AI & SOCIETY (2021) 36:797–805
https://doi.org/10.1007/s00146-020-01014-x

OPEN FORUM

AI and Spinoza: a review of law's conceptual treatment of Lethal Autonomous

Moa De Lucia Dahlbeck[1]

Received: 2 April 2020 / Accepted: 2 July 2020 / Published online: 10 July 2020

Abstract
In this article I will argue that the philosophy of Benedict Spinoza (1632–1677) may assist us in coming to terms with some of the conceptual challenges that the phenomenon of Artificial Intelligence (AI) poses on law and legal thought. I will pursue this argument in three steps. First, I will suggest that Spinoza's philosophy of the mind and knowledge may function as an analytical tool for making sense of the prevailing conception of AI within the legal discourse on Lethal Autonomous Weapons Systems (LAWS). Then, I will continue the argument with the aid of Spinoza's political philosophy which partly complicates the picture as it seems to disqualify a normative process grounded directly upon the means stipulated for achieving a robust understanding of AI. Based on these two separate discussions I will conclude by outlining a composite critique – from the twofolded perspective of Spinoza's ethical and political discussions – of the ongoing negotiations of a new Conventional Weapons Convention (CCW) protocol on LAWS.

Keywords Intelligence · AI · Spinoza's philosophy of the mind and knowledge · Ethics · International humanitarian law · Lethal autonomous weapons systems

1 Introduction.

In this article I will argue that the philosophy of Benedict Spinoza (1632–1677) may assist us in coming to terms with some of the conceptual challenges that the phenomenon of Artificial Intelligence (AI)[1] poses on law and legal thought. I will pursue this argument in three steps. First, I will suggest that Spinoza's philosophy of the mind and knowledge may function as an analytical tool for making sense of the prevailing conception of AI within the legal discourse on Lethal Autonomous Weapons Systems (LAWS). I will do this by way of reading the perception within the specific legal discourse on LAWS (such as this unfolds in the negotiations of a new protocol to the United Nations' Convention on Certain Conventional Weapons) in the light Spinoza's metaphysical account of the mind and knowledge. This account enables an investigation into the reasons for why that discourse seems unable to advance beyond the preliminary question of determining what exactly is the nature of the object it is confronting and regulating (cf. International Committee of the Red Cross (ICRC) 2018; Group of Governmental Experts on Lethal Autonomous Weapons Systems (GGE) 2018; Docherty 2016; Roff 2014; Acheson 2017). Then, I will leave behind the question of the LAWS discourse's understanding of intelligence for a while, to continue the argument with the aid of Spinoza's political philosophy. This part of the argument will at first complicate the task of coming to terms with AI within law and legal thought as Spinoza's political philosophy seems to disqualify a normative process grounded directly upon the findings of the metaphysical investigation related to understanding. Finally, against the background of my review of both Spinoza's metaphysical and political treatment of the question of intelligence, I will conclude my argument by pointing to some problematic aspects of the current process of negotiating

✉ Moa De Lucia Dahlbeck
m.dahlbeck@law.gu.se

[1] School of Business, Economics and Law, Gothenburg University, Vasagatan 1, 405 30 Göteborg, Sweden

[1] Most discourses on Artificial Intelligence start in a basic definition of AI 'as a growing resource of interactive, autonomous, self-learning agency, which enables computational artifacts to perform tasks that otherwise would require human intelligence to be executed successfully' (Taddeo and Floridi 2018, pp. 751–752). Likewise, autonomy is usually understood as the ability '[i]n its simplest form, … of a machine to perform a task without human input. Thus an "autonomous system" is a machine, whether hardware or software, that, once activated, performs some task or function on its own.' (Scharre 2015).

a new Conventional Weapons Convention (CCW) protocol on LAWS.

My principal motive for using Spinoza's philosophy of the mind as a backdrop for an investigation into the LAWS discourse's current inability to respond (uniformly) to the phenomenon of AI, is because this discourse clearly illustrates how law's practical problems with AI are grounded in a conceptual difficulty to determine whether to understand AI as a threat or as an asset to the values protected by law (see Cath 2018; Floridi 2018; Nemitz 2018; Armstrong and Ray 2019; Braun 2019). I believe that Spinoza can assist us in making sense of this indetermination as he provides a detailed explanation of the epistemological conditions for human knowledge within the realm of an entirely naturalistic (and pan-psychist) account of intelligence (Della Rocca 2008, pp. 108–118; cf. Marshall 2013, pp. 129–133). To this end, it seems reasonable to investigate the idea of a non-human intelligence and its effects on law-making from the perspective of a philosophical account of intelligence that does not reject the attribution of a mind to inanimate objects (Ibid).

Spinoza recognizes only one faculty of the mind (see Nadler 2006, pp. 155–157; Della Rocca 2008, p. 118; LeBuffe 2010, p. 16). Spinoza's concept of the mind has been described as 'explained in the same mathematical way that one might explain the motions of bodies in space' and as 'something like the workings of the watch. Just as bodies in motion are to be explained by simple mechanistic laws, so too are the workings of the human mind' (Marshall 2013, p. 3). Spinoza's mechanistic explanation of the mind moves him to attribute one single function to it: to produces ideas (E2d3).[2] Ideas, however, can be more or less adequate (E2d4) meaning that they can have more or less of the intrinsic properties of a true idea (the external denomination of which is its correspondence with an object (cf. E2p43s). Adequate ideas are ideas that explain objects through these objects' own causal connections and power to change, whereas less adequate ideas produce an understanding of the object explained that is 'confused and mutilated' in so far as it is based on the effects caused on a particular body by the object explained by these ideas (E2p43s and E2p29c).[3] According to Spinoza's metaphysical scheme, God is the only being that understands every single thing according to adequate ideas (according to reason) (E2p32). All other things, since they are finite and, therefore, subject to external impacts, are bound to produce a mixture of adequate and confused ideas of their surroundings; i.e., to understand themselves and other things mainly affectively – according to how they themselves reacts to and are affected by the power of another thing – rather than according to that thing's reasons (E2p11c). A finite thing – always existing within a particular context and constellation of the world – will always initiate its perception of things from its own perspective within that context and constellation (E2p23d). This does not mean that it is impossible for a finite thing to produce ideas that explain an object according to (its) reason, but it does mean that such a representation will come neither more naturally nor be emotionally more powerful (convincing) than opaque, and perspectival ideas of the same object. In other words, a perfectly autonomous capacity to produce ideas – in so far as the mind does not rely on the impacts of other things in its production of ideas – is a capacity that is exclusive to God (or nature). Accordingly, no finite being can ever hope for perfect autonomy in the sense of a perfectly rational use of ideas; all they can do is strive towards such understandings (Sangiacomo 2015; Dahlbeck and Lucia 2020). According to Spinoza, for all things that are not infinite and eternal (like God) autonomy is a matter of relations, a quality that comes in degrees and never absolutely. Moreover, since Spinoza is committed to both monism (that there is only one substance) and philosophical naturalism (to treat all identical things according to the same explanatory principles) finite things – in so far as they are not God – must be understood as substantially equal, both in terms of intellect and extension (E1p11 and E3 Preface). The most important consequence of this is of course that it seems that Spinoza would consider it a logical fallacy to think of AI as completely autonomous, in relation to human or any other being. An AI will, just as a human being, be determined to understand things according to both ideas reflecting their reasons and those reflecting its own reaction (affects) to these reasons.

This is the philosophical backdrop against which I will evaluate the LAWS discourse's understanding of AI. I will argue that from the perspective of Spinoza's account of the mind and knowledge the LAWS discourse's apparent problem of finding a shared understanding of AI (as either a

[2] E2d3 reads: "By idea I understand a concept of the Mind that the Mind forms because it is a thinking thing." All references to Spinoza's works are to Curley 1985 and 2016 and I employ their method of referring to the parts of the texts. Passages from the Ethics are referred to according to the following form of abbreviation E –Ethics, ax-axiom, c-corollary, d-demonstration, def-definition, L-lemma, p-proposition, post-postulate, s-scholium. Example: E2p7s = Ethics, Part 2, Proposition 7, scholium. The references to the Theological Political Treatise (TTP), the Political Treatise (TP) and the Treatise on the Emendation of the Intellect (TdIE) are supplemented by references to Gebhardt's edition Spinoza Opera, according to the following form: G II/208/25-30 = Gebhardt, vol. 2, p. 208, lines 25–30.

[3] In the Cartesian tradition, clear and adequate ideas have been associated with a faculty of understanding or knowing and opaque or inadequate ideas have been associated with a faculty of the will or judgment (consisting in the ability to choose which one of two conflicting understandings is adequate).

threat or an asset) seems to be caused by the fact that it entertains an understanding of intelligence which lends itself to very different interpretations of things and events. Before I go further into this, let me just give a brief description of the background to the LAWS discourse's problem of finding a common starting point for negotiations. The inability within this discourse to reach an agreement as to what autonomous intelligence means, has by one commentator been described (although in relation to the wider debate on AI-technology) as follows:

> The word "autonomy" is used by different people in different ways, making communication about where we are headed with robotics systems particularly challenging. The term "autonomous robot" might mean a Roomba to one person and a Terminator to another! (Scharre 2015, quotation marks in original).

The same ambiguity about the essence of AI understood as autonomy can be detected in the ongoing negotiations on a new CCW protocol on the use of LAWS as well. Commenting on these negotiations, Chris Jenks refers to this indecisiveness as setting.

> The conditions for a dialogue bordering on incoherence. So much so that it would be tremendous progress for the international community if there was a complete and utter lack of consensus regarding whether to develop and employ LAWS, but agreement as to what was meant by LAWS. But as of now, we cannot even agree on what we are discussing (Jenks 2016).

The result of the situation described in this passage is the complete inability to agree upon whether to endorse, regulate or prohibit LAWS (c.f. Ekelhof 2017, p. 312). Jenks argues emphatically that even though a 'constructive LAWS dialogue requires a shared and coherent understanding of machine or system autonomy', the international community has, as of yet, 'neither, and perhaps even worse, continues to engage in overly broad and conceptually confusing inquiries' (2016). Despite affirming the importance of achieving a shared understanding of what distinguishes the intelligence of LAWS from that of human beings, Jenks goes on to argue for the futility of efforts to try to do so. He backs away from the initial reiteration of the naturalness and reasonableness of 'the desire to define autonomy' by stating that 'such efforts will inevitably be counterproductive' as they increase the confusion and distracts the dialogue from dealing with its real problems (Jenks 2016). Although the nature of these "real" problems is never explicated by Jenks, I assume that he associate it with the practical and technical aspects of LAWS, since he prescribes a focus on the weapons systems' practical functions of selecting and engaging in targets as a means for overcoming the current stand-still in the negotiations (Jenks 2016).

Finding Jenks' dismissal of his own call for a 'shared and coherent understanding of machine or system autonomy' unconvincing, I will instead hold on to that call as my point of departure for this article's first part on the LAWS discourse's understanding of AI. In what follows I will look specifically at the LAWS discourse's conceptualization of the distinguishing feature of AI in the light of Spinoza's accounts of the mind and knowledge.

2 The fear and hope of the LAWS discourse: Spinoza on rational and emotional intelligence

As already mentioned, discussions within the realm of international humanitarian law on how to approach LAWS indicate the existence of a variety of attitudes towards AI among the discourse's stakeholders.[4] These attitudes vary from a complete faith in AI as an asset to human cognition, to fear of AI as an unstoppable super rationality, resulting in either the recommendation to prohibit LAWS or to embrace it through existing regulation. Among those who argues that LAWS can be controlled by existing or new regulation, AI seems to be perceived as a possible means for human beings to complement their largely emotional nature of thought with a superior capacity to understand according to reason (see Arken 2013). Among those who argues for a complete prohibition, it is the very same feature – of a superior reason – that appears threatening to human values. In either way, the intelligence embodied within LAWS is clearly being portrayed in accordance with its (potential) effects on human beings such as this is perceived from the particular perspective of the stakeholder who is speaking.

The Human Rights Watch (HRW) constantly refers to AI in terms of different human emotional states. For instance, it writes that "although fully autonomous weapons would not be swayed by fear or anger, they would lack compassion, a key safeguard against the killing of civilians" (2015). To the HRW, this lack is of such fundamental significance that LAWS ought to be prohibited. Likewise, during the negotiations of a new CCW protocol, particular LAWS are regularly discussed in terms of whether they allow for a "meaningful human control" or, to what extent they can be made to feign something that objectively reminds of an "appropriate human judgment" (CRC 2016). The International Committee

[4] These are principally the discussions pursued under the auspices of 1977 Additional Protocol I to the 1949 Geneva Conventions about the scope and range of Article 36 and discussions under the auspices of the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects as amended on 21 December 2001 (CCW).

of the Red Cross (ICRC), has questioned whether LAWS can assess proportionality in a manner that satisfies the IHL standard since this standard is 'a question of common sense and good faith' (1987, pp. 679, 682).

These examples have all been interpreted as confirmations of the legislative problems related to AI as being caused by cognitive misunderstandings between a fully rational and a less than fully rational being (see Acheson 2017). Thus, what seems to stand in the way for a shared starting point of the LAWS-negotiations is the difficulty to make sense of AI's apparent lack of the aspect of human intelligence that disrupts rationality: i.e., the tendency to perceive things according to the imagination (according to things' effects). To this end, it has been argued that what distinguishes artificial from human intelligence is the non-presence in the former of a spontaneous, and, therefore, incalculable, way of reacting to the unexpected events that are so typical for war (Robinson 2015). In sum, it seems fair to say that most current attempts to discuss normative approaches to AI and LAWS do so on the premise that the distinguishing feature of AI in relation to human intelligence is the former's lack of a contextual and reactive (emotional) intelligence. In order then to answer the question of why the LAWS discourse has not yet been able to advance beyond the preliminary question of what AI is (a threat or an asset) from Spinoza's perspective, I suggest that we begin here: in the discourse's different conceptualizations of AI that all take as their point of departure the aspect of human intelligence that disrupts rational thought.

For Spinoza, there are, as mentioned, not two separate faculties of the mind. To think is to do one single thing: to produce ideas that explain objects. As mentioned above, sometimes ideas will not explain an object so much according to the object's own powers to change (affect) things around it (and these powers' natural causes, in turn), but more according to the impacts of these powers on the mind that produces the ideas. When ideas correspond to their objects in the sense that they explain them through their constituting affective powers and their causes, then, the mind produces ideas that are clear and represent their objects through (their) reasons (E2p43s). However, we should not forget that the human mind is inclined towards perceiving a thing according to how that thing affects itself, rather than according to its own constituting powers (E2p19, 23, 25, 26, 27, 28). In fact, Spinoza's entire ethical project can be said to motivated by the human mind's cognitive difficulty to produce clear and adequate ideas despite the ethical advantage (which I will explain further in short) of doing so (TdIE 13-14 G II/8-9).

According to Spinoza, the cognitive difference between affective, or emotional, understandings and rational understandings is not that they stem in different faculties but that the first is constructed around ideas that focus on the mark left upon the contemplating being in its encounters with the thing contemplated, whereas the second focuses on the causes of the thing contemplated. He writes:

> I say expressly that the Mind has, not an adequate, but only a confused [NS: and mutilated] knowledge, of itself, of its own Body, and of external bodies, so long as it perceives things from the common order of nature; i.e., so long as it is determined externally, from fortuitous encounters with things, to regard this or that, and not so long as it is determined internally, from the fact that it regards a number of things at once, to understand their agreements, differences, and oppositions. For so often as it is disposed internally, in this or another way, then it regards things clearly and distinctly, as I shall show below (E2p29s, emphasis added).

In so far as the mind contemplates things, not based on the ideas reflecting the affective changes of its own constitution resulting from its encounters with the things contemplated, but based on the causal powers of the things, it obtains an understanding of them based on their reasons. This understanding is more reliable – and thus more useful (than an affective understanding of a thing) – when it comes to making evaluations of what kind of interactions are good or bad for the evaluating thing's own affective power (to preserve in being). In other words, there is a practical ethical problem related to the fact that human cognition is inclined towards its own, particular and affective perspective. According to Spinoza, this problem emerges from the metaphysical premise that the mind's principal object is its own body[5] and that a complete liberation from its own particularity when thinking, therefore, is impossible.

> For the Mind does not know itself except insofar as it perceives ideas of the affections of the body (by P23). But it does not perceive its own Body (by P19) except through the very ideas themselves of the affections [of the body], and it is also through them alone that it perceives external bodies (by P26). And so, insofar as it has these [ideas], then neither of itself, nor of its own Body (by P27), nor of external bodies (by P25) does it have an adequate knowledge, but only (by P28 and P28S) a mutilated and confused knowledge, q.e.d. (E2p29c).

[5] In E 2p11 Spinoza writes that "the first thing that constitutes the actual being of a human Mind is nothing but the idea of a singular thing which actually exists". In proposition 13, he clarifies that the actually existing thing, of which the mind is a reflection, is the human body. In other words, the mind's nature or essence is to be the idea – the representation – of its own body.

This passage aptly shows how the finite mind is always bound to begin to understand things through mutilated and confused ideas; ideas that represent the thing reflected together with that thing's effect on the body and mind of the thinking being. This constitutes a practical problem for human beings in so far as ethical wellbeing (i.e., the successful striving to persevere in being), according to Spinoza, amounts to the achievement of a clear (rational) understanding of the world (TdIE 13-14). It is thus part of the individual's striving towards ethical wellbeing to try to counter the natural inclination towards perspectival and confused (irrational) understanding (E4p18s).

The metaphysical explanation of the relation between wellbeing and cognition begins in Spinoza's definition of the essence of human (or any other finite thing's) nature as a power to persevere in being; to be active (E3p7 and E4d8). This explains why Spinoza uses the term passive when describing a mind that forms ideas based on external impacts on the thinking thing. It also explains why ideas explaining a thing based on that thing's constitution for Spinoza are evidences of an active mind; one that is relatively successful in its striving towards persistence in being. The output of the human natural inclination towards passivity together with the goal of increased activity amounts to Spinoza's normative ethics: adequate, clear ideas give more stable and trustworthy representations of what acts and things are beneficial for an individual to interact with to persevere in being. In short, they are objectively better at revealing what things are good and what things are bad for the thinking thing's chances at persevering in being, given its own affective powers. Relying on ideas formed out of passive encounters for these kinds of decisions may lead to miscalculations and acts that are detrimental for individuals. It is precisely this scenario that Spinoza describes in the preface of the Theological Political Treatise (TTP) where he emphasizes that confused ideas, because they do not take as their objective a thing in itself but rather its effects (as these have been perceived by someone), may produce very different understandings of one and the same thing. One and the same object may be the object to both fear and hope.

> [1] While the mind is in doubt, it's easily driven this way or that – and all the more easily when shaken by hope and fear, it comes to a standstill. At other times, it's over-confident, boastful and presumptuous. (…) [3] If, while fear makes them turn this way and that, they see something happen which reminds them of some past good or evil, they think it portends either to a fortunate or an unfortunate outcome; they call it a favorable or unfavorable omen, even though it may deceive them a hundred times. (TTP preface 1 and 3).

It is along these lines, then, that must we try to explain the LAWS discourse's complete stand-still if we wish to do so from the perspective of Spinoza's philosophy of the mind and knowledge. To this end, the discourse is unable to advance because it takes as its point of departure an understanding of intelligence that lends itself to different – not even necessarily similar – perceptions of AI. According to Spinoza's account of the relationship between irrational ideas and hope and fear, only inadequate (irrational) ideas can give rise to disparate perceptions of a thing. They do so since they represent an object based on how it has affected another body – not based on its causes. Thus, we may conclude that the first step to take for the LAWS discourse to proceed to a more productive phase would be to try to establish an idea of intelligence that does not lend itself to multiple interpretations of the affective powers of AI. Such an idea is only possible to form in so far as the affective powers of AI are appreciated and represented in accordance with their own causal order, and not in accordance with the impacts of these powers upon another thing's particular composition of affective powers.

Spinoza's account of the mind gives us a few hints of how to assess the AI's affective powers adequately. As I have already mentioned, Spinoza is committed to both monism and philosophical naturalism which means that there is only one substance in the world and that all finite things are to be treated according to the same explanatory laws and principles, metaphysically speaking. In addition to this, Spinoza establishes a metaphysical parallelism which stipulates that everything that happen in one expression of substance – an attribute, to use Spinoza's own term – happen too in all other expressions simultaneously. There are only two known such expressions according to Spinoza: extension and thought. In other words, parallelism commits Spinoza to hold that whatever happens in the body happens simultaneously in the mind and vice versa. Putting together all of these metaphysical premises to understanding a new particular kind of finite intelligence, we get the following point of departure: all finite things are equipped with a mind whose terms of being and principles of activity function in one and the same way. There cannot be, then, a substantial difference between AI and human intelligence. In so far as the human mind strives towards its own preservation (and more adequate ideas) so does the AI. In so far as the human mind is substantially conditioned to a relative autonomy and a less than perfect rationality, so is the artificial mind.

How then can we explain the difference that we undeniably perceive when considering the intelligence within LAWS in relation to the intelligence of human soldiers? According to Spinoza we must do so on the same terms that we use to distinguish one human mind from another: through relying on its essential power to strive to persevere in its being (E3p7) which produces a mind that is 'more capable, the more its body can be disposed in great many ways' (E2p14).

Attributing a mind to a machine, weapon or other artefact implies to attribute all of the components (emotional and rational) identified within the human mind to it. What may differ is their respective complexity and degree of ability to produce clear and adequate ideas. However, even if it could be established that AI's essential power is more prone to produce clear (rational) ideas than the human being in general is, this does not make AI immune from other things' affects in its production of ideas (as the idea of AI as a perfectly rational being presumes). The LAWS discourse's idea of AI (as a perfectly rational being) reveals to this end how it is formed out of mostly passive affects. To the extent that the discourse's ideas of AI as a perfectly rational being are associated with either hope or fear, we can say then that this idea of AI is inadequate. The explanation of why the LAWS discourse cannot come to terms with whether AI constitutes a threat or hope for the values protected by law perhaps lies, then, in the overwhelming passive affects of the discourse's members at the moment of conceptualizing the nature of AI.

3 Spinoza's political response to human emotional intelligence

An analysis such as the present, of a specific discourse's approach to the idea of intelligent machines from the perspective of Spinoza's philosophy, could arguably have been brought to an end here: at the point where we have discussed why the discourse's current approach leads to conceptual confusion rather than a shared starting point for negotiations. The reason for pursuing the argument a bit further is straight-forward: Spinoza's understanding of what human beings should do to improve their ability to objectively evaluate what affective connections are ethically beneficial for them is not immediately translatable into normative terms. In fact, Spinoza's political philosophy can be read as completely cancelling out what I just have argued about his own recommendations to a discourse struggling with incommensurable understandings of a thing. In his normative ethics, Spinoza stipulates what is best for individual human beings in so far as their ethical freedom is concerned. This is to say that the normative ethics contains advice on how to make an individual's mind more active (free) and produce, as a consequence, more rational ideas. This is what I have reviewed above, in the context of LAWS. In the political philosophy Spinoza provides advice on how to organize the human social condition so that individuals may pursue the just mentioned normative ethics as with as few external disturbances as possible (Sangiacomo 2015). In Spinoza's own words, political theory is separated from ethics because 'freedom of the mind is a private virtue and the virtue of the state is security' (TP 1, 6). Thus, even though interrelated through one being a means to the other, Spinoza's normative ethics is one thing and his normative theory is another completely, in so far as the latter needs to strive towards an independent end to make the first at all realizable (Dahlbeck and Lucia 2020).

It is according to this logic that the two different ends expressed by Spinoza in relation to civil society (in his political works) usually are explained: peace and security is society's primary end in so far as it is a precondition for the overall end of human freedom (cf. TP 5, 2 and 5; TTP 3, 20 and TTP 20, 12 and see Curley, 2016: 506, footnote, 14). To this end, one of the more important premises of Spinoza's ethical thought works as a limiting condition for his normative thought too. Human beings are inclined towards sociability as perseverance is too hard in isolation (E4p35s). Not only does this mean that humans must take this sociability into account individually when calculating their own striving towards wellbeing (as I have described in the section above), but a stable society is an absolutely necessary condition for mapping out and facilitating a successful individual striving in the first place. As such, a well-ordered society is the most important condition for a successful striving towards the human highest ethical good of freedom (i.e., rationality) (TTP 20, 12). This is probably why the TTP's discussion on the means to this freedom does not revolve around how to use reason to combat passions (as in the Ethics) but around human beings – those being governed as we well as those who govern – limited cognitive ability and how to get them to live so that philosophical freedom can flourish.

In the TTP, Spinoza clarifies why this turns obedience into the necessary starting point for a good political society and not reason, even though reason is what such a society will allow, encourage, and sometimes, even transmit.

> Though the voice the Israelites heard could not give them any philosophical or mathematical certainty about God's existence, still, it was enough to make them wonder at God, insofar as they had previously known him, and to motivate them to obedience. That was the purpose of the manifestation. God did not want to teach the Israelites the absolute attributes of his essence. (He did not reveal any of them at that time.) He wanted to break their stubborn heart and win them over to obedience. So he addressed them with the sound of trumpets, with thunder, and with lightning, not with arguments. (TTP 14, 36, emphasis added).

Spinoza prescribes that political and legal activity should be directed towards obedience, rather than the (ethical) problem of what is the best way of living for humans. To be directed towards obedience implies for Spinoza that political theory is dedicated to examining the composition of affective powers of the people that is to be guided into a peaceful and stable living condition, and to formulate its strategy for peace and stability based on the results of this examination rather than the idea of

an already stable and peaceful society. In the preface to the Political Treatise (TP) Spinoza describes how normative philosophy tends to erroneously take the components of stable and well-functioning society for the means to guide naturally instable and affective individuals towards the establishment of such.

> Philosophers conceive the affects by which we're torn as vices, which men fall into by their own fault. That's why they usually laugh at them, weep over them, censure them, or (if they want to seem particularly holy) curse them. They believe they perform a godly act and reach the pinnacle of wisdom when they've learned how to bewail the way men really are. They conceive of men not as they are, but as they want them to be. That's why for the most part they've written Satire instead of Ethics, and why they've never conceived a Politics which could be put to any practical application, but only one which would thought a Fantasy, possible only in Utopia, or in the golden age of the Poets, where there'd be absolutely no need for it. (TP 1[1], G III/273/4-17, emphasis added).

The close relationship between the end of ethical freedom and the end of a stable social context is rendered more explicit in the following excerpt from the TTP:

> A social order is very useful, and even most necessary, not only for living securely from enemies but also for doing many things more easily. For if men were not willing to give mutual assistance to one another, they would lack both skill and time to sustain and preserve themselves as far as possible. […] Everyone, I say, would lack both the strength and the time, if he alone had to plow, to sow, to reap, to grind, to cook, to weave, to sew, and to do the many other things necessary to support life […] which are also supremely necessary for the perfection of human nature and for its blessedness. […] Now if nature had so constituted men that they desired nothing except what true reason teaches them to desire, then of course a society could exist without laws; in that case it would be completely sufficient to teach men true moral lessons [the divine law], so that they would do voluntarily, wholeheartedly, and in a manner worthy of a free man, what is really useful. But human nature is not constituted like that at all. It's true that everyone seeks his own advantage – but people want things and judge them useful, not by the dictate of sound reason, but for the most part only from immoderate desire and because they are carried away by affects of the mind which take no account of the future and of other things. That's why no society can continue in existence without authority and force, and hence, laws which moderate and restrain men's immoderate desires and unchecked impulses. (TTP 5, 18–22, emphasis added)

A few interesting things can be noted on the basis of this excerpt. First of all, it demonstrates that human nature for Spinoza, while geared towards the striving for ethical freedom, is not sufficient on its own to guide us reliably towards this. This is because human nature is naturally inclined towards irrationality, i.e., understandings based on passively caused ideas of things' explanations. Second of all, it clarifies that it is the inclination towards irrationality in human nature that establishes the need for law and ordered society in the first place. This proposition can be read in conjuncture with an explanation made by Spinoza in the Ethics to the end that '[n]o affect can be restrained by the true knowledge of good and evil insofar as it is true, but only insofar as it is considered as an affect' (E4p14). And because '[a]n affect cannot be restrained or taken away except by an affect opposite to, and stronger than, the affect to be restrained' (E4p7), the cited excerpt is concluded with the affirmation that it is insufficient to rely solely on the power of truth and reason in political matters. In sum, because human beings are inclined towards irrational ideas, the state – working for the mutual well-being of all – cannot rely on the rational superiority of its measures when communicating these to the people as they simply will not be convincing to someone governed by emotions.

So, insofar as rationality tends to be conquered by powerful passions in a state of nature, the mark of a good state is that it does not allow for passive affects and ideas that inhibit its citizens' rational behaviors. To do this, however, the state's orders aimed at countering the passions that are detrimental to peace and stability must be convincing for those who are governed by them. They will only be convincing to the extent that they are cast in terms that corresponds to the level of perspectival and irrational understanding of things that dominate among a given people (Dahlbeck and Lucia 2020). As the cited passage above reveals, human beings that are left alone to determine what is good for their perseverance will rarely succeed in this because they are not – none of us are – susceptible to rational reasons because of their rationality.

The immediate consequence of all of this for the question of AI and law is that even though we may convince those negotiating a new CCW protocol on LAWS, that a shared and coherent understanding of AI depends on ideas that explain AI in accordance with its own affective powers (causes), there is no guarantee that this understanding will form the basis of a protocol that is efficient. Efficiency – in terms of actually curbing the affective powers of those affected by the protocol in a way that makes them act in

accordance with the stipulated norms – depends on how well the lawmaker has understood the affective powers at work in the governed people, not the object of the law. What makes a law good qua law, for Spinoza, is the extent to which it reflects the mentality (the specific composition of passively and actively produced ideas) of its subjects and the extent to which it communicates its norm in terms that are convincing for this particular mentality (TTP 5, 26).

For a regulation of the use of AI to be efficient in this sense, then, it is not enough to juxtapose the law-makers inadequate understanding of the object of regulation with a more adequate one (in accordance with the outline in the previous section). An efficient regulation depends also upon a thorough study of the particular composition of the inadequate passive understandings of AI (why it invokes fear in some and hope in others) that are at work in the particular legal practices affected. Only by combining these two methods can the discourse begin to formulate a protocol that makes the legal subjects behave towards AI in way that is beneficial for the goal in relation to which the protocol ultimately is a means: peace and security.

4 Conclusions

To conclude my argument on making sense of the problem of AI in law with the help of Spinoza's philosophy, I wrote initially that I would contemplate his ethical and political discussions of intelligence and human cognition together so that I could identify separate but inter-related grains of critique of the law-making process taking place under the auspices of the CCW negotiations on a LAWS protocol. Having studied the LAWS discourse first through the lens of Spinoza's account of the mind and knowledge and then through that of his political theory, I suggest the following critical notions with respect to the legal treatment of AI.

First of all, the dominating idea of AI within the law-making discourse seems to be confused and perspectival, which makes it difficult to consider seriously what kinds of human measures and acts are good and bad in relation to AI, in the light of the overall ethical goal of increased human wellbeing. A clear and adequate idea of AI – i.e., an idea reflecting the nature of AI's constitutive powers – would be more useful for the legal community in so far as such would allow for a more adequate evaluation of AI and its affective powers in relation to those within the community and its goal of increasing people's wellbeing. However, and this is the second critical notion, the means Spinoza envisions for a successful individual ethical striving towards wellbeing are not the same as those he envisions for the establishment of a stable civil society, although the latter is an absolute precondition for the first.

Put differently, coming to terms with AI in law and legal thought is thus not a question of determining who is right among those who thinks AI is a threat and those who thinks it is an asset. We know now that Spinoza would classify both as inadequate and passive perceptions of AI in that both reflect more of the thinking being's own affective constitution than they do the nature of AI. Neither, however, is it a question of merely installing an understanding of AI according to its natural causes and powers as the basis of legislation. Even if such an understanding would help the lawmaker in its assessment of how to deal with AI in order for AI to disturb as little as possible individuals' striving towards ethical wellbeing, it is not enough in itself to secure a good legislation, i.e., a legislation that encourages peace and security as an external condition for ethical wellbeing. In the light of Spinoza's political philosophy, then, coming to terms with AI in law and legal thought must rather begin with an examination of the specific conceptualizations that associate AI with fear and hope respectively, so that the negotiations on a new CCW protocol on LAWS can proceed with these affective powers in mind. To judge by Spinoza's normative theory, the fear and the hope generated in human beings by their affective encounter with AI is more dangerous and detrimental for peace and stability than AI taken on its own appears to be. This is why good laws are laws that succeed in 'moderating and restraining men's immoderate desires and unchecked impulses'. In the light of this, the legal discussion on how to regulate human interactions with AI must perhaps endorse for legal measures and norms to adaptable and varying according to the specific desires and impulses that dominate within the different particular contexts in which they are to function. To check reactions to an idea of AI originating in fear does arguably require different methods than those aimed at moderate those stemming in hope.

Acknowledgements Open access funding provided by University of Gothenburg. I am especially grateful to Gregor Noll, Valentin Jeutner, Johan Dahlbeck and the two anonymous reviewers for their valuable comments and detailed suggestions for improving the essay. I thank Riksbankens Jubileumsfond for funding. The essay forms part of my post-doctoral research within the research project Digital Krigsföring: Ansvar, Avsikt och Rättsäkerhet (project-id: P17-0719:1).

References

Acheson R (2017) Losing control: the challenge of autonomous weapons for laws, ethics, and humanity. CCW Report, Reaching Critical Will. Civil Society Perspectives on the CCW Group of Governmental Experts on lethal autonomous weapons systems, 13–17 November 2017. https://www.reachingcriticalwill.org/disarmamentfora/ccw/2017/laws/ccwreport/12166-ccw-report-vol-5-no-3. Accessed 11 Dec 2018

Arken R (2013) Lethal autonomous systems and the plight of the non-combatant. AISB Q 137:1–9

Armstrong H and Ray J (2019) A working model for anticipatory regulation. Nesta. https://nesta.org.uk/report/a-working-model-for-anticipatory-regulation-a-working-paper/. Accessed 10 Oct 2019

Braun R (2019) Artificial intelligence: socio-political challenges of delegating human decision-making to machines. IHS Work Pap Ser 6:24

Cath C (2018) Governing artificial intelligence: ethical, legal and technical opportunities and challenges. Philos Trans R Soc A Math Phys Eng Sci 376(2133):20180080

Dahlbeck J, De Lucia DM (2020) The moral fallibility of spinoza's exemplars: exploring the educational value of imperfect models of human behavior. Ethics Educ. https://doi.org/10.1080/17449642.2020.1731106

Della Rocca M (2008) Spinoza. Routledge, New York

Docherty B (2016) Making the case: the dangers of killer robots and the need for a pre-emptive ban. In: Human Rights Watch; Harvard Law School. International Human Rights Clinic. Human Rights Watch, New York. https://www.worldcat.org/title/makingthe-case-the-dangers-of-killer-robots-and-the-need-for-a-preemptive-ban/oclc/970365994. Accessed 11 Dec 2018. ISBN: 9781623134310 1623134315

Ekelhof M (2017) Complications of a common language: why it is so hard to talk about autonomous weapons. J Conflict Secur Law 22(2):311–331

Floridi L (2018) Soft ethics, the governance of the digital and the general data protection regulation. Philos Trans R Soc A Math Phys Eng Sci 376(2133):20180081

Jenks C (2016) The distraction of full autonomy & the need to refocus the CCW LAWS discussions on critical functions. In presentation delivered as part of the 2016 Informal Meeting of Experts on Lethal Autonomous Weapons as part of the Convention on Certain Conventional Weapons treaty. https://www.semanticscholar.org/paper/The-Distraction-of-Full-Autonomy-%26-the-Need-to-the-Jenks/cb624f975b1fc38f70e8784390509b9ff1551972

LeBuffe M (2010) From bondage to freedom: spinoza on human excellence. Oxford University Press, Oxford

Marshall E (2013) The spiritual automaton: spinoza's science of the mind. Oxford University Press, Oxford

Nadler S (2006) Spinoza's ethics: an introduction. Cambridge University Press, Cambridge

Nemitz P (2018) Constitutional democracy and technology in the age of artificial intelligence. Philos Trans R Soc A Math Phys Eng Sci 376(2133):20180089

Robinson R (2015) 3 human qualities digital technology can't replace in the future economy: experience, values and judgement. Blog post. https://theurbantechnologist.com/2015/04/12/3-human-qualities-digital-technology-cantreplace-in-the-future-economy-experience-values-and-judgement/. Accessed 20 Apr 2020

Roff H (2014) The strategic robot problem: lethal autonomous weapons in war. J Mil Ethics 13(3):212

Sangiacomo A (2015) Spinoza and relational autonomy: an outline. In: Eckert M, Cunico G (eds) Orientierungskrise. Herausforderung des Individuums in der heutigenGesellschaft. Roderer, Regensburg

Scharre P (2015) Between a roomba and a Terminator: what is autonomy?. In: War on the rocks texas national security review. https://warontherocks.com/2015/02/between-a-roomba-and-a-terminator-what-is-autonomy/20/04/20. Accessed 20 Apr 2020

Taddeo M, Floridi L (2018) How AI can be a force for good. Science 361:6404

Official reports and Legal Instruments

Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons (2016) Report of the 2016 informal meeting of experts on lethal autonomous weapons systems. In: Fifth review conference of the high contracting parties to the convention on prohibitions or restrictions on the use of certain conventional weapons which may be deemed to be excessively injurious or to have indiscriminate effects, 10 June 2016

Group of Governmental Experts on Lethal Autonomous Weapons Systems (2018) Report of the 2018 group of governmental experts on lethal autonomous weapons systems. Geneva, 9–13 April. https://www.unog.ch/80256EDD006B8954/(httpAssets)/20092911F6495FA7C125830E003F9A5B/$file/2018_GGE+LAWS_Final+Report.pdf. Accessed 11 Dec 2018

Human Rights Watch (2015) Mind the gap: the lack of accountability for killer robots. Report of the HRW, 9 April 2015. https://www.hrw.org/report/2015/04/09/mind-gap/lack-accountability-killer-robots#page. Accessed 11 Dec 2018

International Committee of the Red Cross (1987) Commentary on the additional protocols of 8 June 1977 to the Geneva conventions of 12 August 1949, Martinus Nijhoff Publishers, Geneva

International Committee of the Red Cross (2018) Ethics and autonomous weapon systems: an ethical basis for human control?. https://www.icrc.org/en/document/ethicsand-autonomous-weapon-systems-ethical-basis-human-control. Accessed 11 Dec 2018