What Should We Know When Interacting with Machines? A Critique of Daniel Dennett's Idea

Rafał Michalczak[1] and Maciej Próchnicki[2]

[1] Department of Legal Theory, Jagiellonian University, Kraków, Poland
Institute of Philosophy, Jagiellonian University, Kraków, Poland
michalczak.rafal@gmail.com
[2] Department of Legal Theory, Jagiellonian University, Kraków, Poland
Institute of Philosophy, Jagiellonian University, Kraków, Poland
maciej.m.prochnicki@gmail.com

Abstract. In this paper we present a criticism of Daniel Dennett's argument about very strict legal regime concerning the duties of the programmers and developers of artificial intelligence systems, presented in the book From Bacteria to Bach and Back: The Evolution of Minds. The argument, which postulates strict legal regime for intelligent machines, is based on the uncertainty about the potential negative effects of the development of intelligent systems. The proposal includes severe sanctions, as well as the necessity to provide the maximum possible information about the system to the user, or the party to a contract. Firstly, we try to argue that these claims are to a large extent incompatible with the general idea presented in the book, which could be described as Dennett's strange inversion of reasoning. Dennett analyzes ideas of notable thinkers, inter alia Darwin and Turing, and proposes a concept of competence without comprehension as a key factor to understand the evolutionary process of mind-shaping. In the second part of the paper, we also try to evaluate legal-theoretical and practical consequences of adapting such solutions, mostly in the context of criminal and civil law. We claim that there are, no real ratio legis behind the type of crime proposed by Dennett, and that the informational duties should be shaped in accordance with human cognitive capabilities, rather than being excessive.

Keywords: Artificial Intelligence and Law · Machine Law · Legal Philosophy · Legal Responsibility · Criminal Law · Civil Law · Philosophy of Mind.

1 Introduction

Advances in the development of new technologies, especially in artificial intelligence systems, provide a great challenge for the modern legal systems. In most cases, the lawmaker cannot foresee the shape of new systems and problems arising thereof, which most frequently makes it stand one step behind the need for regulation. However, the advancement of intelligent systems raises many concerns, doubts, or even fears, not only among common folk, but also engineers,

scientists and thinkers specializing in the field [1]. Could the legal reforms go ahead and provide a solution to, at least some, of these problems? There is a lively debate, and when one of the most prominent and influential contemporary philosophers and cognitive scientists makes a strong statement in that case, it is surely worthwhile to analyze it thoroughly.

In his newest book, "From Bacteria to Bach and Back: The Evolution of Minds" [2] (hereinafter referred to as FBB), Daniel Dennett presents a complex and comprehensive view of minds (not only human), taking an evolutionary approach. Drawing from the most notable ideas presented in his earlier works, Dennett presents conceptual milestones in evolution of cognitive capacities from the age of archea to the age of postintelligent design. At the end of his journey, he formulates a few remarks on the future of humans and machines. On the margin of them, one really strong and bold statement, concerning the issue of machine law – legal problems arising from the development of artificially intelligent beings – should be noticed.

> When you are interacting with a computer, you should know you are interacting with a computer. Systems that deliberately conceal their shortcuts and gaps of incompetence should be deemed fraudulent, and their creators should go to jail for committing the crime of creating or using an artificial intelligence that impersonates a human being.

In the following paragraphs, Dennett discusses potential duties required from the creators of intelligent machines. Most of them have an "informational" character (i.e. the program should always inform you that it tries to "read your mind" – e.g. assuming your preferable choices, due to the use of algorithms and statistical data), but some of them impose more practical tasks on developers of the program (there should always be an option to turn off the intelligent features of the program). However short this fragment of the book is, it raises a lot of important issues about law and new technologies. In this paper, we will try to elaborate on these issues and provide a constructive critique of some of the arguments raised in FBB. Dennett claims that only strict regulation of intelligent machines can be a solution to potential dangers. This claim is based on the premise of our limited comprehension of artificial systems, as well as limited capacities of these systems – they cannot perform as well as we want them to, due to the lack of comprehension by themselves. However, they can perform in fraudulent way, faking their gaps, and therefore deceive us. Our starting point will be the claim that ideas and arguments presented in sections of FBB that precede the fragment cited above suggest something opposite in relation to the legal aspects of the expansion of the intelligent technologies. This would mean that the whole argument is inconsistent in general. Secondly, we will focus on the more specific points made by the author of FBB, analyzing the proposal of duties that should be imposed on AI creator. Our analysis will cover the viewpoint of civil law (mostly consumer law), as well as criminal law. Our argumentation will concern mostly examples from Polish legal system (as an instance of a continental legal system), but they could be *mutatis mutandis* generalized to other legal systems, including common law ones.

2 A Strange Re-Inversion of Reasoning

FBB has a few recurring themes. One of them is the idea of "strange inversion of reasoning". Introducing this term in the article from 2009 [3], Dennett quotes Robert MacKenzie Beverley, 19th century critic of Charles Darwin's work.

> In the theory with which we have to deal, Absolute Ignorance is the artificer; so that we may enunciate as the fundamental principle of the whole system, that, IN ORDER TO MAKE A PERFECT AND BEAUTIFUL MACHINE, IT IS NOT REQUISITE TO KNOW HOW TO MAKE IT. This proposition will be found, on careful examination, to express, in condensed form, the essential purport of the Theory, and to express in a few words all Mr. Darwin's meaning; who, by a strange inversion of reasoning, seems to think Absolute Ignorance fully qualified to take the place of Absolute Wisdom in all of the achievements of creative skill.

Dennett continues to use this "tool for thinking" [4], describing also the work of other influential thinkers, such as Alan Turing. Moreover, he also uses the term to analyze the work of David Haig [5]. What is the idea behind all these strange inversions? In accordance to Darwin, it is the statement, that in order to "to make a perfect and beautiful machine, it is not requisite to know how to make it" [2]. The idea by Turing, on the other hand, can be expressed as the following: "In order to be a perfect and beautiful computing machine, it is not requisite to know what arithmetic is." What is the essence of strangeness of these inferences? It is the counterintuitive idea that one can be able to do something without understanding how it works – the competence without comprehension. It is a general meta-rule for all strange inversions of reasoning to put competence before comprehension, because as Dennett notes, comprehension consists of competences [2]. It could be called consequently Dennett's strange inversion of reasoning. This brainchild of his is the key to understand the process of evolution. According to Dennett, the evolution is a kind of intelligent design, which does not require designer, relying on a set of free floating rationales instead [6]. Organisms learned to make use of affordances – potential possibilities offered by environment [7] – and gradually came to the cognitive capacity of comprehension as a result of increasing biological complexity. This is closely connected with another phenomenon: consciousness, which could be perceived in Dennett's approach to it as a "user-interface".

However, Dennett denies the conjecture that advancement in AI would bring us some kind of superintelligence [2]. Hence, although popularity of the learning algorithms, as well as evolutionary ones, is growing, it seems that AI is far from obtaining comprehension. Instead, Dennett points out several problems connected to the complex machines. What is interesting, he points out that the claim, that superintelligent machines will overtake our role as the rulers, is not the main danger. The real problem, according to Dennett, lies in the idea that we will overconfidently assign them excessive comprehension and as a result cede out authority to them, while in fact their competence would be much lesser. This seems quite the contrary to Dennett's strange inversion of reasoning: if we

endorse the idea of competence without comprehension, the comprehension both by the machines, as well as of the machines, seem a secondary question, as long as they perform in the right way. Why worry?

Of course, we would not like to deny that there is a vital problem there. The growing complexity of AI systems, which contain algorithms that are far from being transparent and clear, is a serious issue. This is especially problematic if these systems seem to perform better than human experts, although even their creators could not fully explain how do the process of decision-making realized in the machine works. The legal concerns of admissibility of such algorithms, and responsibility for their use were recently a subject of discussion in the European Union. The new General Data Protection Regulation [8], regulates "automated individual decision-making, including profiling" – in the article 22 of GDPR. This provision includes several rules, "right to explanation", obliging the creators of intelligent systems to algorithmic transparency, above all. However, the introduction of this law will not solve our problems. Conversely, current shape of GDPR raises a lot of questions, regarding for example the character of non-discrimination through algorithmic decision making, as noted by B. Goodman and S. Flaxman [9].

3 Legal Duties of Programmers and Developers

Let us now take a brief look at a more specific legal problems, which are implicitly mentioned in FBB.

3.1 Criminal law

Dennett's proposal include harsh punishments (incarceration) for creators of systems that intentionally hide their shortcuts and gaps of incompetence, that result in misperceiving those systems as humans. This would mean a criminalization of a deed of creating a system that passes a Turing Test [10] and hides it artificial identity. Creating this kind of crime raises two problems. Firstly, the criminal law is generally seen as ultima ratio – a measure to be used only as a last resort when it comes to shaping social attitudes. The negative results of overcriminalization may be varied [11], including e.g. overload of courts. The general principle of the criminal law is that the severity of punishment should match the severity of damage done to some legal values (such as, inter alia, life, health, public safety) – to put it simply: the degree of harm. It seem that a proposed crime would only protect against personal uneasiness of some people, thought that they were interacting with human. That is not a harm, which could justify depriving people of their freedom. Alternatively, Dennett's proposition could be also seen as a proxy crime [12], i.e. the concept of criminalizing a deed that is not harmful per se, but is suspicious, as it can often lead to a serious one. It this case hiding the true, artificial identity, of an agent, could lead to a severe fraud or extortion. The idea behind proxy crimes, however, is to provide an easily traceable, in the evidentiary sense, substitution for the primary crime (which are often hard to

prove due to the high evidentiary threshold – beyond any reasonable doubt). Anyhow, that is not the case here, because of two reasons: firstly, tracing the author of a malicious, fraudulent AI system could be as hard as proving fraud itself. Secondly, incarceration is often considered as the harshest measure (along the capital punishment, where it is legalized) [13], mostly reserved for serious felonies only [14]. Moreover, a question of *mens rea* should be assessed. For example, in the Polish legal system, a felony can only be committed with intent, whereas misdemeanors can also be committed negligently or recklessly, if the provisions say so. To achieve the rational level of deterrence, the proposed crime would require to show that the author of AI acted with intent in the sense of *mens rea*, which in this context may be problematic (since frequent presence of "bugs" or mistakes in programming). Criminalizing recklessness or negligence, i.e. buggy codes, would definitely mean imbalancing the degree of punishment with the blameworthiness of the crime (since frauds or extortions can be committed generally only with intent). To sum up, this approach could be considered some kind of philosophical-technological version of "penal populism" [15], apparently calming the society about the problems and threats of new technologies through overcriminalization.

3.2 Civil law

Most of Dennett's apprehensions are in relation to various branches of civil law. They could be analyzed, for instance, in two areas: capital market law (where AI algorithms are already quite popular), and consumer law (to which Dennett's propositions seem most applicable).

Capital market law. Let us take a popular example of application of Artificial Intelligence to the commercial and corporate law. Although there are many problems with machines making contracts, highlighted in a seminal article "Can computers make contracts?" by R. Allen and T. Widdison [16], concluding contracts by an algorithmic proxy became a fact. Nowadays, a lot of transactions in the stock market are not done by humans. This is a result of using so-called high frequency trading (HFT)[17] or low latency trading [18]. Both are the type of algorithms, that perform transactions automatically. It would be quite problematic to inform potential clients that they are not selling their shares to humans. Applying the rule "you should know that you are interacting with machine" would cause a lot of potential trouble: delay caused by an actor deliberating whether he want to trade with machine, necessity to divide offers into those made by humans, and those made by machines, and so on. It would probably mean the need to separate markets into human and machine, which would impair trade as a result. Moreover, it could create discriminative conditions for people that could not use a machine to perform transactions in their name.

Consumer law. Dennett argues that every advertisement (of an intelligent system) should contain a full list of all potential limits, shortcomings, gaps, and

cognitive illusions used by the system, similarly to the list of potential side-effects in drugs. Once again, we can see an important legal issue here and, once again, we can see a departure from Dennett's strange inversion of reasoning. It seems that such a list would inform a human user, who then should rationality make his choice, whether to use and interact with the agent. But that would require of a human agent advanced knowledge not only in law, but also information technology and cognitive psychology. However, it seems that imperfect generalisations would work here in a sufficient way (compare the role of folk psychology in everyday social contacts). The legal problem behind the absurdly long lists of caveats (accompanying drugs, user licenses, and other contracts) is the precision and adequacy to cover all of the outcomes. On the other hand, this makes most of them completely incomprehensible to anyone (excluding a small group of specialists). As a result, the principle of fairness of the law is put at risk. Recently, one of the truck shops in New Zealand, Zee Shop, was held responsible [19] for producing incomprehensible contracts, thus violating Credit Contracts and Consumer Finance Act. The company failed to present essential information and express them clearly and concisely. Yet, the legalese cannot be abandoned as a whole, because the precision of the legal language is often valued more than its understandability [20]. The concept of competence without comprehension comes real handy in here. As noted by Judge David Wilson in the case of Zee Shop "how people managed to navigate their way through those [clauses of contracts], remains mysterious." [19]. We cannot think of a more radiant example of competence without comprehension. Consumers in fact function, and should function, in some kind of Legal Umwelt. This environment should be tailored every-time specifically in compliance with the character of the legal institution (for instance, information duties of the producers of robots, which serve as aid to children, may differ from those, which provide services for adult users). How consumer environment should be shaped? Although providing excessive lists of exceptions may be a base for developing competence without comprehension, there may be better ways to improve it. Let us look at the example from FBB: the chess rule of 50 moves. According to the rule, if players perform 50 moves without taking a piece or moving a pawn, the game is considered a draw. It serves as a clear evidence, that players could not do anything productive, rather than endlessly repeating some sequences of moves. Although there is a theoretical exception to it, from the practical standpoint that kind of situation would be almost impossible to occur in a serious game. Hence, FIDE – World Chess Federation, decided to keep the rule in the official chess provisions. Our legislators should follow the example of FIDE. It seem like a two-step approach may be a solution. While starting to interact with an agent, we should firstly adapt pragmatic rules, which are best suited for our cognitive capabilities.They could then refer to a comprehensive list of all possible contingencies, that we know of.

4 Concluding Remarks

The issue of the regulation of new technologies, and the role of law in preventing potential dangers related to rapid development of intelligent systems cannot be overstated. Yet, the simple strategy of severe criminalization and putting burdensome duties on the creators of intelligent machines seem like a misplaced and inefficient idea, which could lead, for example, to partial paralysis of capital markets. The issue of the liability of the AI systems is more multi-layered and problematic. For example, one of the most promising ideas, include creating a new, specific kind of legal personality, tailored for AI agents [21] [22] [23].

Another one of FBB *leitmotifs*, the Second Orgel Rule, firmly states that "the evolution is smarter than you". It is not clear, why that should not be the case of the future social co-evolution of legal culture and technology, since law can be understood per se as a product of human evolution [24]. Societies as a whole, can be beneficients of free floating rationales [25]. While thinking of the legislation about the new technologies, we should note that lawmaking is not only top-down, but also a bottom-up process – the role of judiciary and legal doctrine (which may be understood in the terms of Darwinian selection of legal concepts) in this context may be also a better idea than imposing harsh regulations *a priori*.

Rather than providing every possible information on the AI agent, we should try to focus on these pieces of information that really make any difference. Which would those be, that is to determine by community of engineers, scientists and lawyers.

Acknowledgements. We would like to thank Bartosz Janik and Piotr Bystranowski, as well as two anonymous reviewers, for their insightful comments, which helped in the preparation of the paper. Additional thanks to Marta Dubowska for proofreading of the initial version of the text. We declare that no competing interests exist.

References

1. Bostrom, N.: Superintelligence. Dunod, (2017).
2. Dennett, D.: From bacteria to Bach and back: The evolution of minds. WW Norton & Company, (2017).
3. Dennett, D.: Darwin's strange inversion of reasoning. Proceedings of the National Academy of Sciences 106.Supplement 1 (2009): 10061-10065.
4. Dennett, D.: Intuition pumps and other tools for thinking. WW Norton & Company, (2013).
5. Haig, D., Dennett D. Haig's strange inversion of reasoning(Dennett) and Making sense: information interpreted as meaning (Haig). (2017).
6. Dennett, D., The free floating rationales of evolution, in "Rivista di filosofia, Rivista quadrimestrale" 2/2012, pp. 185-200, doi: 10.1413/37254
7. Gibson, J.J.: The Theory of Affordances. In: Shaw, R., Bransford, J. (eds.): Perceiving, Acting, and Knowing, Lawrence Erlbaum Hillsdale, NJ, (1977).

8. General Data Protection Regulation. http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679
9. Goodman B., Flaxman S.: European Union Regulations on Algorithmic Decision Making and a "Right to Explanation". AI Magazine. 38, 50-57 (2017).
10. Turing A.: Computing machinery and intelligence. Mind. 49, 433-460 (1950).
11. Larkin Jr., P.L.: Public Choice Theory and Overcriminalization. Harvard Journal of Law & Public Policy, 36, 715-793 (2007).
12. Bystranowski, P.: Retributivism, Consequentialism, and the Risk of Punishing the Innocent: the Troublesome Case of Proxy Crimes. Diametros. 53, 26–49 (2017) doi: 10.13153/diam.53.0.1099
13. Wróbel W., Zoll A.: Polskie prawo karne. Znak, (2010).
14. Czabanski J., Zagrozenie kara oraz orzecznictwo za najpowazniejsze przestepstwa w ujeciu miedzynarodowym. Prokuratura i Prawo. 1, 99-126 (2008).
15. Pratt, J.: Penal populism. Routledge, (2007).
16. Allen, T., Widdison, R.: Can Computers Make Contracts?.Harvard Journal of Law & Technology. 9, 25- 52 (1996).
17. Gomber, P., Haferkorn, M., High-Frequency-Trading, Business & Information Systems Engineering. 5, 97-99. (2013). https://doi.org/10.1007/s12599-013-0255-7
18. Hasbrouck J., Saar G., Low-latency trading. Journal of Financial Markets .16, 646–679, (2013).
19. Truck shop sentenced for incomprehensible contracts. (2017) http://www.comcom.govt.nz/the-commission/media-centre/media-releases/2017/truck-shop-sentenced-for-incomprehensible-contracts/
20. Skoczeń I.: Implicatures Within the Legal Context: A Rule-Based Analysis of the Possible Content of Conversational Maxims in Law, In: Araszkiewicz, M., Banas, P., Gizbert-Studnicki, T., Pleszka, K. (eds.) Problems of Normativity, Rules and Rule-Following Springer International Publishing (2015).
21. Dahiyat E.A.R., Intelligent agents and liability: is it a doctrinal problem or merely a problem of explanation? Artificial Intelligence and Law. 18, 103–121 (2010).
22. Sartor G., Cognitive automata and the law: electronic contracting and the intentionality of software agents, Artificial Intelligence and Law. 17, 253- (2009) DOI: 10.1007/s10506-009-9081-0
23. Michalczak R.: Czy programy moga miec intencje? O odpowiedzialnosci agentow programowych. In: Samonek, A. (ed.) Teoria prawa. Miedzy nowoczesnoscia a ponowoczesnoscia, Wydawnictwo Uniwersytetu Jagiellonskiego, (2012).
24. Załuski, W.: Evolutionary Theory and Legal Philosophy. Edward Elgar Publishing. (2009).
25. Schliesser, E., On the Significance of Dennett's Free-Floating Rationales for Social Science (I) (2017), http://digressionsnimpressions.typepad.com/digressionsimpressions/2017/03/on-the-significance-of-dennets-free-floating-reasons-for-social-sciene.html