FREE WILL: WHO CAN KNOW?

Zafer Kılıç
zafer.kilic@boun.edu.tr

1. Introduction

In our day, the factors behind the difficulty of settling the free will issue is mainly twofold: on the one hand, we witness the advances in social and behavioral sciences, scope of which contains human beings, continue to unravel causal explanations for human action, leaving less and less room for free will. On the other hand, our experience furnishes us with an unassailable feeling that we are free to choose what we think and do.

Given this predicament, where should we put our esteem? Which source of knowledge is more reliable if we are to obtain an answer? As implied in the title, I will inquire as to the characteristics of knowledge that is required to license us to make justifiable claims regarding the freedom of our will. Particularly, I will try to epistemically establish the importance of a priori knowledge in making justifiable claims on free will, and provide metaphysical reasons to doubt whether our knowledge is of that sort.

My thesis is that we, as human beings, cannot have the proper kind of knowledge to make justifiable claims regarding our freedom –or lack thereof– of will as per the argument below:

P1: In order for a human being to know whether they have free will or not, their understanding of the laws of the universe they inhabit should be reliable so that they can make necessarily true propositions regarding they have freedom with respect to those laws.

P2: Human understanding of the universe is based on synthetic a priori knowledge as Kant suggested. This understanding incorporates certain innate principles and experience of the world.

P3: We cannot be sure about either the reliability of our innate knowledge (whether they are provided by natural or supernatural sources) or whether our experience accurately reflects the actual universe.

C1 (from P2 and P3): Our understanding of the universe and its laws is not necessarily true.

Therefore (from P1 and C1);
C2: Human beings cannot make justifiable claims regarding whether they have free will or not.

I have come to propose the above thesis through the following stages described in the respective sections of this paper: In Section 2, I explain why I adhere to Kantian epistemology and philosophy of mind –rather than empiricism– and how Kant's views make sense with respect to cognitive science by explaining the ways Kant's certain ideas have formed the philosophical groundwork of cognitive science. In Section 3, I briefly introduce the positions in free will debate and discuss the relevance of definition of free will to my thesis. In Section 4, I elucidate my argument and present the main discussion behind it over a simulation example, along with the evaluations of some possible objections. In Section 5, I make my concluding remarks.

2. Understanding of the Mind and Cognition

Prior to any discussion regarding the matter of free will, I must specify the view of the mind I submit to. As a student of cognitive science, I accordingly favor the understanding of the mind adopted by it. In cognitive science, the mind is likened to a computer, and tried to be understood based on the computational theory of mind as a working hypothesis.

As it happens, some of the most important premises of contemporary cognitive science are based on Immanuel Kant's philosophical insights (Brook, 2014). Therefore, I think it will serve well to make my main argument more clear if I address briefly in which ways Kant's views are adopted by cognitive science and fit to the computational theory of mind. But first, I will explain my reasons as to why I deem Kant's notion of the mind and cognition more plausible compared to that of David Hume and empiricists in general.

2.1. Roots of Knowledge: Kant versus Hume

Under this chapter, I will explain why I consider Kant's theory of mind more plausible compared to Hume's empiricism. For this comparison, I picked Hume as the foremost representative of empiricism, who conveyed the essential ideas very clearly and concisely. As for Kant, though definitely not an empiricist, he cannot be deemed a rationalist either; his approach was to conjoin the two schools of thought by giving rationality and experiences their respective roles in human knowledge.

The fundamental point where Hume's and Kant's views diverge can be recognized as the following: Hume thinks our experience of the world is shaped by the world itself, while Kant thinks our experience of the world is shaped by the mind.

Indeed, a remarkable portion of Western philosophy since the end of the eighteenth century can be linked back to these rival views of Hume and Kant. I will start by briefly describing Hume's empiricist views.

Hume's account of causal connection give the central role to experience; the role of the mind is merely to associate ideas. Due to observation of constant conjunction of events, the mind develops the habit of generalizing this relation, and expects it to recur. All of which means there is no necessity in causal relations other than that which our minds ascribe to it.

As a strict empiricist, Hume maintained that the content of every thought is gleaned from the experiences which affirm it. Unless referenced to the sensory impressions, no belief can be licensed as true. In empiricism, one's knowledge of the world is restrained to knowledge of their own subjective point of view.

The problems with this empiricist picture will become apparent when we look at the current findings provided by cognitive science research after we look into Kant's opposing ideas.

Profoundly influenced by Hume's ideas, Kant was convinced that causal necessity cannot be accounted for logically or empirically. He affirmed that it was us, the subjects, who ascribed causal connection to the universe. Where he diverged from Hume is that he did not think that our belief in causality was the manifestation of mental habit, but there should be a specific source of knowledge in our minds, a priori knowledge.

Kant held that no coherent theory of (objective) properties, objects, events, causal relations, substance, time or space can be produced that is not already also an account of our (subjective) perceptions, concepts, and judgements concerning such things. In the last analysis, he concurred, the nature of the reality we know is inseparable from the nature of the mind that knows it. He contended that the "objects of the senses must conform to the constitution of our faculty of intuition." This was what Kant called his "Copernican Revolution".

According to Kant, underlying our understanding of the universe, there are basic principles and judgements such as the truths of geometry and arithmetic, the principle of causality, and the principle of object permanence (the judgement that objects do not come into existence out of nowhere, and not stop existing). Kant proposed that these have some important characteristics, such as (Bell, 2003, p. 728):

i. They are necessarily true, and cannot be either justified or falsified by appeal to contingent facts or perceptual experience (that is, they are *a priori*).
ii. They are not merely logical truths or truths by definition (that is, they are *synthetic*).
iii. They are essential to our understanding of reality.

According to Kant's distinction, the knowledge we possess of this type of truths is *synthetic a priori knowledge* and in Kant's idealism there are two premises involving it: first, no knowledge, or understanding, or meaningful experience would be possible if it was not for synthetic a priori knowledge; second, it cannot be acquired through sheer experience of the world.

There are two points about which Kant is very insistent. The first one is that there is an utter distinction between sensibility and understanding (corresponding to intuitions and concepts). The second one is that involvement of both sensibility and understanding is absolutely necessary in any knowledge that is accessible to us. That means Kant denies the possibility of gaining any knowledge that is solely sensory or, equivalently, the possibility of gaining any knowledge that is purely conceptual, as he exquisitely stated "thoughts without content are empty, intuitions without concepts are blind" (Kant 1929: B xviii).

Now, we may turn to some scientific research, which I think is relevant to comparison of Kantian and Humean understandings of the mind and cognition.

But before we delve into scientific studies, recall that Kant's contention was that there are principles (that are based on a priori concepts) which lay beneath our understanding of the universe such as intuition of physics, causality, and the principle of object permanence. Trying to understand the origins of human knowledge regarding objects, substances, and mathematical concepts, cognitive development researchers directed their enquiries toward the least cognitively developed agents they could find: young infants. If some of that knowledge could be found in young infants –who have had neither enough time nor variety of stimulus to base their understanding on experience– then it would stand to reason to think that they have some a priori knowledge of the sort Kant affirmed.

There are several studies on infant cognition but I will explain a review study due to its comprehensiveness and extensive coverage.

In a 2011 study, psychologists vanMarle and Hespos reviewed infant cognition research of last 30 years. They found that an intuitive understanding of certain physical laws were already maintained by infants that are 2 months of age, about time when they become responsive to moving objects and can be tested with eye-tracking technology (vanMarle & Hespos, 2011, p. 20). For example, infants at this age expect unsupported objects should fall and concealed objects should not cease to exist. In a test, researchers moved a container with an object placed in it and 2-month-old infants knew that the object moved with the container (ibid., p. 21). This strongly indicates that the infants possess an expectation of object permanence.

This innate knowledge of basic physics keeps developing as the infants gain experience by interacting with the world. At 5 months of age, babies start to distinguish the properties of solid objects from those of liquids (ibid., p. 21). This is also in alignment with Kant's view, according to which innate knowledge and experience together ground human understanding of the universe.

Regarding their findings, vanMarle states that they contend that infants are furnished from birth with expectations regarding the objects in their environment, and this is a skill they were never taught (Thornhill, 2012).

In addition to empirical evidence, there is an enduring theory from linguistics which I think supports and fits into Kantian picture of cognition: *universal grammar* (UG),

postulated by the renowned linguist Noam Chomsky. The fundamental postulate of UG theory is that linguistic capacities are hard-wired in our brains. In support of his theory, Chomsky remarks that the experiences available to language learners are way too scarce to account for their knowledge of their language. He thinks that we must assume that individuals possess innate knowledge of a universal grammar which captures the common structure of natural human languages to explain language acquisition (Chomsky, 1975). I think this capacity Chomsky talks about may well be one of the basic principles that Kant thought we have ingrained in our minds, hence an example of synthetic a priori knowledge. Empiricists' account of knowledge acquisition seems to fall short of explaining linguistic abilities of humans as long as the universal grammar theory continues to stand.

In light of the above discussion, I consider it reasonable to assert that the findings of infant studies and the most esteemed theory of language acquisition (UG) strongly indicate that we are endowed with a priori judgements and principles which constitute our foundation of understanding universe as Kant conceived. Contrary to what Hume thought, humans do not solely rely on observation for associating ideas, but make use of already present a priori concepts such as time, space, causality, and object permanence.

2.2. Kant and Cognitive Science

In the previous chapter, I presented some research and a theory which I think clearly supports Kantian understanding of the human mind. Those findings already indicate some aspects in which Kant's views and cognitive science are matchable. I will now explicitly point out these aspects of Kant's philosophy that found place in contemporary cognitive science. Indeed, Kant's influence is so palpable that he has been called "the intellectual grandfather of contemporary cognitive science" (Brook, 2014, p. 61).

To start with a central doctrine of Kant, which has become orthodox in cognitive science, one can mention the doctrine that *representation requires concepts as well as percepts*. Stated in more contemporary terms, the idea is that to make discrimination among things, information is required to ground the discrimination; however, for information to be useful to us, we have to apply capacities to discriminate on it (ibid., p. 65).

Kant's another influence on cognitive science has been methodological. His method of transcendental argument has become a major, if not the major, method that has been employed by cognitive scientists, as they try to infer the conditions that are necessary for some phenomenon to come about. (In contemporary terms, this method, at its core, amounts to *inference to the best explanation*.) This method is important in cognitive science as it helps forming an explanatory bridge between observable behavior and unobservable psychological antecedents (ibid., p. 65).

Currently in cognitive science, functionalism is the most widely accepted philosophical view of the mind (ibid., p. 66). The idea behind functionalism is, in essence, that to model the mind we should model what it does and can do, which means to models its functions. The main function of the mind in representational models is to shape and transform

representations. This matches very well with Kant's picture; he also held a representational model of the mind, and he also saw the mind as a system of functions that applies concepts to percepts (ibid., p. 66).

Finally, to outline the most essential ideas of Kant adopted by cognitive science (ibid., p. 67):

i. His epistemological vision that experience requires both perception and conception, that is we make sense of what we perceive according to the concepts rooted in our minds
ii. His transcendental argument method
iii. His picture of the mind that employs functions operating on concepts to shape representations

It should be noted that (i) implies that Kant's notion of synthetic a priori fares well with cognitive science, as also indicated by the infant cognition studies and universal grammar theory discussed in the previous chapter. I think, when we combine these aspects, they grant us enough reason to concur that Kant's philosophy constitutes a suitable framework for thinking about the mind and freedom of will.

3. Which Free Will are We Talking About?

In this section, I will summarize the main positions one can take with respect to free will dispute and explain their relevance to my thesis, which concerns justifiability of our claims regarding our freedom of will.

There are two main divisions to the dispute of free will: compatibilism and incompatibilism. The former holds that free will and determinism are compatible; we can have free will even though we live in a universe governed by deterministic laws. Taking the latter position, one can side with one of these two camps: libertarianism and hard determinism. Libertarians hold that we have free will which requires the laws of universe to not be deterministic. Hard determinists hold that universe is governed by deterministic laws and that denies us free will.

Considering above introduced positions, I do not see the conflict between compatibilism and incompatibilism is relevant to my project. I think, more crucial to my project is the dispute between compatibilism and libertarianism, as the latter require indeterminism for free will and the former does not. This point of divergence is relevant for my purposes as I need to specify some criterion for freedom to which I can respect in the context of my simulation example. Therefore, I am to evaluate both views considering their definitions of free will.

Compatibilists submit to determinism and they interpret "freedom" different than libertarians do, so that it can be fitted into the determinist picture. They do that by redefining "free will" in a non-metaphysical manner, as freedom to act in accordance with one's will, without any compulsion or coercion or whatnot (different versions of compatibilism offer different solutions but this is the main gist). The important point is

not whether your decisions and actions are determined, they claim; but that they are *your* decisions and actions. According to compatibilists, even though what you have done was determined, you could have done otherwise, if you had wanted otherwise (McKenna, 2012). Compatibilists hence put the emphasis on the will rather than freedom.

As opposed to compatibilists, a well-known proponent of libertarian free will Robert Kane affirms it is necessary that there be metaphysically real alternatives for our acts, but it is not sufficient; our acts could be random unless they are in our control. The control is found in what Kane calls "ultimate responsibility" (UR) (Kane, 2002). *Ultimate responsibility* requires that agents must be the ultimate originators and sustainers of their own ends and goals. There must be more than one way for a person's life can turn out. More crucially, the person's willing acts must be the arbiter of which way it turns out. Kane defines UR in detail as follows (Kane, 1996, p. 35):

> An agent is ultimately responsible for some (event or state) E's occurring only if (R) the agent is personally responsible for E's occurring in a sense which entails that something the agent voluntarily (or willingly) did or omitted either was, or causally contributed to, E's occurrence and made a difference to whether or not E occurred; and (U) for every X and Y (where X and Y represent occurrences of events and/or states) if the agent is personally responsible for X and if Y is an arche (sufficient condition, cause or motive) for X, then the agent must also be personally responsible for Y.

Considering the two different understandings of free will discussed above, I do not think the interpretation of free will in compatibilism affect the issue I am dealing with in any significant way. I am more concerned with the matter of whether you can have the knowledge regarding whether your actions are determined or not.

Therefore I will take Kane's libertarian definition of free will since it calls for the strictest criterion, which I think is the only one that accounts for our subjective experience of freedom. Besides, if an agent can justifiably confirm or refute the prospect of his having libertarian free will, I believe his judgement should also be justifiable for assessing compatibilist or any other accounts of freedom of will.

4. Can We Simulate Free Will?

In this section I present my main argument and my underlying reasoning via an example of simulation.

I argue that in order for us, as agents of a system (the universe), to make judgements about freedom of our wills, we should possess analytic a priori knowledge about laws of the universe we live in. Given that our understanding of the system is based on synthetic a priori knowledge, we are in no position to make such judgments because we do not know if we can rely on our knowledge. That is because our synthetic a priori knowledge – which is the foundation of our understanding of the universe– is not reliable since (i) we do not know its source and metaphysical properties of it; (ii) we do not know our

experience of the world accurately represents the world itself. Below I state my thesis in its argumentative form ('P's stand for premises and 'C's for conclusions):

P1: In order for a human being to know whether they have free will or not, their understanding of the laws of the universe they inhabit should be reliable so that they can make necessarily true (true in every logically possible world in which the determining laws obtain) propositions regarding they have freedom with respect to those laws.

P2: Human understanding of the universe is based on synthetic a priori knowledge as Kant suggested. This understanding incorporates certain innate principles and experience of the world.

P3: We cannot be sure about either the reliability of our innate knowledge (regardless of whether it comes from natural or supernatural sources) or whether the experience accurately reflects the actual universe.

C1 (from P2 and P3): Our understanding of the universe and its laws is not necessarily true.

Therefore (from P1 and C1);
C2: Human beings cannot make justifiable claims regarding whether they have free will or not.

Concerning Premise 2, I have already explained my reasons for submitting to Kant's epistemology and philosophy of mind in section 2. As for the rest of the premises and conclusions, I will venture to justify them in the following main discussion involving a simulation example.

4.1. Elaboration of the Main Argument through a Simulation Example

In support of my argument presented above, I will present my reasoning based heavily on a simulation example, especially regarding the relation of the programmer and the simulated agents, as well as comparison of them in terms of their free will and knowledge about it. I will look for correspondence and analogies between this simulated universe and our universe, and see if their incorporation provides helpful insights into the issue of free will.

In essence, the purpose of the simulation example is this: to investigate the possibility of free will and knowledge thereof in the frame of a relatively more comprehensible universe and hence more comprehensible epistemic relationship between its creator and agents. I will try to derive ideas that can be assessed considering the similarities (and differences) between this simulated universe and that of ours and in terms of source and justifiability of knowledge.

There are more specific and important points regarding the simulation example which are as follows:

- The simulation does not have to be based on an electronic system. The purpose is to imagine a system over which the creator has complete authority and happens to be a human so that we can relate and think at a lower order of abstraction, i.e. we do not need metaphysical explanations. To make it easier to imagine, you can think of the computer game series "the Sims" (theSims.com, 2017) or the movie series "the Matrix" (en.wikipedia.org, 2017) with a modified version of the simulation (Matrix) in which real human brains are not hooked up to the simulation as agents, instead they are also simulated as is the rest of the universe.
- The philosophical function of the example is to present us with a system with respect to which we are epistemically better-grounded compared to when we think about our own universe and free will.
- I preferred to imagine a simulation of a whole universe inhabited by agents instead of a singular agent with artificial intelligence (AI) so that we can examine the characteristics of the knowledge of the simulated agents regarding the simulated world. The simulation should be viewed as a virtual universe similar to ours, inhabited by individual AI agents with the cognitive capacities are on par with humans in the frame of computational theory of mind. We are not interested in their subjective mental experience so much as information processing aspects of their 'minds'. A practical way to qualify them would be assuming that they could pass the *Turing test* (Turing, 1950). Thence it should be noted that the AI in the simulation example is clearly idealized.
- The programmer of the simulation is different than an ordinary human being in that he has such mental capacities that he knows and can access every governing logical law of the simulation; i.e., he is idealized in the sense that he is omniscient with respect to laws of the simulation he created.

I think the above introduced simulation example accommodates a much clearer view of the epistemic relationship between the agent, the universe, and the creator (or any other phenomenon as the arbiter of laws of the world and their knowability) compared to our own case as humans. By overviewing a lower-order system –with regard to which we are epistemically more reliable than we are with regard to our own system–, we might find ways to extrapolate the insights we gain to higher-order systems such as our own universe.

Now, I can start with the depiction of the example and then move on to the discussion based on it. Consider a programmer and a simulation he has created using a sufficiently capable computer. The simulation consists of a virtual universe (similar to ours) and virtual humans with AI (hereinafter I will call them "simulants") inhabiting it. Basically, a computer carries out logical operations to run a program (such as the simulation at hand). These operations are represented digitally and realized physically by means of some electronic system. Granted, the programmer may not possess the knowledge of all possible physical states of the system, but the electronic system can be precisely expressed as logical operations, i.e. the whole system can be defined in logical terms. Therefore, having written the code, the programmer knows a priori every logical sentence (in practice he would rather know and use compiled statements via a

programming language, which is a means to implements logical operations, but that is not a problem for our purposes) that defines the system. This means he does not need any experience to know that the simulation will work as he programmed it. Since his understanding of the system is based on a priori knowledge, his propositions based on those logical statements would be necessarily true. Thus the programmer is justified in making claims regarding the simulants' freedom of will, regardless of whether his own will is free or not. At least, he would be justified in claiming that the simulants do or do not have free will relative to the system of which he is an agent, i.e., our universe. Having proposed this, I need to specify some criterion according to which the programmer can make judgements about the simulants' freedom of will. In compliance with Kane's libertarian free will definition I previously stated I would respect (see section 3), I specify the criterion as the following: a simulant can be declared to have free will if he is *ultimately responsible* for some event that deviates from the laws of the simulation determined by the programmer.

Based on the above specified criterion, the knowledge of the programmer would allow him to make claims regarding the simulated agents by observing the compatibility of their behavior with the laws he set, hence knows a priori. Now, concerning our universe and our freedom of will, a candidate who may have that kind of justification could be some entity like Laplace's demon, which is conjectured to see the future and the past of the universe with certainty as he knows all the governing laws of our universe and the momentary positions of all the things in it (Laplace, 1820). I argue only such an entity would have a status –with respect to our universe– that is epistemically equal to the status –with respect to his simulated universe– of the programmer. That brings us to the point of the discussion I have made thus far (and the ground of the first premise of my argument): I do not think any human being can possess or access to such knowledge. Even if we granted a human being a perfect understanding of the laws of the universe, his knowledge is not as reliable as in the manner the programmer's knowledge is as per the Kantian framework I have submitted to. The programmer determined the laws of the simulation using nothing but his own knowledge, so he is justified in expecting the simulation to obey his rules as he has a priori knowledge regarding the simulation. This means that the simulation is a logical manifestation of the knowledge that is in the mind of the programmer; his knowledge is not contingent. Such precise correspondence is not the case for the relation between human knowledge and human universe; however accurate a human's knowledge of the universe may be, there is reason to doubt it since it is not necessarily true.

Now we can consider the issue further by taking the viewpoint of the simulants so as to compare with our situation with respect to our universe. It seems to me that the question "What can the simulants know as to the nature of the simulation they inhabit?" would not give way for a promising investigation in this case, since the programmer has directly determined the laws of the simulation and thus has the power to dictate the limits of knowledge of the simulants. A more productive question may be "Is it possible for the programmer to design a simulation in which the simulants may possess or access to the knowledge which would allow them to precisely understand the simulation?" In order to do that, the programmer must somehow help the simulants to see the

simulation the way he sees it. Note that the programmer has the distinct advantage of not being a part of the simulation; he does not need to self-refer to understand it. He would have to ingrain the electronic concepts as well as the logical concepts into the minds of the simulants, whose existence is based on the system that has been founded on those very concepts. That would be akin to us trying to understand the very concept that provides us with our mental faculties, as we are somehow trying to do throughout this paper. If we came to know the exact structure of our brains, would that knowledge afford us the understanding of our mental phenomena? Or, likewise, if the simulants understood the exact structure of the electronic and computational system on which the simulation is running, could they understand their own experience?

In search for answers to the above posed questions, we could inspect the issue through a Kantian lens. In Kantian terms, one could say the programmer knows about the noumenal world (the world that includes computation and electronics), whereas the simulants have only knowledge regarding the phenomenal world (the one they experience within the simulation). In the Kantian picture, our mental faculties only provide us with knowledge of phenomenal world (things as they appear to us) not of noumenal world (thing-in-themselves). We have concepts based on judgments of perception, which are subjective; however, judgments of experience we make according to those concepts are objective. Yet, their objectivity only implies their necessary universal validity in the phenomenal world. This may or may not hold true for the simulants as the programmer can furnish them with such concepts that their experience reflects nothing like the simulated world they live in. I leave further inquiry to the next chapter, in which I will deal with an objection that can be raised in this vein.

Now, I will offer my reasoning behind Premise 3 which purports that we cannot be sure about the reliability of our synthetic a priori knowledge, which is innate. The reasons for doubting synthetic a priori knowledge would depend on the ontological stance we take with regards to the source of that knowledge possessed by humans: naturalism or supernaturalism. Let us investigate the issue of reliability of knowledge from both perspectives.

In the naturalist picture, I think the most plausible candidate for the source of synthetic a priori knowledge is evolution for it is the best scientific theory we currently have, which affords us great deal of explanatory power regarding life on earth. The mechanism of evolution is such that the characteristics (this includes inherited knowledge like synthetic a priori knowledge which is of interest to us) of populations are passed on through generations based on their fitness value. It can be plausibly asserted that fitness value does not necessarily correlate with accuracy of the knowledge[1]. It may probably be the case that the knowledge underlies our understanding of the universe is selected so as to facilitate our survival even though it does not provide us with accurate understanding. (This claim can be supported by many well-studied shortcomings of the human mind, which are generally known as "cognitive biases," as termed by Kahneman

[1] Alvin Plantinga has proposed an argument against naturalism by appealing to evolution in a similar manner (Plantinga, 1993).

& Tversky, 1974.) Consequently, in the naturalist picture, we can identify at least one aspect that grants us reason to doubt our synthetic a priori knowledge.

As to the super-naturalist side, one shall consider we are created by a divine being and he is the one who bestowed us with synthetic a priori knowledge. Now, we cannot really know the intentions or the agenda of this divine being. He might have put the knowledge in our minds to make our experiences more pleasant, to make us worship him, and so on; the possibilities are endless. Thus we can conclude that it is also the case for the super-naturalist that he cannot rely on the accuracy of his synthetic a priori knowledge, or the understanding he has developed based on it.

4.2. Assessment of Possible Objections

Now I will evaluate some possible objections that can be brought forward against my argument.

To start with, I suppose an objection can be formed by appealing to some views of Kant himself, several ideas of whom I adopted for the purposes of my argument. I have already conceded that humans have synthetic a priori knowledge; however, I disagree with Kant on its metaphysical status. But others might not.

For instance, one can question whether a human agent really ought to have the sort of knowledge which Laplace's demon possesses in order to justifiably make claims about his free will. Let us consider this question in relation with Kant's views. Kant held that we know a priori that all events are determined in the phenomenal world. Along this line, it can thus be argued that we can be justified if we claimed we have not free will, not in the libertarian sense anyway. Thus it would follow that we do not need to specifically know the laws of nature –neither the simulants need to know the laws of the simulation–, only that they are determined.

According to what I understand, Kant considers our freedom that we experience in the phenomenal world to be unproblematic. We may feel like we can introduce spontaneity into the causal chains to conduct experiments. According to Kant, this freedom does not qualify as absolute freedom, but a sort of "second causality" to ground moral imperatives. Therefore, our freedom can be seen as an uncaused spontaneous cause among the phenomenal world, but it is not determined by the phenomenal world. Thus we have moral duties, Kant concurs.

However, I see some difficulties here. This phenomenal experience gives us no knowledge regarding the source, limit, or demands of our experiential freedom. It is allegedly not bounded by practical causality and hence it must be grounded in the noumenal realm. There it can be subject to "logical relations" aside from temporally causal relations. The will or "soul", like all the other things, has a double existence as phenomenal and as a "thing-in-itself". But having affirmed that we have no access to entities of the noumenal world, how can Kant know? Or more relevant to my thesis, how can any agent know and make justifiable claims regarding their freedom of will? Let us first see about Kant's answers and then I will present my answers and evaluation.

In the Preface to Critique of Pure Reason (2E), Kant admits that we cannot know we have that freedom since we cannot know about things-in-themselves. However, he claims, we can nevertheless think them hypothetically, provided that they are not self-contradicting.

To sum up, in Kant's picture, the grounds by which we can infer our freedom are the following: (i) we feel this freedom in our phenomenal experience; (ii) it is not self-contradicting–that is, according to our available understanding which is based on the phenomenal world through our synthetic a priori concepts ("pure categories of understanding"); and (iii) as the source of our moral duties deduced via transcendental method.

To criticize them in order, I think (i) is not a solid ground as phenomenal experiences may not accurately reflect the noumenal world, though we may concede it at least gives us a reason to expect that we have free will. As for (ii), I must point out we can have countless non-contradictory thoughts which may not be representative of the noumenal realm. Their being non-contradictory depends on the credibility we assign to the source of our synthetic a priori knowledge which shapes our judgments (recall the discussion in the previous chapter concerning the reliability of the source of this knowledge in the cases we assume naturalist or super-naturalist views). Aside from faith based justifications, I do not see how we can justifiably rely on them. Finally, (iii) requires us to give moral duties the unique status Kant gives them; however, I do not see why our moral intuitions would not be susceptible to doubt in the same way our other concepts are, especially if we assume a naturalist viewpoint rather than a super-naturalist one.

If we get back to our simulation example, the programmer may provide the simulants with all the three grounds –except for (i), which is a subjective mental experience hence cannot be accounted for in computational theory of mind– offered by Kant to grant us justification for deducing free of will, without actually allowing them any free will. He can adjust the AI of the simulants so that the idea of free will is not self-contradicting (for example, by not furnishing them with a notion of determined universe). He also can implement moral duties as strict constraints in their concepts, as hard rules they should obey, regardless of whether they have free will or not. Moreover, I think it is conceivable that an assumed creator of our universe could employ similar methods in designing us.

To conclude the assessment of this objection, Kant's views in this matter seem to me suffering from self-serving bias; they aptly serve to Kant's project which is to be make justifiable rational claims regarding freedom of will and practical reason, to evade epistemological and moral chaos. Therefore, I do not deem the above discussed objection based on these views reasonable.

Next, I want to touch upon Leibniz's compatibilism as it can constitute the foundation for another possible objection to my thesis. In Leibniz's view, God determines human action, but he claims that they are nonetheless free, in the sense that the actions are not necessary –hence they are contingent–; the contrary of some action could be done in another possible world (Look, 2013). Let us examine how this approach can be adapted to our simulation example. Firstly, in the simulation picture, the programmer takes the

role of God; that is the most apparent translation. Now, following Leibniz's argument, one can claim that actions of the simulants are free as they can be programmed to do otherwise in another possible simulation. But I find this claim untenable, for the reasons I will explain. I think the critical point of difference between a Leibnizian universe and the simulated universe is the method of creation employed by the creator. In the Leibnizian picture, each individual substance has a complete individual concept and that concept contains all predicates that are true of the individual's past, present, and future (ibid.). (I think it is clear that this is hardly the case for our simulants). Therefore, I understand Leibniz contends God created humans in an exceptional manner, as he wrote, "For they are not bound by any certain subordinate laws of the universe, but act as it were by a private miracle" (Lawrenz, 2010, p.143). Hence, I think, it follows that they are not conceptually inherent with respect to the universe. By "conceptually inherent," I mean that they can be analyzed into concepts of the universe they take place and can be explained by those. (So, "not conceptually inherent" should be considered to imply, for example, that they cannot be the products of some process that occurs in the universe –such as evolution–, hence essentially products of the universe itself.) However, on the contrary, the simulants are conceptually inherent to the simulation; they can be formed as discrete units, defined via certain logical propositions –bounded by the logical laws of the simulation–; therefore they are still based on the same substrate (physically and logically) as the entirety of the simulated world. The implication is that the programmer cannot make an exception –with respect to general laws of the simulation– for any simulant. He can alter them, but still based on the laws he programmed. He cannot, in Leibniz's words, "incline [their] souls without necessitating them," (Look, 2013) for any inclination he causes in the simulants would be a necessary inclination since it is caused based on the laws of the simulation. Moreover, it might not be the case that humans are created in an exceptional manner in our universe either, unless you concede Leibniz's particular theistic views. If you assume a naturalistic viewpoint, the arguments I made regarding the simulated universe become valid for our universe as well. Furthermore, they would remain true even if you assume a theistic view which only dictates the divine creation of a lawful deterministic universe, without making an exception for humans. Consequently, barring his particular theistic assumptions, Leibniz's compatibilism does not seem to be applicable in a deterministic world; be it our universe, or a simulated universe.

Regarding Lebniz's ideas, another way of inspecting free will in the context of simulation example would be to imagine that the programmer created the simulants guided by their complete concepts. Hence, in itself, actions of a simulant would be contingent; as it would be necessary that if the programmer creates the simulants by instantiating his concept guided program, those actions will be performed. Then the question I must answer would be whether this action can be regarded as free? Now, in the computational framework, here is how I imagine an agent's "complete concept" would be formed: The programmer defines a set of modal propositions which contain all possible predicates of the simulant in all possible conditions (I think this is the way the actions become contingent, given some other conditions in another possible simulation the simulant would act differently). In this this way, I can say, at best, he furnished the simulant with a will, but I do not think that will is free (a classical compatibilist would probably say it is free, though,

affirming that the complete concept of an agent is who he is, it is what constitutes his "will" and he is free to act as he wills). Because, still, the programmer will never be surprised by the actions of the simulants.

Another way of objecting to my thesis could be established on the possibility that the programmer may transmit his complete knowledge of the simulation to the simulants so that they can unmistakably know whether they have free will or not, hence can be justified to make claims about it. Doing that in the design stage seems problematic as I discussed in the answer to the previous example, so now let us consider another way, which would be applicable to us humans as well.

Suppose, a special simulant that is devised and controlled by the programmer himself, an avatar to represent him in the simulation (hereinafter to be called "the Avatar"), is introduced into the simulated world. Through his avatar, the programmer tries to convince some regular simulant (I will refer to him as "the Chosen One") that the world he lives in is a simulation which is totally determined. How would he convince the agent? Also as importantly, would the agent be justified to believe him?[2]

Note that the Chosen One knows nothing of the world outside the simulation; for him, his simulated world is as real as it gets, provided that the programmer designed the simulants in such a way that their beliefs are somehow in correspondence with the world they live in. Introduction of the Avatar accommodates a potential for deviations in the laws of the simulated world. Because now that the simulated world and the 'real' (the programmer's) world are causally linked, there ought to be essentially same order of freedom in both worlds. However, since the Chosen One's appraisal of experience is still governed by his concepts, the unlawful events caused by the Avatar would be proper miracles to him, just as the information that contradicts with his knowledge regarding his universe. If aptly programmed, he could come to believe the Avatar and he would be justified to do so. And is that not what we humans would demand to believe some prophet (claiming to be God's avatar) who tries to convince us of a divine being –if it was not for our capacity to believe on faith? (Though, still, holy books of many religions offer numerous accounts of past miracles as reasons to believe in their gods; and it makes metaphysical sense.) And I think that is a capacity which distinguishes us from the simulants who, having artificial intelligence, are bounded by the computational theory of mind; they cannot be expected to believe anything that contradicts their categories of judgement. But we can say this for the Chosen One because we know that his mind is

[2] By definition of our simulation, it is not as though the Chosen One is free whether to believe the Avatar or not. But we granted the simulants artificial intelligence and as per computational model of the mind, they process information then assume mental states and conduct actions depending on their concepts of understanding, experience, and individual traits that were specified by the programmer. Thus, the Chosen One's beliefs are not free but contingent –they were necessary in the absence of the Avatar since all the information and experience that was accessible to him was determined by the laws of simulation then– in the sense that they depend on his concepts and the information he is exposed to. (His concepts are still as they were determined by the programmer, but the information available to him is not, as it can come from outside the simulation now through the Avatar.) The Avatar is simply trying to make him assume a mental state in which he believes that his world is simulated and hence determined by exposing him to specific information and/or incidents.

also governed by the laws of the simulation. If we happened to witness a genuine miracle, our whole understanding of the universe would become doubtable; let alone our sense of freedom. But, still, we would end up epistemically better off, as we will have been freed of our incorrect beliefs regarding the laws of the universe. But this enlightenment would work only in that way of negation, as the inaccurate beliefs dissolve as they are contradicted by experiential and/or epistemic evidence that we witness. Because having witnessed that our knowledge of the universe was illusory or incomplete, we would have reason to doubt if our understanding is complete on the same grounds as we begin with. Thus, the conclusion I arrive at is this: there is no amount of knowledge, except for the complete knowledge, that can allow us to make unmistaken inferences and justifiable claims regarding the nature of our being –including our free will.

5. Conclusion

I have inquired as to what sort of knowledge humans need to make justifiable claims regarding free will. I defended the thesis that humans do not have the sort of knowledge which would allow them to make such claims. Adopting the view of mind based on cognitive science and Kant's philosophy of mind, first I laid out the characteristics of that knowledge with the help of a simulation example I devised. Then, upon investigating the epistemic relations between the different sources of knowledge and the agents of a system (such as the relation between the programmer and the simulated agents as well as god and humans), I claimed that knowledge bearing those characteristics cannot be accessible to human beings.

I think discussions about what we can and cannot know about freedom of will can guide us in where to put our philosophical and scientific efforts for future studies, and hopefully help us achieve a more accurate understanding of many issues that are intertwined with freedom of will.

References

Bell, D. (2003). Kant. In Bunnin, N. & Tsui-James, E. P. (Ed.), *The blackwell companion to philosophy* (2nd ed.). Oxford: Blackwell Publishing.

Brandon C. Look: Leibniz's Modal Metaphysics. The Stanford Encyclopedia of Philosophy. 2013

Brook, A. (2014). Kant and cognitive science. *Estudos Kantianos [EK]*, *2*(02).

Chomsky, N. (1975). *Recent Contributions to the Theory of Innate Ideas*, reprinted in S. Stitch (ed.), Innate Ideas, Berkeley, CA: California University Press.

En.wikipedia.org. (2017). *The Matrix*. [online] Available at: https://en.wikipedia.org/wiki/The_Matrix [Accessed 11 Jan. 2017].

Hespos, J. & vanMarle, K. (2011). *Physics for infants: characterizing the origins of knowledge about objects, substances, and number*. WIREs Cogn Sci 2012, 3: 19-27. doi: 10.1002/wcs.157

Kane, R. (1996). *The significance of free will.* Oxford: Oxford University Press

Kane, R. (2002). *Free will*. Malden, MA: Blackwell.

Kant, I., 1929 [1781, 1787]: *Critique of Pure Reason* (1st edn (A), 1781, 2nd edn (B), 1787) (translated by N. Kemp Smith). London: Macmillan.

Laplace, P. S. *A Philosophical Essay on Probabilities*, translated into English from the original French 6th ed. by Truscott,F.W. and Emory,F.L., Dover Publications (New York, 1951) p.4

Lawrenz, J. (2010). *Leibniz: the nature of reality and the reality of nature: a study of Leibniz's double-aspect ontology and the labyrinth of the continuum.* Newcastle: Cambridge Scholars.

McKenna, M. (2012). Contemporary Compatibilism: Mesh Theories and Reasons-Responsive Theories. In Kane, R. (ed.) *The Oxford Handbook of Free Will: Second Edition.* NY: Oxford University Press.

Plantinga, A. (1993). *Warrant and Proper Function.* New York: Oxford University Press. doi:10.1093/0195078640.001.0001. ISBN 0-19-507864-0.

Thesims.com. (2017). *The Sims*. [online] Available at: https://www.thesims.com/ [Accessed 11 Jan. 2017].

Thornhill, T. (2012). Coogee coo-blimey: Babies are born 'with knowledge of intuitive physics'. Retrieved January 02, 2017, from http://www.dailymail.co.uk/sciencetech/article-2091217/Babies-born-knowledge-intuitive-physics.html

Turing, A. (1950). *Computing Machinery and Intelligence*, Mind, LIX (236): 433–460, doi:10.1093/mind/LIX.236.433, ISSN 0026-4423, retrieved 2008-08-18

Tversky, A. & Kahneman, D. (1974). *Judgement under uncertainty: Heuristics and biases.* Science. 185 (4157): 1124–1131. doi:10.1126/science.185.4157.1124. PMID 17835457.