paper analyzes the legal desirability of assorting legal liability to a non-human intelligence. It argues that the question of the liability of such agents is defined, primarily, through the meeting of psychosomatic conditionals. Moreover, it poses that these conditionals are, partially, determined by consciousness. It concludes that to define whether legal liability is desirable or not one must set standards for non-human consciousness and, depending on the level of consciousness of that non-human intelligence, judge if the agent should be legally liable or not.

Keywords: Consciousness, Non-Human Consciousness, Non-Human Liability.

Resumo

A responsabilização jurídica de inteligências não-humanas é um assunto complicado que tem preocupado filósofos há muito tempo. No entanto, com os avanços atuais na pesquisa e na tecnologia, o problema se aproxima cada vez mais. Este artigo analisa a se é juridicamente desejável atribuir responsabilidade a uma inteligência não-humana. O texto argumenta que a questão da responsabilidade de tais agentes é definida, principalmente, através do encontro de condicionantes psicossomáticos. Além disso, o artigo postula que esses condicionantes são, em parte, determinados pela consciência. Conclui que, para definir se a responsabilidade é desejável ou não, deve-se estabelecer padrões de consciência não humana e, dependendo do nível de consciência dessa inteligência não humana, julgar se o agente deve ou

[1] Lecturer in the Foundations of Law Department at Maastricht University. PhD in Law at Maastricht University (UM), The Netherlands; PhD in Law at Universidade de São Paulo (USP), Brazil. Master of Sciences, Master of Laws (M.Sc. LL.M.) at the Federal University of Paraíba (UFPB). Maastricht University, Law Department, P.O. Box 616, 6200 MD, Maastricht, The Netherlands.

não ser legalmente responsável.

Palavras-chave: Consciência, Consciência Não-Humana, Responsabilidade Não-Humana.

Introduction

An elephant, an autonomous military robot, and an adult human male walk into a bar. Without any reason or planning, each of these unusual and incidental companions kills a person inside the tavern. The police arrive and find three dead bodies. A criminal lawsuit follows against each of the murderous agents. If evidence confirms our story, should the judge convict all of them? Is holding a trial for them even rational? Our common sense tells us that, in principle, there does not seem to be any reason for the man to be summarily absolved of the crime at hand; he should be prosecuted and, depending on the circumstances, be condemned for homicide. On the other extreme, unless this is an exceptional elephant with human-like consciousness, a normal pachyderm should not be put under trial. Its lack of conscious *agency* is reason enough to question even if the killing should be technically considered criminal activity. (If a trainer domesticated it to commit felonies, then the elephant was a mere instrument; the human is the one who should be prosecuted for the murder.)

Our focus finally lies on the robot. Considering that it was not being controlled by some third party while performing the violent deed, perhaps its actions were just a result of the previous programming. It was built and designed for killing; conceivably, it could have malfunctioned. In this case, the question of liability falls on its programmers or the manufacturer who made the machine. However, for the sake of argument, consider that the judge finds out that the killing robot is a recently developed autonomous unit with sufficient "artificial intelligence" (AI) to decide for itself what it should do next, whom to attack and where to target to maim or kill. Should the automaton be sentenced and sent to reprogramming or disassembly? Robot-jail? Perhaps, the answer lies in the opposite direction. Even if the machine is autonomous, should the liability still lie entirely with its human creators?

The legal liability of non-human intelligence is a matter that has concerned philosophers and legal scholars for a long time. Notwithstanding, with the exponential advance of AI technologies for autonomous vehicles and unmanned drones, researches on animal intelligence, and even the prospect of finding extraterrestrial life, the problem of defining the breadth of liability to non-humans seems to be growing ever closer. In this paper, I intend to analyze the desirability of asserting legal liability to a non-human intelligence. I hold that the question of the liability of such agents is defined, primarily, through the meeting of psychosomatic conditionals. Moreover, I argue that these conditionals are, partially, determined by identifying consciousness in non-human intelligence. However, I hold that the answer to the question of the desirability is not a binary "yes" or "no" one size fits all solution. The matter is considerably more convoluted, considering that the set of agents that meet the criteria for non-human intelligence is theoretically complex. In this sense, I believe that to

define whether legal liability is desirable or not we must set standards for non-human consciousness and, depending on the level of consciousness of that non-human intelligence, judge if the agent should be legally liable or not.

To do so, I will first present a thorough analysis of the different definitions of responsibility according to Hart's theory *(Section 2)*. Under the same theoretical reference, I will consider the conditionals for legal liability *(Section 2.2*, specifically) and correlate the responsibilities with the conditions for liability in a taxonomic rank *(Section 3)*. Thereafter, I will argue that there is no controversy for attributing causal-responsibility for non-human agents *(Section 4)* and demonstrate that role-responsibility and the liability of legal persons do not solve the issue of assigning liability for non-human intelligence *(Section 5)*. Following, I will analyze the prominence of the psychosomatic conditionals for liability *(Section 6)*, point out why we must consider the "desirability" instead of "legal possibility"*(Section 7)* and, finally, present the reasons why I hold that consciousness is critical to solve the matter of defining liability to non-human intelligence *(Section 8).*The arguments presented in this paper are not explicitly related to any individual jurisdiction or legal framework, although there is a chance that some of the grounds I mention are more closely linked to a Civil Law background. I will do my best to bring the Common-Law correlates.

Four definitions of responsibility

To define liability, I will use the highly influential set of distinctions first presented by Hart (1968, pp. 210-230)— also used by Hage (2017;2016) among others — dividing responsibility into four heads of classification: *(i)* Role-responsibility, *(ii)* causal-responsibility, *(iii)* capacity-responsibility, and *(iv)* liability-responsibility. While at first glance it may seem that only the fourth definition interests us, as it will be made clear in the following pages, an understanding of the different concepts of responsibility is essential to understand liability and my overall argument fully.

(a) Responsibilities one to three

As the name implies, *(I) role-responsibility* is the sort of responsibility related to the role the person holds. Such is the case of a captain who may be deemed *responsible* for the sinking of his ship even if s/he was not directly involved with the negligent maneuvering of the vessel. Through the same definition, the Queen's Life Guard is responsible for the protection of the Queen, parents for their child's upbringing, and a host for the well-being of his guests. The classification at hand suggests a generalization that an individual is said to be morally or legally responsible when s/he holds a distinctive role within a social organization to which specific duties are attached to in favor of others. Therefore, role-responsibility is not a special kind of responsibility, but rather *grounds* for the existence of moral or legal responsibility in an unspecified sense: S/he is said to be responsible for fulfilling these duties under the role s/he occupies.

(ii) Causal-responsibility occurs when someone or something is the causal explanation or

justification for some consequence. E.g., the coconut that fell on top of someone's head was responsible for his/her death; the bumpy road was responsible for the vehicle crash; the motor-cortex in the brain is partially responsible for the control and execution of voluntary movements of the body. In all these cases, we could replace the word "responsible" for "causes" or "produces". The causal sense of responsibility is more a question of the factual outcome of a particular chain of events than a matter of distinct analysis of intention. It is purely a statement concerned with the contribution of an entity (falling coconut) to the consequences of a specific state of affairs (death by trauma). In this sense, in the narrative presented in the introductory paragraphs of this paper, we could say that the elephant, the robot, and the man were *responsible* for the three deaths. However, arguably in the case of the human and debatably for the automaton, more than just causal-responsibility, there is the question of another more critical analysis of the innate responsibility of these agents.

(iii) Capacity-responsibility, in its turn, is generally used to assert that an individual has the capacities of reasoning, understanding, and control of conduct for his/her actions. In this sense, a person has capacity-responsibility over his actions because s/he can understand what conduct legal rules or morality requires, and, simultaneously, s/he can deliberate, reach decisions, and conform to the decisions made. Therefore, capacity-responsibility is related to a complex set of psychosomatic characteristics of an individual and, thus, may be diminished by a temporary or permanent mental illness, a bodily deficiency, sickness or wound, and, arguably, the consumption of mind-altering substances. In this way, we say that an intellectually disabled person is not responsible for a particular action, whereas if an intellectually sane individual practiced the same act, s/he would be deemed responsible, i.e., mentally able to restrain him/herself from that action. Similarly, a non-human animal, a young child, or someone who is sleepwalking may not be deemed legally or morally responsible for his/her actions on account of their lack of capacity-responsibility.

(b) Responsibility four: liability-responsibility

Lastly, *(iv) liability-responsibility* can be either moral or legal. I will focus on the general theoretical foundations for legal liability considering our present focus (thus, if "liability" appears without qualifiers, read it as "legal liability"). Nevertheless, *mutatis mutandi,* most of the fundamental aspects presented hereafter are valid for moral liability-responsibility. Generally, legal liability has a more extensive reach than its moral counterpart.(Some exceptions apply, as the infamous drowning child example tells us; while in most legal systems there is no *legal* liability for someone who refuses to shove a child's face out of a puddle in which it is drowning, there seems to be a grave *moral* liability for someone who refuses to do it while perfectly capable of doing so).[2]

The Law requires that individuals act or abstain from acting in a certain manner. One who

[2] "When lawyers are asked to offer examples of the difference between law and morality, they are very likely to say, out of ancient law school tradition, that we have no legal duty to shove a child's face out of a puddle in which it is drowning as we stroll by. The example is powerful because the moral duty the law refuses to enforce is so uncontroversial. The threat to the child is at one extreme of harm, and the effort required of us at the other extreme of cost." (Dworkin, 2011, p. 276).

infringes on a legal norm's command is, in principle, subject to liability, which, in turn, may be ordered by another norm to make compensation or punishment (sometimes both for the same action). In other words, s/he who breaks the law will be made to pay for his/her transgression. That is a commonsense and even trivial account of liability; contemporary legal systems are substantially more complicated arrangements of rules. Someone may be legally punished for what an entirely separate individual may have done. Consider a hotel-owner who may be directly liable to pay monetary compensation for damages caused by one of his employees against a lodger; or a factory-owner who may also be directly liable for an accident that maimed one of his workers even if that misfortune was caused by carelessness from another employee. In these cases of "vicarious" or even "corporate responsibility", that liability is derived not from a plain matter of causal-responsibility, but rather an inference of role-responsibility that the Law decided to turn into legal grounds for liability.

The question of legal liability for some action or harm is generally concerned with the meeting of certain conditions that are not exclusively related to psychological states, as we have seen above. These criteria that were also first presented by Hart(1968, p. 217) divide legal liability into a second-level threefold classification: *(iv-a)*Mental or psychological conditions; *(iv-b)* causal or other forms of connection between act and harm; *(iv-c)* personal relationships.

(b1) Mental or psychological conditions for liability-responsibility

In Criminal Law, one of the most frequent issues raised is whether the accused person was mentally and psychologically apt, i.e., satisfied*(iv-a) mental or psychological conditions* that fulfill the requirement that s/he had the capacity to understand what is required by the Law to do or not to do, deliberate and decide what to do, as well as to control his/her conduct in light of these decisions. Neurotypical human adults are generally assumed to have these capacities, that is why we assume that the man from our introductory plot should be put under trial and, potentially, condemned for the homicide. However, if during the presentation of proofs, the judge found out that our human adult was not in his full capacities (imaginably he was suffering from a severe schizophrenic hallucination), he may be absolved of his crimes because he should not be deemed *responsible* for his actions (a similar result could be achieved if a young child had practiced the criminal act). In this case, "responsibility" must be understood under the heading capacity-responsibility discussed above. (Notwithstanding, there is nothing that stops the judge from absolving the mentally ill man from criminal liability, but ordering his admission into a psychiatric hospital. However, that would not be strictly deemed as punishment, but rather medical treatment or even a safety precaution for society's and the man's own safety.)

Common-Law systems call the mental element of one's intention to commit a crime by action or lack of action "*mens rea"* ("guilty mind"), an overarching category. On the other hand, Civil Law systems generally divide between two main types of psychological conditions: Questions related to overall capacity as matters of capacity-responsibility, which are called matters of imputability, while questions on the presence of knowledge and intention are described within the topic of dolosity, fault or malice. In both systems, the matter of a person's criminal liability is closely related to their mental and psychological conditions, i.e., culpability.

If an individual has these capacities impaired, it is possible that s/he may not be deemed liable. However, it is also relevant to point out the existence of criminal "strict liability", i.e., liability for which the *mens rea* does not have to be proven concerning other elements of the criminal act. Therefore, in some cases the accused may not be deemed *culpable* in common usage of the term; however, they may still be considered legally liable. A contemporary theory of strict liability is called "objective imputation" *("imputación objetiva")*, where someone may be deemed criminally liable if his action creates a risk that is at least potentially within the sphere of the action performed. Be it as it may, as stated, in general, the *(iv-a) mental or psychological conditions* category of *(iv)liability-responsibility* is closely related to *(iii)capacity-responsibility*.

(b2) Causal or other forms of connection with harm for liability-responsibility

However, matters of liability-responsibility are not limited to psychological or mental conditions. Questions of causal nexus between the agent's act and the harm done to the victim are also significant. While the issue of link between the person who willfully pulls the trigger firing a pistol against his foe and his foe's death may seem trivial, in some cases a person accused of a crime may not be deemed liable for the injury done if there is some other form of connection or if the relation between the defendant and the harm is too distant. Should the salesperson be also liable for murder because he sold the weapon of the crime to the person who fired it? The answer varies according to the kind of liability (tort, criminal) and the legal system. It seems clear that these matters of *(iv-b)causal or other forms of connection with harm* are a category of *(iv) liability-responsibility* that seems to correlate with *(ii) causal-responsibility.*

(b3) Personal relationships

(iv-c) Personal relationships or relationship with the agent as the final category for liability-responsibility also has some proximity to the above-considered category *("iv-b").* Generally, in tort and criminal law, a minimum required condition for punishment is that the person to be punished should have him/herself done what the law forbids. Thus, a minutest causal connection is required between who committed the crime and who is liable for it. At first, it does not seem reasonable for "Alfred" to be punished for the wrong that "Bertrand" has done. However, if these two individuals share some prior relation, perhaps the punishment of the former for the actions of the latter agent may not seem to be so absurd. Consider the examples of the hotel-owner and the factory-owner above *(Section 2.2)*; they had a previous relationship to their employees–– they were responsible for their supervision, i.e., had the right, ability or duty to control them––and, thus, could be considered liable for their actions. In this case, it seems that the *(iv-c) personal relationship* category of *(iv)liability-responsibility* shares some common ground with *(i) role-responsibility* considered before.

Taxonomy of responsibilities and conditions for liability

As mentioned above *(Section 2)*, to properly understand liability, it was fundamental to

analyze the overall senses of responsibility. As we have examined, there are diverse types of criteria for liability-responsibility. Arguably, one of the most noticeable parameters is mental or psychological conditions. However, there are also causal or other connections between the agent and the harm done as well as matters of inter-personal relation between the parties involved in the scheme of liability.

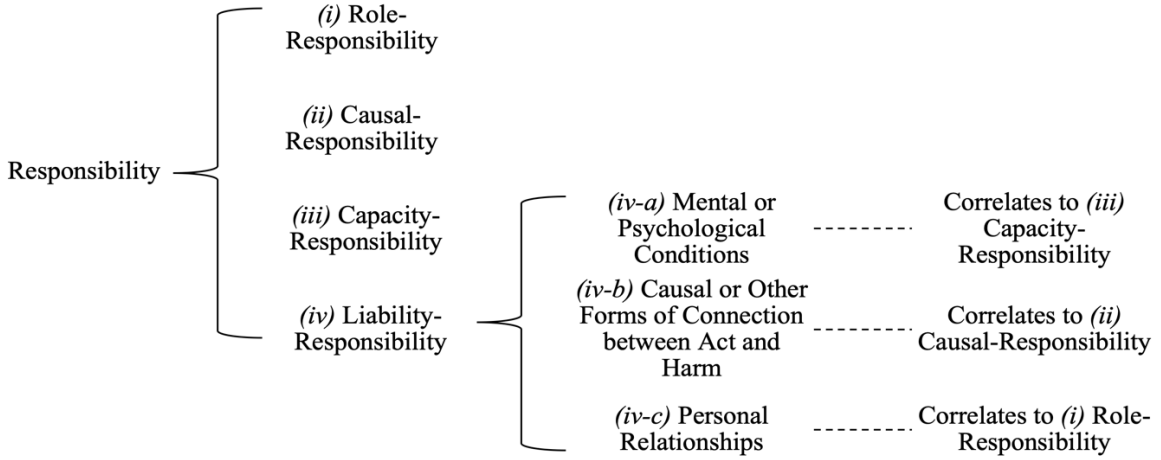

Figure 1: Taxonomic Rank of Responsibility and Conditions for Liability
Source: The Author

There are multiple conditions attributed to liability and varying meanings of responsibility. Considering that the sense of liability is derived from the content of the different concepts of responsibility which, between themselves, have no evident common features, our analysis suggests that the usage of the word "responsibility" to describe all of the disparate genders is, as Wittgenstein would have called it, *a family resemblance* (1986, p. 32).Therefore, we use the word "responsibility" to describe varying kinds of responsibility that are dissimilar from each other and, maybe, not even have a consubstantiating characteristic in common. These are different language games with altering rules for each. Nevertheless, the loose architecture of meaning indicates the conventional nature of human language. In this sense, family similarity serves as an analogy for the means of connecting specific uses of the same word ("responsibility").

Hence, "responsibility", can be understood as the responsibility placed upon an individual on account of the role s/he occupies *("i")*; the responsibility as the causal link between agent and result *("ii")*; or responsibility as the psychosomatic conditions that allow one of being capable of reasoning, understanding, and controlling the conduct for one's actions*("iii")*. Likewise, the legal iteration of responsibility, i.e., liability-responsibility *("iv")* is verified when the legal criteria are met. The varying definitions of responsibility derive the content of these conditions and their respective correlations to the requirements of liability, as stated above. Therefore, at least partially, the definition of liability-responsibility depends on the substantive meaning of the other three responsibilities.

Lack of controversy surrounding causal-responsibility

It does not seem that every one of the responsibilities that we analyzed pose difficulties on being attributed to non-humans. As mentioned above *(Section*2.1*)* causal-responsibility can be attributed to anything, even a coconut falling from its tree could be said to be responsible for someone's injury or death. Therefore, it does not seem to present any complexity to say that a non-human intelligence (such as an AI) can be (causally) responsible for some harm (e.g., a self-driving vehicle was *responsible* for crashing into a crowd). In this sense, considering the correlate between *"ii"* and *"iv-b"* (i.e., causal-responsibility and causal connection between act and harm as a condition for liability), it seems plausible to assert that at least this preliminary criterion for liability could be met with no significant controversy.

I draw attention to the fact that the assertion made in the previous paragraph is not sufficient to conclude *tout court* that non-human intelligence is liable. Most legal systems require that more than just the causal condition of liability-responsibility be met. Otherwise, the coconut that fell on someone's head or the animal trained to attack on command would be considered legally liable for their actions. Causal-responsibility serves as a minimum requirement. However, by itself, it is not enough to lay on to someone or something legal liability. My assertion is merely that*(ii)* causal-responsibility and, thus, *(iv-b)* causal connection between act and harm can be attributed to almost any object. Consequently, our attention must focus on responsibilities "*i"* and "*iii"*.

Role-responsibility and the liability of legal persons

Regarding *(i)* role-responsibility and its counterpart *(iv-c)* personal relationships, they are a quality attributed to a previous set of responsibilities that increases the scope of a person's liabilities not only to their own actions but to the actions of others. They are grounds for a new level of responsibility and liability. The legal justification for this kind of liability seems quite straightforward. In one way, it is justified because it induces the individual in the position of power to be even more vigilant in fulfilling his/her supervisory duties. Simultaneously, it turns restitution for damages more efficient. It seems intuitive that a hotel-owner or factory-owner (mentioned in *Sections 2.2* and *2.2.3)* has deeper pockets than one of their workers (especially considering that, contemporarily, in most cases, large business-owners are legal persons).

The moral grounds for holding role-responsibility – which influences the legal norm for this kind of liability – is that the party to whom the responsibility is attributed to is, in a *folk psychology* conceptual scheme, seen as capable of being ascribed that responsibility.[3] Folk psychology is a handy legal tool. Mainly because the regulator needs to create a set of legal

[3] By folk psychology I mean a set of fundamental capacities that enable humans to describe their behavior and the behavior of others, being able to explain this behavior, also predicting and anticipating further actions as well as producing generalizations on human behavior. Thus, opposed to a scientific explanation of human behavior, folk psychology is the commonsense human capacity to explain and predict the behavior and mental state of other people (STICH; RAVENSCROFT, 1994).

fictions that allow the law to be as far-reaching and efficient as possible. Consider the case of a legal person such as a corporation or a state. Technical analysis of cognitive science tells us that it is ludicrous to attribute free will, malice, intention, negligence, or any other human state to a legal person. However, the Law creates these fictions for regulatory purposes. Through narratives, it becomes possible to hold a legal person liable for civil and even criminal misdeeds that are attributed directly to them instead of having to find the specific individual or group of human associates that were causally-responsible for the transgression. The justification is pretty similar as the one given for the liability through personal-relationships and role-responsibility: It is far easier to identify the legal person whereas in some cases isolating the individuals that committed the act is not so straightforward; similarly, legal persons have more liquidity, making it easier to get restitution in the case of damages.[4]

However, the matter of legal (and moral) liability for legal persons is not the same as arguing for the liability of non-human agents. The actions of legal persons are always traceable to the actions of a human individual. Even if the process of identification is complicated or even impossible in practice, in all cases the acts were practiced by a person or a collective, but always human persons. Including the case of a legal person who has for shareholders other legal persons. If we dig deep enough, on the decisive end of things, we will always find human beings. Therefore, the liability of a legal person does not pose a difficulty to the commonsense conception of agency. The fact that they are hiding under one or several curtains of red tape does not change the fact that, in the end, legal persons are simply humans in bureaucratic disguise (Brozek and Jakubiec, 2017).

It does not seem that the matter of liability of non-human agents will be solved through the analysis of the liability of legal persons. They are different monsters altogether. Moreover, causal-responsibility and its correlate, as we have seen *(Section4)*, do not pose any controversy (anything can be causally responsible for something else). Role-responsibility and the personal relationship condition for liability may seem at first to be an independent matter, but, in reality, it is merely a liability by proxy: Someone (or something) is liable for someone else's behavior. The responsible party is given that responsibility because s/he is seen to hold a distinctive role within a social organization. It is not a special kind of responsibility, then. It is a moral or legal ground for the attribution of responsibility and, likewise, liability through personal relationships. As one of its kinds, vicarious liability is attributed to a "third party" that is perceived as having the right, ability or duty to control its subordinate. The "superior" in the case of vicarious liability becomes liable for the actions of the violator. However, this is only sustainable because the now responsible party is (at least through folk psychology) seen to be *capable* of being subject to that responsibility. It seems unlikely for a legal institute to prescribe a case of vicarious liability to a *prima facie* incapable category of subjects, such as stating that newborn infants should be liable for their parent's failure to pay the invoice on medical charges or that a laptop computer should be liable for its user's online criminal activity.

[4] Brozek and Jakubiec (2017) present a similar and very compelling argument.

Prominence of the psychosomatic conditionals for liability

Our previous analysis and the inferences we made along the way keep sending us back to *(iii)capacity-responsibility* and the correlate *(iv-a) mental or psychological conditions* for liability. It seems to show that, as Hart had pointed out(1968, p. 221), the psychosomatic elements are the most prominent for defining liability. According to the author, this is made clear through what he considers a Cartesian depiction of the agent:

> If we conceive of a person as an embodied mind and will, we may draw a distinction between two questions concerning the conditions of liability and punishment. The first question is what general types of outer conduct *(actus reus)* or what sorts of harm are required for liability? The second question is how closely connected with such conduct or such harm must the embodied mind or will of an individual person be to render him liable to punishment?(HART, 1968, p. 221)

Hart's second question interests us the most at this moment. To what extent must the embodied consciousness be the author of the conduct or the harm in order to turn it liable? Is it enough for the body to perform certain movements or is it required that it possess a certain capacity of control with intention? The exact answer to these questions is given by each individual legal system with its particular rules that define the reach of liability (e.g., should the age of adulthood for criminal liability be defined at sixteen, eighteen or twenty-one years?). Nonetheless, these are all inquiries on whether the accused person is mentally and psychologically capable of understanding what is required by law, deliberate and decide what to do, as well as control his/her conduct in light of this decision.

One of the main reasons why it seems incongruous to put the elephant from the opening paragraphs under trial and, in case of conviction, condemn it to a strict sense legal punishment is because, for most people nowadays, that animal is not capable of understanding what the Law requires from it. Thus, it is not able to deliberate, contemplating what the legal norm demands, intentionally comply with the Law. On the other hand, if an exceptional elephant came about, being able to communicate with humans through trunk gestures akin to sign language and fully understand the complex world around it, including what the human legislature requires from its subjects –– if all of these capabilities were proved to be true beyond any doubt –– perhaps our outlook would change towards this remarkable animal. Imaginably, in this farfetched case, people would not think that this gifted elephant should be spared so hastily from trial.

What is the difference between elephants one and two? I believe that it is, most of all, a disparity between psychosomatic capacity that is, hypothetically (and fictitiously), present in elephant number two while absent from elephant one (and every other elephant known to have existed). If we replace "elephant" for "AI", "extraterrestrial intelligent life form" or only "non-human intelligence" it then seems clear that, if any non-human intelligence came about, the key element to define whether they should be considered liable for their actions is the psychosomatic capacity for reasoning and deliberation over their actions. That seems to be the fundamental requirement for liability, as we have seen, and, therefore, should also be the

parameter for the liability of non-human intelligence. In this sense, the matter of the desirability of liability of these non-human agents necessarily passes through the definition of their psychosomatic capacities. If these are absent on a given agent, it is evident that it should not be considered liable.

Desirability instead of "legal possibility"

Before we move on further, I must comment on some reservations that I believe are in order when we talk about folk psychology and the specific matter of popular opinion on whether non-humans should be liable. The elephant as the animal of choice for our example is not without reason. In 1916, Mary, an Asiatic elephant who was forced to work as a circus performer in the United States, was hanged to death. She was accused and condemned for killing one of her trainers after he prodded her ear with a hook. The accounts vary, and the real story is mixed with sensationalist tabloid depictions of violence (she was even called "Murderous Mary" in some representations). Also, it is not clear if Mary was actually put under some sort of judicial trial or if her captors simply slew her. However, what I believe that contemporarily would be considered an irrational settlement for what happened (and an embarrassing depiction of what humankind is capable of), was considered at that time an adequate arrangement by some. Unfortunately, similar stories of elephants being executed after killing their captors are not uncommon; being popular circus animals, these animals were constantly subject to stressful situations and sometimes, defending themselves, end up injuring or killing their handlers. Other than Mary, the elephants Topsy and Ziggy also had tragic faiths; after what was considered a foul offense, they were each condemned to life imprisonment or death.

Historically, it has not been unheard of for the Law to treat animals, plants, and other inanimate objects in the same way as humans and, in particular, be punished for their deeds. As Kelsen (1949, pp. 3-4) recalls, in Ancient Greece, there was a special court whose function was to prosecute inanimate things such as a spear used as a murder weapon. This could be considered a residue of the animism of the primitive man who was in the habit of endowing anything with humanlike features such as a "soul". These entities were, therefore, turned into *agents* by the Law because people–– and the regulator –– saw them as being capable of being considered legal subjects. Kelsen (1949) indicates that, in his view, the "civilized peoples", i.e., the contemporary society, have long overgrown this characteristic of attributing *agency* to entities that are devoid of any real capacity of having intention. Yet, the dreadful account of what happened to Maryat the beginning of the 20th century seems to weaken this optimistic diagnosis. Bearing this in mind, to put it simply, folk psychology and popular opinion may be useful as an instrument to point out the way for what the Law should do. However, it should be far from the only guide for lawmakers.

It may come as a surprise to non-lawyers, but in sensible terms it makes little sense to talk about "legal possibility". As the aforementioned examples tells us, theoretically, anything can be put into the law (the case of the melee weapon being condemned by a tribunal seems to be

a blatant example of how far the regulator can go). Thus, anyone or anything can be made into a legal agent. This is precisely what Kelsen's legal theory tells us about what the Law *is* (and not what it *should* be): the Law can be anything and have any content imaginable (Kelsen, 1967). Generally, more restrained "soft positivists"[5], such as Hart, hold that legal systems are not so unhindered as Kelsen would describe them. Hart's doctrine states that the legal rule of recognition may incorporate the conformity with moral principles or substantive values as criteria for the validity of the legal norm (Hart, 1994, p. 250). Even then, the Law still has considerable leeway for regulation. In this sense, it seems fruitless to talk about the legal *possibility* of conferring liability upon non-human intelligence. It is far more productive to talk about what would be *desirable* for the Law to do. In other words, considering its overarching purposes, *should* the Law consider certain non-human intelligence liable? That inquiry seems more prolific.

The question of consciousness for defining liability

According to our inferences above (especially *Section 6),* to answer the question of desirability we must consider what is expected from the Law in the specific matter of allotting liability. Fortunately, we have already contemplated this matter while considering the different conditionals for liability and their correlates in the different kinds of responsibilities (*Section 2).* Those are the criteria for defining liability for *human* agents, however. Now, the matter at hand is to judge whether the conditions as mentioned earlier should also be assigned to *non-humans,* or if we should propose different criteria. Yet, in light of what we have examined, I see no justification so far for devising different conditions for non-humans. I have anticipated my hypothesis. We shall see if it holds.

(a) Overcoming speciesism

Similar to the mentally brilliant elephant case seen above *(Section 6)*, let us consider the following experiment. By chance, you find a long-lost friend from your childhood. You both start catching up on what happened in your lives throughout all these years, and he tells you that he is currently working in a multinational law firm. It is an exciting job, and the pay is good, but the working environment has been tense lately. Recently, two of his coworkers got into a heated argument over their favorite movie director. The discussion ended quickly. One of them drew a loaded gun and fired it at the other. If we stop with the experiment now and I ask your judgment about what you have been told, your preliminary conclusion would probably be that the person who killed his/her colleague is in the wrong here. One of your friend's colleagues killed the other over a petty quarrel; no reasonable person would think that it is sensible to murder because of preference in cinema. Moreover, if we are to believe that the story is a perfect factual depiction of what happened, we can infer that all signs point that the killer had

[5] Soft positivists, according to Dworkin (1986), are positivists that allow for morality to figure among the tests of the validity of law.

a clear causal connection between action (pulling the trigger) and harm (death by a gunshot wound). Also, if no evidence comes to prove the contrary, it seems that an average lawyer is mentally capable enough to understand his/her actions, deliberate and contain him/herself over a disagreement on which director is the best in the movie business. (In the narrative, there is no need to analyze their personal relationships as a condition for liability. The killer and the victim were at the same level in the firm's hierarchy. Role-responsibility is not required in this case, given that the criminal has a direct causal link to the harm done. Perhaps the family of the deceased may wish to file for damages against the legal person or the partners of the firm, but that is unimportant right now given that, in principle, it does not withdraw personal liability from the murderer.) In all, the conditions for liability were met.

We proceed with our experiment and now you ask your friend: "So, what happened to your former coworker, is s/he in jail?". "No", he responds. It turns out that the murderous colleague was a bioengineered humanoid placed in the firm to test if its social skills were enough for him to pose as a human. The story became a lot more interesting; your friend should have started his account emphasizing the android-aspect. Inquisitively, you start asking about this technological wonder. Does it look human? Does it act like a person? Is the AI capable of passing the Turing test?[6]The answers are all affirmative. In fact, throughout your questioning, it seems that this humanoid is virtually identical to a human, including cerebral functioning. The only difference is that, instead of being naturally born, it was synthetically engineered in a laboratory. Hence, on account of not being strictly human, the android was summarily absolved of all criminal charges. Does that seem like a desirable iteration of the Law? I do not suppose it is.

There does not seem to be any reason for there to be a strict differentiation on the sole excuse that this agent is not a human. If the android fulfills all the criteria for liability, except for a debatable underlying condition of being "human", I believe that liability is due if a thorough inquiry attests that this organism is indeed mentally capable. The prominence of the psychosomatic conditions was already mentioned above *(Section 6).* As I have pointed out on that moment of our investigation, the *(iii)* capacity-responsibility and its correlate *(iv-a*) mental or psychological conditions for liability are the outlining factors and the foremost steps for us to start inquiring if a particular entity (human or not) could be considered the *agent* of a specific act. When we are analyzing human conduct, most legal systems assume that adults are mentally capable unless proven contrary. If we are faced with a diverse set of non-human agents, however, it does not seem that we should assume so quickly that they are psychosomatically capable. On the other hand, the opposite solution that proposes that we should, exclusively, only hold humans liable as an *a priori* and uncontestable metaphysical truth seems unthoughtful. I am unable to notice any sound justification that creates an unsurpassable wall between the categories of individuals that are human and non-human. Under the circumstances of the friend's tale of the murderous humanoid, it seems that the most rational course of action is to ascertain how conscious that organism is of its actions. If it turns out that we have

[6] The Turing test is a test of an AI's ability to exhibit intelligent behavior that is indistinguishable from that of a human (Turing, 1950).

uncontroversial proof that the android is as psychosomatically capable as any human adult, why should it be left off the hook? It seems that any argument for an absolute distinction between humans and non-humans is plain speciesism.

Perhaps you might consider that my experiment was too extreme.[7] Maybe the humanoid organism is far too human not to be considered human altogether. It is a wacky exercise, and, possibly, the challenges the Law will face in the foreseeable future will not be even close to the borderline android-human experiment. Most probably, we will be confronted with questions regarding the liability of neural networks, autonomous vehicles, and military drones, as well as ethical issues over primates and other non-human animals. I agree entirely. It is a bizarre experiment, indeed. But it is useful precisely because it is extreme. We have confronted this imaginary case precisely because this humanoid, as a marginal organism, shows us that the dividing line between human and non-human liability is artificial. When faced with this agent possessing human-like psychosomatic capacities, we are confident that it is clearly unsound to summarily absolve it of any potential wrongdoings, no matter how wicked, on the ingenuous account that it is not human. Realizing that any idea of a definite dividing line was merely defined on account of our experience –– as inhabitants of the early 21st Century, we have never met any non-human with human-like mental capacity –– we are ready to face the question on which standard should be used to define whether a non-human could be considered an agent for the effect of liability.

(b) Psychosomatic conditionals as consciousness

To say that that particular humanoid should be liable is not the same as declaring *tout court* that any non-human should be liable. We have already gone through that (*Sections 6* and *7)*, and I concluded that the mental or psychological conditions are the prominent criteria for defining liability. Nevertheless, how are we to point out how these psychosomatic conditions may be perceived in non-human agents? As I have pointed out, folk psychology and popular opinion cannot be used as the only guide to solving this problem. It seems that this matter must necessarily be dealt with by a thorough analysis of consciousness. In other words, when dealing with the issue of non-human liability, the test of their psychosomatic capabilities should go through a review of their consciousness potential. The relation between liability and the mental state of consciousness has already been examined by Moore (1980, p. 1583) while considering the relations between the conscious and the unconscious self, specifically the effects of unconscious action in liability. Through Moore's analysis of the varying definitions for responsibility we are presented with a somewhat different scheme that divides responsibility into "causal responsibility", "answer-ability", "culpability" and "liability". However, a closer reading shows us that, in fact, these diverse responsibilities are conditions for the ascription of liability similar to the Hartian definitions of responsibility and conditions for liability we have examined before *(Sections 2* and *3)*. Moore's "causal responsibility" correlates to *(iv-b)* causal or other forms of connection between act and harm; "answer-ability" can be both a matter of

[7] Dworkin (2011, p. 283) presents a similar argument in defense of "crazy cases".

(iv-a) mental or psychological conditions or *(iv-c)* personal relationships. For what interests us the most, Moore's usage of "culpability" and "liability" relate to *(iv-a)* mental or psychological conditions; "liability" being the final condition, when reasons for the imposition of legal sanctions are met, the person is said to be "liable".

In this regard, while analyzing the underlying metaphysical moral basis for criminal liability, Moore provides us a set of principles for liability under which fault is properly ascribed to persons for their behavior. I believe that these principles serve as a fitting framework for what is the adequate psychosomatic conditions for liability. Hence, fault is properly ascribed when *(p_1)* a being is sufficiently accountable for his actions that he may be counted an agent; (p_2) the legal norms may fairly obligate such agents; *(p_3)* an act is done, a harm is caused, and with which mental states culpability is to be found; *(p_4)*it is ascribed while considering circumstances that could serve as a justification or an excuse for having caused that harm (Moore, 1985, p. 12). Analogous to the conclusion I have arrived concerning the prominence of the psychosomatic conditions for liability (*Section 6)*, while considering the effects of culpability, Moore states that if there are actions or intentions *present* when thought they were *absent* or these conditions were *absent* whereas we prove they are *present*, there will be some direct consequence over the ultimate question of the liability of that agent (Moore, 1980, p. 1586-1587). It seems clear to me that these parameters of liability have a close correspondent to the different types of consciousness.

(b1) Varying types of consciousness

Consciousness, however, is not an easy concept to define. Searle (1990, p. 635) says that, by consciousness, he means the subjective state of awareness and sentience that begins when we wake up in the morning and continues through the day while we are awake until we fall into a dreamless sleep, coma or death.

Nevertheless, for the purposes herein, I will use Pessoa Jr.'s distinction of consciousness into four categories (Pessoa Jr., 2019, p. 21-22). *(c_1)* *Sentience-consciousness*, can be understood as the phenomenal consciousness of the passive subjective experience, i.e., what it is like to feel phenomenological qualities such as the warmness of heat, the gradient of color, or the stinging of a pain. Sentience may come in varying levels from the "unconscious" perception to the "self-conscious" and reasoned experience in which one is attentive to what is felt. *(c_2)* *Reasoning-consciousness*, in its turn, is the ability to create mental representations and planning, which generally involves language. Reasoning may also vary from "unconscious reasoning" to a purposive attitude when one intentionally engages in reasoning. *(c_3)* *Deliberation-consciousness*, is the kind of consciousness connected to action. Block (1995, p. 229) uses the term *access consciousness* for the state of the conscious availability to interact with other states and of the access that the conscient self has to its content. Deliberation-Consciousness allows one to have access to mental representation for practical reasoning and rational guiding of actions. Finally, *(c_4)* *Introspection-Consciousness* is the set of phenomenal consciousness and reasoning that leads to a higher state of consciousness where one is introspectively aware that s/he is in that state, i.e., self-consciousness which may lead to the concept of self-hood. At this

time, I am unable to indicate which of these concepts for consciousness have primacy on the definition of liability. It seems that all of them (especially the latter three concepts) have a close connection to the required psychosomatic conditionals for liability.

(b2) Non-human consciousness?

Should we be discussing non-human consciousness altogether? Both the exceptionally intelligent elephant and the murderous humanoid are entirely fictitious cases (respectively, *Sections 6* and *8.1*). I have never been introduced to a non-human whom I judged to have human-like consciousness. Also, I do not know of anyone in their right mind who has. It seems safe to say that contemporarily there is no authentic non-human intelligence on the same level as an average human being. There is even a chance that we will never be able to develop *real* AI. Very convincingly, Searle argues that programs are defined by purely formal processes, while there is more to having a mind than having syntactical rules. Minds deal with meaning; they are semantical in the sense that they have more than formal structures; they have content. Even considering that these machines will go through remarkable technological progress across the years to come, computer processes are still syntactical (Searle, 2003, p. 28-41).

Nevertheless, Searle also seems to argue that, while syntax is not sufficient for semantics and no computer program by itself is adequate to give a system a mind, there is nothing that prevents an alternative technological breakout that allows a man-build artifact to produce mental states equivalent to human mental states. That artificial mind, however, will not be a computer (Searle, 2003). There is still a chance that such a technological prowess will never be achieved by humanity. Besides, there is a reasonable possibility that we are all alone in the universe, we will never find intelligent life elsewhere, and finally (however unlikely) we may prove that non-human animals are entirely "unconscious automata". Even then, I still hold that –– theoretically – it is still practical to discuss non-human consciousness and, consequently, non-human liability.

Conversely, functionalism may be right, i.e., it may be true that mental processes are just brain processes and the replication of these processes through the appropriate means may be able to replicate conscious states. In this sense, by functionalism I mean the set of doctrines that hold that what makes something a mental state of a certain kind is not dependent on its internal constitution (its substance), rather on the role it plays, i.e., its function within the system it is a part of (Schwartz, Begley, 2009). The far-reaching physical similarities we can observe between humans and other primates may be indicative that at least some non-human animal scan achieve a higher level of consciousness beyond basic sentience. Moreover, even if these similarities are not present in all organisms, i.e., even if they are incapable of sharing identical mental states on account of unsurpassable physiological differences, some of them may be able to share a higher-level property indeed. Hence, the terms for mental states may be understood as designators that denote the same items (the higher-level role properties) (Levin, 2018). For what it is worth, the 2012 "Cambridge Declaration on Consciousness" (Panksepp et al., 2012) seems to recognize this possibility regarding non-human animals, stating that the lack of a neo cortex does not appear to impede an organism from experiencing what can be called *affective*

states. "Convergent evidence indicates that non-human animals have the neuroanatomical, neurochemical, and neurophysiological substrates of conscious states along with the capacity to exhibit intentional behaviors" (Panksepp et al., 2012). Accordingly, humans may be not the only species on the planet capable of processing their neurological substrates to generate consciousness.

(b3) Levels of consciousness

If the assumptions above are correct and we end up proving that some animals or even biological beings outside the Animalia kingdom are conscient to some level. Or if we find some sort of extraterrestrial intelligent life (or if they find us). Or, finally, if we develop some machine that is capable of semantical content capable of being considered real AI. Would that result in automatic human-like liability for these agents in potential? I believe it should not. The psychosomatic conditionals for liability – which, as I have posed, should be considered intrinsically connected with consciousness – are, arguably, not the same for every organism. Similar to the differentiation that most legal systems have between minors and adults, it seems more reason able to establish a distinction for liability based on the level of consciousness of non-human intelligence. Thus, even if we end up proving that primates or cephalopods have functioning higher-level consciousness, it seems safe to conclude that they should not be considered legally liable for their actions. Likewise, if our technological advancements allow us to create AI eventually, these machines may have varying levels of consciousness between each other. Maybe an AI put in charge of running a nuclear power plant or a global database will have a higher level of consciousness than an AI developed to attend domestic chores. At this period in time, however, a specific grade for the setting of standards seems closer to speculation. Nevertheless, considering the connection between consciousness and liability I have proposed, it seems right that the increase in the former should proportionally affect the latter.

Among other proposals, Tononi's "Information Integration Theory" (IIT) (Tononi, 2004) poses that the integration of information is one of the most substantive functions of consciousness, proclaiming as far as information integration could be considered sufficient for consciousness regardless of the substrate in which it is exercised (not necessarily biological, thus). Therefore, IIT, defines consciousness as integrated information, and that its *quality* is given by the informational relationships generated by a complex relationship of elements (Tononi, 2008, p. 217). In this sense, consciousness could be reduced to a purely information-theoretic property of systems represented by the letter "φ" or simply "*Phi*". This coefficient, in turn, could be mathematically measured to indicate not only the mere information in the parts of a system, but the information contained in the organization of that system itself (Van Gulick, 2018). This measure of consciousness varies in quantities and qualities through several degrees so that even a simple system could be conscious to some point, given that the level of consciousness is determined by the totality of informational relations within the integrated set (Van Gulick, 2018). Some initial experiments have conducted computational tests estimating Tononi's Phi coefficient to measure the integrated information within the "OpenCog" cognitive architectures while reading short documents and guiding a robot in carrying out dialogue

interaction. The research showed that a preliminary comparison of the variation of Phi with the behavior of this cognitive system has shown sensible patterns (Iklé et al., 2019).

At this time, it is not clear whether the Phi measure is entirely satisfactory. However, it seems to demonstrate that the use of instruments to measure the level of consciousness of an agent, especially a non-human intelligence, is technically possible. It seems that through these measurements, one could define the level of consciousness of a given agent and, thus, prospectively or retrospectively, ascertain the level of liability that should be assorted to that agent. Nevertheless, it seems clear that the matter is more intricate than a pure "consciousness" correlates (or does not correlate) to liability. A *certain level* of consciousness is necessary to fulfill the conditionals for liability. In other words, there should be standards within the levels of consciousness that correspondingly affect liability.

Conclusions and future research

Conclusion 1: To some extent, in Hart's theory of responsibility, the definition of *(iv)* liability-responsibility depends on the substantive meaning of the other three responsibilities. The content of each condition for liability is derived by the varying definitions of responsibility and their respective correlations to the criteria of liability, as stated above *(Sections 2 and 3). Conclusion 2*: Considering the correlation between *(ii)* causal-responsibility and *(iv-b)* causal connection between act and harm as a condition for liability, these criteria can be met by a non-human intelligence with no significant controversy. Anything or anyone can be causally responsible for some act *(Section 4). Conclusion 3*: The matter of legal liability for legal persons is not the same as the argument over the liability of non-human agents given that, in the former, the actions of the are always traceable to the efforts of a human individual *(Section 5).Conclusion 4*: The crucial element to define whether a non-human agent should be considered liable is the psychosomatic-capacity over their actions understood under the heading *(iv-a)* mental or psychological conditions for liability *(Section 6). Conclusion 5:* There is no reason for a strict differentiation between human and non-human on the sole argument that one agent is or is not a human; the *(iii)* capacity-responsibility and its correlate *(iv-a)* mental or psychological conditions for liability are the outlining factors and the foremost steps for us to start inquiring if a certain entity – human or not – could be considered the agent of a certain ac*t (Section 8.1). Conclusion 6*: The parameters of liability, especially *(iv-a*) mental or psychological conditions, have a close correspondent to the different types of consciousness*;* therefore, the increase in consciousness should proportionally affect the degree of liability of a given agent *(Section 8.2).*

For the future, further research is necessary to shed more light on the specific correlation between the different concepts of consciousness and which of them has prevalence over the rest on the matter of defining liability according to the outline presented in this paper. Likewise, it is not clear whether the specific Phi coefficient is appropriate for the purposes of determining liability. It seems advantageous to research whether this measurement technique is fitting (or not) and for what reasons, as well as considering different types of instruments.

References

BLOCK, N. 1995. On a Confusion About a Function of Consciousness. *Behavioral and Brain Sciences*, **18**(2):227–247.

BROŻEK, B.; JAKUBIEC, M. 2017. On the Legal Responsibility of Autonomous Machines. *Artificial Intelligence and Law*, **25**:1-12.

DWORKIN, R. 1986. *Law's Empire*. Cambridge, Belknap Press of Harvard University Press.

_____. 2011. *Justice for Hedgehogs*. Cambridge, Belknap Press of Harvard University Press.

HAGE, J. *Should Autonomous Agents Be Liable for What They Do?* Rochester, NY: Social Science Research Network, 1 dez. 2016. Disponível em: <https://papers.ssrn.com/abstract=2885488>. Acesso em: 4 dez. 2019.

_____. 2017. Theoretical Foundations for the Responsibility of Autonomous Agents. *Artificial Intelligence and Law*, **25**(3):255-271.

HART, H. L. A. 1968. *Punishment and Responsibility*. 2. ed. Oxford, Oxford University Press.

_____. 1994. *The Concept of Law*. 2. ed. Oxford, Clarendon Press.

IKLÉ, M. et al. 2019. *Using Tononi Phi to Measure Consciousness of a Cognitive System While Reading and Conversing*. AAAI Spring Symposium: Towards Conscious AI Systems, 1-6.

KELSEN, H. 1949. *General Theory of Law and State*. Trad. Anders Wedberg. Cambridge, Harvard University Press.

_____. 1967. *Pure Theory of Law*. Trad. Max Knight. Berkeley, University of California Press.

LEVIN, J. 2018. Functionalism. *In*: E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Fall 2018 ed. [s.l.] Metaphysics Research Lab, Stanford University.

MOORE, M. S. 1980. Responsibility and the Unconscious. *Southern California Law Review*, **53**(6):1563-1678.

_____. 1985. The Moral and the Metaphysical Sources of the Criminal Law. *Nomos*, Criminal Justice, **27**:11-51.

PANKSEPP, J. et al. (eds.). 2012. *The Cambridge Declaration on Consciousness*. Disponível em: <http://fcmconference.org/img/CambridgeDeclarationOnConsciousness.pdf>. Acesso em: 20 dez. 2019

PESSOA JR., O. F. 2019. *Qualitative Sensations: What is the Nature of Subjective Sense Impressions?* Website. Disponível em: <http://opessoa.fflch.usp.br/sites/opessoa.fflch.usp.br/files/MindBrain-19-Ch03.pdf>. Acesso em: 19 dez. 2019.

SCHWARTZ, J. M.; BEGLEY, S. 2009. *The Mind and the Brain*. New York, Regan Books.

SEARLE, J. 1990. Who is Computing with the Brain? *Behavioral and Brain Sciences*, **13**(4):632–642.

_____. 2003. *Minds, Brains and Science*. 13 ed. Cambridge, Harvard University Press.

STICH, S.; RAVENSCROFT, I. 1994. What is Folk Psychology? *Cognition*, **50**(1):447-468.

TONONI, G. 2004. An Information Integration Theory of Consciousness. *BMC Neuroscience*, **5**(1):42.

_____. 2008. Consciousness as Integrated Information: A Provisional Manifesto. *The Biological Bulletin*, **215**(3):216-242.

TURING, A. M. 1950. Computing Machinery and Intelligence. *Mind*, **LIX**(236):433-460.
VAN GULICK, R. 2018. Consciousness. *In*: E. N. ZALTA (Ed.). *The Stanford Encyclopedia of Philosophy*. Spring 2018 ed. [s.l.] Metaphysics Research Lab, Stanford University.
WITTGENSTEIN, L. 1986. *Philosophical Investigations*. Trad. G. E. M. Asconmbe. 3. ed. Oxford, Basil Blackwell.

Submetido: 02/03/2020
Aceito: 25/07/2023