

**On the impossibility of free will
in a probabilistic universe**

by

Olav Torvik Sem



The University of Bergen

FILO350

Supervisor: Sorin Ioan Bangu

Department of Philosophy

Spring 2024

Table of contents

Introduction 4

Part 1: Is free will metaphysically possible? 5

1.1 A Strawsonian account of metaphysical challenges 5

1.2 Narrowing down the position of the free will optimist – why we live in a probabilistic universe 8

1.3 An objection from partial responsibility 15

1.4 An objection anticipated – the claim of existence being gradual 21

1.5 Identifying general misconceptions about free will scepticism 24

Part 2: Why a probabilistic universe does not point to the existence of free will 30

2.1 A theory of probabilism 30

2.2 A theory of probabilism continued 37

Part 3: How could the free will optimist resurrect his case for the existence of probabilistic free will? 43

3.1 The “could have done otherwise”-condition and transworldly committedness 43

3.2 Predictability and patterns 50

3.3 How does the free will optimist give a coherent account of internal control of the will? 53

3.4. Elaborating on control over the will – the importance of desires 65

Conclusion and final thoughts 69

Bibliography 71

I first wish to thank my mom,
Who is always warm-hearted and calm
And subsequently also my dad,
To whose existence I am awfully glad
And thank you my little brother
I would never want no other
I now want to thank all my friends
For all my worries they cleanse
Thanks Seb, Bas, Ed, and Jo
O, I appreciate you all - please know
And finally I will thank Sorin, of course
A most reliable and knowledgeable source
My supervisor so helpful and clear
Your guidance, I will always hold dear
And in a possible world,
Far away hurled
I would thank my wife
The highlight of my life

On the impossibility of free will in a probabilistic universe

- **Introduction:**

The subject, structure, and progression of the thesis

In broad terms, the aim of this thesis is to explore the question of whether free will is possible. More specifically, I will – primarily - defend two separate claims. With respect to this thesis' structure, it will consist of three parts.

The first claim is that it is difficult to see how free will could exist in any possible world, and I will in this thesis attempt to show that there are substantial problems with establishing a coherent theory of free will. Essentially, I focus on an argument for the metaphysical impossibility of free will (as originally presented by Galen Strawson). Part 1 of the thesis will thus explore the question of whether free will, metaphysically, can be said to exist. I will in this thesis' first part also touch upon some general misconceptions about free will scepticism, as it seems most reasonable to introduce these misconceptions early in the thesis.

The second claim is that the existence of free will, in the actual world, seems impossible. That is, it seems impossible given the conclusions I arrive at. Among other things, I defend the claim that the universe is probabilistic, and that the factors of indeterminism necessarily present in a probabilistic universe are better accounted for by what I label “chancy components”, as opposed to what can be labelled “free will components”. Thus, part 2 of the thesis will explore the question of the possibility of free will in a probabilistic universe.

After having defended the two claims above (that is, that free will seems unlikely both metaphysically as well as in our actual world), I will in Part 3 discuss some objections and arguments which may or may not “resurrect” the case for the existence of free will in probabilistic universes. I conclude here that although there are certainly philosophically respectable ways to defend the possibility for the existence of probabilistic free will, I am not convinced that the arguments for this conclusion outweighs the reasons we have to be sceptical of the possibility of there existing free agents in probabilistic universes. Stated more clearly, I do think that no arguments in favour of the existence of free will in our actual world - or indeed any possible world - stands. I will attempt to show that all the arguments against the impossibility of free will in our world, and possible worlds, either are subject to convincing counterarguments, or leaves us - at best - in a position of being agnostic as to the possibility of free will.

Although some of my arguments, if valid and sound, seem to dictate that I must arrive at the conclusion that free will is indeed *impossible*, I wish to emphasize that this thesis primarily is focused on pointing out the *unlikelihood* of free will. Or, as its title strongly suggests, that there are indeed many problems with establishing a coherent and convincing account of free will. What I argue is that the cumulative case of all of the arguments presented in this thesis seems to point to the fact of free will *not* existing. I do, however, wish to merely claim that – as far as I can tell – free will appears unlikely. While I believe that free will is indeed not possible, my main goal in this thesis is not to prove this, but merely to show some of the reasons why I find it hard to believe that we are indeed free agents.

- **Part 1:**

Is free will metaphysically possible?

In this section I will lay out, and subsequently discuss, Strawson's argument(s) against the existence of free will. Strawson claims not only that moral responsibility – and free will - is lacking in the actual world, but that it is also *metaphysically impossible*. He thus claims that in no possible or conceivable world can there be such a thing as freedom of the will.

1.1 A Strawsonian account of metaphysical challenges

The simplest way I think one can summarize the basic argument - as originally brought forth by Strawson¹ - is in the following way.

P1: S does what S does because of who S is²

P2: But S cannot be ultimately responsible for who S is

C: Therefore, S cannot be ultimately responsible for what S does

¹ Strawson 1994, 13-14.

² What P1 entails is that it takes into account every state of affairs which is true of S' desires, "wants" and mental characteristics, at *t*.

The main point Strawson relies on, I believe, is the second premiss. He claims that there seems to be a fundamental impossibility in being responsible for who one is, at any time. We act upon preferences that are the result of previous mental states, and those previous mental states are again reliant on yet another series of mental states, and so on. So, to be responsible for action A, at *t*, one must be in control of *who one is*, and what one's mental states are, at *t*. But due to the infinite regress this would imply, it seems impossible for an agent to be responsible for what characterizes him, or for what he "is".

One could of course ask how it can be demonstrated that an infinite regress of "self-determination" would be necessary to be truly responsible for who one is. The point is that given P1, S must be able to decide what S' preferences, inclinations and "wants" are, to be able to act truly freely. But to decide this at any point in the regress, S would already need preferences to begin with. The absurdity of this is thus that one cannot truly decide on the crucial aspects of one's nature without already being a certain way. Hence, an infinite regress would be required in order to set up one's preferences without these preferences themselves being outside of the agent's control.

Strawson further argues that in order to be responsible for who one is – due to the infinite regress mentioned – one would need to be one's own cause (*causa sui*).³ This would – on the face of it - presumably escape the problem of having to exist prior to oneself, in a temporal sense, to be ultimately responsible for one's nature. But nothing, Strawson argues, can be *causa sui* (especially not human beings). Because in order to be one's own cause, one would of course *already* have to exist. And, thus, the idea of *causa sui* presents us with the ultimate bootstrap paradox, and what Nietzsche (as Strawson notes) described as "the best self-contradiction that has been conceived so far".⁴

From this Strawson argues that:

- (i) It is impossible to be *causa sui*
- (ii) One would need to be *causa sui* in order to be ultimately morally responsible
- (iii) From which it follows that one cannot be ultimately morally responsible

³ Strawson 1994, 5.

⁴ Strawson 1994, 15.

In relation to C and (iii), Strawson contends that free will is impossible because no matter the explanation for whatever it is that we do, the explanation in question can only come from a binary set of possibilities. Either we act as we do because of randomness, or we act because of “non-random influences for which we are not responsible”.⁵ Strawson’s position is that it seems plausible that a deterministic universe is incompatible with free will. At the very least, one cannot be free in the sense of *making a choice* - between A and B, at *t*, say - if the universe is determined.⁶

On *indeterminism*, in the form of complete randomness, it seems uncontroversial that one cannot have any kind of meaningful freedom of choice. Strawson’s conclusion is that in either case, whether the universe is governed by randomness, determinism, or some combination (such as what one can call *probabilism*⁷), the concept of free will is an inherently incoherent one. And the reason is, again, that there is a fundamental impossibility in the form of ultimate responsibility, which is dependent upon self-determination of one’s mental states – which again accounts for an agent’s actions.

The impossibility of free will, as Strawson sees it, is about questioning the claim that an agent can be responsible for *either* the totality of his mental constitution, a crucial portion of the sufficient reason for an action happening, or a mere “sliver-mental constitution”.⁸ Either is impossible on the basic argument, because it is the impossibility of *ultimate* responsibility to any degree, or of anything, which is what the basic argument holds. Hence, Strawson does not claim that one cannot be responsible *enough*.⁹ Rather, what he is saying is that a statement like this does not make much sense. Strawson’s claim is of a more absolute nature. More specifically, *any* imaginable account of ultimate responsibility is, on the basic argument, simply impossible. As Michael Istvan notes on Strawson’s position: “In essence, all Strawson demands of an ultimately responsible-candidate is that A (an action) not flow entirely from a not-up-to-him mental constitution”¹⁰.

⁵ Strawson 1994, 19.

⁶ The compatibilist will of course object that viewing free will as a matter of the ability to «could have done otherwise» is a semantical error, and that freedom only entails not having any external restrictions upon one completing one’s desires.

⁷ I will define this term shortly.

⁸ That is, *any* mental constitution.

⁹ In a way this would seem harder to defend than what the basic argument contends, if only for the arbitrary point at which one would have to draw the line of “responsible enough”.

¹⁰ Istvan 2011, 402.

It is important to avoid the strawman of assuming that the basic argument demands a certain, arbitrary threshold of mental constitutions outside of the agent's control (in order to establish the non-existence of free will), as in the sense of needing the majority, or even a minority of the mental constitutions, to be ultimately traced back to the agent's own free will. The point of the basic argument is to show that there are indeed *no* mental constitutions for which an agent can be ultimately responsible. And accordingly, what is meant by statements like "the impossibility to be ultimately responsible for action A", is simply that there is no mental constitution involved in bringing about A that the subject, S, could possibly be ultimately responsible for.

1.2 Narrowing down the position of the free will optimist – why we live in a probabilistic universe

Why the claim of the impossibility of free will is intuitively compelling

I will argue that the free will sceptic's position is *prima facie* persuasive, partly because I think that some of the implications or related claims of Strawson's arguments are in themselves persuasive. The sceptic's position (such as Strawson's) highlights the difficulty of understanding how exactly free will - as a coherent concept - could exist in any possible world. I think there is, initially, at least one significant problem with the idea of free will - and that is what one might call its "metaphysical characteristics". It is difficult to see how freedom of the will would supposedly function. By this I mean that it is difficult to see how free will could exist in any conceivable universe, where *either* of the following descriptions would be accurate of the universe as a whole.

1. The universe is deterministic
2. The universe is random
3. The universe is probabilistic

Probabilism, randomness and determinism

Before discussing, and elaborating, on 1., 2., and 3., I will clarify what I mean by determinism, randomness and probabilism; respectively.

Determinism

By determinism, I mean that every single event, and every set of events, can only happen in one way (given certain initial conditions)¹¹. Meaning, that for event A, event B must necessarily follow. And for event A, B, and C, say, every possible arrangement of the events could only take one form. For instance, “A is followed by B, which in turn is followed by C”. Not only would there only be one possible arrangement of events – which would take place at set time slices – the events themselves are the way they are by default; meaning that they both: a) must take place or “happen”, and b) must, following a), take place in a given order with respect to each other.

Thus, not only would determinism entail “A, B, and C”, but also *specifically* that particular order and arrangement of the mentioned events. As a clarification, it is merely a necessary, but not sufficient, condition of determinism that events A, B, and C (in no particular order or arrangement) take place in the universe. Determinism also entails a truth condition by way of *the order in which the events happen*. Thus, both the events themselves - as well as the order of the events – can only happen in one way.

Randomness

By randomness I mean, unsurprisingly, something very different from determinism. Randomness, at least as used in this thesis, entails that in a set of time slices, every possible t is equally likely to entail any possible event. In other words, randomness would entail that event A, which itself was brought about, could itself be followed by any possible event.

An interesting implication of every possible event being equally likely to happen is that unless there is an infinite number of possible events, each event in a random universe would actually have a given “chance” of becoming actualized or real. The reason for this is that randomness by definition means that each possible event is equally likely to happen at any given time

¹¹ One such initial condition, when it comes to our universe, appears to be the big bang.

slice. For simplicity's sake, we can imagine that there are 100 metaphysically possible events (or arrangements of atoms, for instance) that could take place in the universe. In that case, it would be true that even granted complete randomness, we would seemingly also have to grant that event 1, for instance, would have a 1 percent chance of happening at a given t ; and in fact in every time slice – since each event at every time slice would be characterized by complete randomness.¹²

Randomness would in this way imply what could be called “raw chance”, since any possible event in a random universe would be equally likely at every time slice to be actualized. In short, a universe characterized by randomness would entail that everything metaphysically possible can happen at any possible time. Further, such a universe would in no way be systematic or pattern-based with respect to the events that would happen; nor their likelihood of happening (outside of any possible event having the same likelihood of occurring).

One must note, however, that a random universe could have the *appearance* of being pattern based. For given complete randomness it is not impossible that the same event happens twice; such as both at t_1 and t_2 , for instance. In theory, a truly fair dice¹³ could – even in a random universe - produce nothing but the number six, for an infinite series of time slices. In fact, pure randomness necessitates the possibility of patterns being possible (since the randomness itself must involve all possible outcomes being conceivable). Hence, it follows that the appearance of patterns must be possible in a random universe. These concerns are of course only “theoretical” ones, since it is overwhelmingly likely that a universe characterized by pure randomness would look very random indeed.

Probabilism

Unlike determinism and randomness, probabilism appears more dubious; and is thus the least intuitive and most controversial of the three terms to define.

Firstly, we must clarify what probabilism is not. Let us imagine a scale, from 0 to 10, say, which represents the probability of a given event taking place. 0 can represent that we know

¹² In this way one could also demonstrate that it is actually possible that apparent probability and patterns could still be due to processes governed by complete randomness. The simplest way to visualize this is the truth of the claim that it is possible to get six eyes a million times in a row, say, when rolling a dice (assuming the dice is completely fair).

¹³ Meaning a dice where the probability of getting any possible number from one to six is *equally likely* every time the dice is thrown.

an event will *not* take place, and so 10 can represent that we know an event *will* in fact take place. On determinism, every possible event will have the “value” of either 0 or 10; meaning that a given event will either certainly take place, or certainly not take place. This would mean that every possible event is equated with every actual (or to *be* actualized) event in the universe; and thus every actual event is a 10 on the scale of probability in a deterministic universe. In a random universe, every possible event – in isolation¹⁴ – would have a value of 5; meaning that every event would be just as likely (in isolation) to occur as not to occur.

What then, of probabilism? Probabilism would (at least in this thesis) describe a universe where any possible combination of the values 1 to 9 were true for any event or events¹⁵; *except* a scenario where every possible event had a value of 5. In other words, probabilism is true for any event(s) where the outcome is not certain, but also not completely random. In the same way that a universe would not be deterministic if one event in that universe were to *not* be determined (that is, not with a value of 10), a seemingly random universe with a single event *not* having a value of 5 (despite every other event randomly having that value), would not be random. It would then more accurately be described as a probabilistic universe, just as a seemingly deterministic universe would not, after all, be determined if one chancy event took place within that universe.

In light of the distinctions made, we could argue for the truth of either 1., 2., or 3. (that the universe is either deterministic, random, or probabilistic) in the following way.

- I. Either an action (or an event) is determined, or it is not
- II. If it is not, then it is either random or not
- III. If it is not random, and not deterministic, then it must be probabilistic (*or* something akin to not being determined and not being random)

¹⁴ With respect to itself, and not taking into account one specific event’s probability of occurring instead of *any other possible event occurring*. Meaning, every specific event, at every specific *t*, has a value of 5.

¹⁵ Here, every event is neither determined to happen, nor to not happen. It is thus probabilistic in that the events both could, and could not take place; i.e. they have a certain *probability* of occurring.

- IV. But from (III) it would seem that there is no alternative to something being probabilistic. For probabilism must simply mean every conceivable state of affairs, say, *which is neither determined, nor random*. From this it appears that probabilism is the only possible alternative to the other two alternatives of determinism and randomness.
- V. And thus, we arrive at the conclusion that free will (that is, on any possible understanding¹⁶) must be deterministic, random, or probabilistic.

There does not seem to be a fourth alternative to (1., 2., and 3.). Presumably, the free will libertarian would want to place himself in the third category. But even with respect to 3., the element of chance *itself* (which at first glance seems to necessarily follow from the terminology of “probabilism”) appears to be a substantial problem for the idea that we *choose* an action as free agents. For would we not only be free if we could *determine* our actions? It does not seem that the concept of chance – which again, seems to follow from probabilistic conditions – could be a welcoming environment for free agents who would want to be *in charge of their actions*.

The question is of course whether there would be a fallacy in assuming, from V., that probabilism *necessitates chance* as an inherent component. Let us assume that either S is determined to “choose”¹⁷ an action, or he chooses it without being determined. If the latter is the case, then is it the fact of the influence of free will which negated the chance-component in the non-deterministic scenario? In other words, is free will such that it is non-deterministic in nature, without being subject to chance itself?¹⁸ If not, then what S decides to do is (at least partly) up to chance, and if so, S is not free – at least in any ultimate or complete sense. The free will optimist could here perhaps counter by saying that S is in control of the probabilistic process of choice, and that the chancy component adjusts to the tune of the agent’s will. But

¹⁶ What I mean to say here is that no matter how one views and defines free will, it would have to take place in the universe, which seems to be either deterministic, random, or probabilistic.

¹⁷ Of course, S would not in fact choose his action here, since he would only be determined to do X; and not be able to do any different than X. Hence, there really is no choice present, since S did not have a choice as to whether or not to perform X.

¹⁸ This is certainly the position most advocates of the possibility of free will would claim.

this kind of position seems farfetched, in that the mere presence of a chancy component makes it impossible for S to *ensure* that S will do, or want, X¹⁹.

If free will, on the other hand, is indeterministic *without* being subject to chance (that is, probabilism is not necessitated by chance), then there seems to be nothing left to account for what probabilism entails; which is that there is some undetermined factors, which are not within the ability to know, control or ensure. Probabilism seems to undermine the idea of free will, since it describes – at least on the face of it – a physical-temporal reality where events are not something to *be* determined (either with, or without, respect to an agent).²⁰

How could an agent have control in a universe which, seemingly by definition, operates based on uncertainty?²¹ In other words, an agent with free will seems to want to determine that which cannot be “determined”. If an agent could “will” his way in a probabilistic universe, it seems that it would not be probabilistic after all; it would simply be determined by the agent. Additionally, the agent could then will his actions in such a way that would be so obviously pattern-based, that there would seem to be something almost certain about his actions; as opposed to something merely probabilistic. All this could of course be purely semantical concerns spinning in its own tracks, but it would seem to me that probabilism appears to imply some component of chance; and necessarily so.

One intuitive way of resisting this line of thought²² is by insisting that while certain actions an agent does are outside of his control – such as being up to chance – *some* actions are within the agent’s control. For instance, one could argue that even though one did not choose to do X at t_1 , one chose to do X at t_2 . But the problem again is that it seems to be true to say that in any possible universe, every possible event must either be random, deterministic, or probabilistic.

¹⁹ I will come back to the issue of the connection between free will and *ensuring* an action or one’s will.

²⁰ Another way to phrase this would be to say that certain outcomes and events, on probabilism, must be unclear.

²¹ Here the free will optimist could of course claim that there is a fallacy in assuming that the only alternative to determinism is uncertainty.

²² Being that probabilism entails chance, and that this means that we cannot be free in the sense of deciding (in an ensuring and absolute way) our actions.

The agent could only *ensure* that he did X freely if he lived in a universe which was neither of the three described alternatives. For to ensure an action freely, one must be able to *guarantee* the outcome of that action; and to be able to do so would require that one lived in a universe where one's actions were neither up to chance, nor deterministic forces. To be free with respect to any action – unless one denies that the criteria of a free action is that it can be ensured to happen by the agent (at his will) – one must present a different alternative to probabilism.

If probabilism does necessitate some component of chance, it would appear that the aspect of “freedom” in the concept of freedom of the will, could just as coherently be that one was “freely determined to do A” (whatever that would mean), as opposed to just being free, in a probabilistic manner, to do A. In either case, whether supposed freedom of the will comes from probabilism, determinism, or randomness – it seems that we are in a fundamental way at a disadvantage to able to *decide* what to do without that decision being subject to external control and limitations by one of the fundamental ways in which the universe must function. The point is that we appear to either be prisoners of random influences, probabilistic influences, or determined influences - and there is no space between these three for free will to occupy²³.

In essence, my point is merely that it is difficult to imagine how free will could be explained, how it could be identified upon inspection, and in what way the laws of nature would have to be like, for there to be any room for the possibility of free will. This is the heart of the case the free will sceptic presents.

Prima facie then, I would argue that the free will optimist²⁴ needs to provide some arguments for the plausibility of *at least one of* the following two claims:

²³ At least given an understanding of free will as necessarily be characterized by an agent's ability to *ensure that his action takes place*. In other words, that it is not up to other (in)deterministic forces whether or not he performs the action.

²⁴ That is, one who argues that free will is metaphysically possible and/or likely (at least in this thesis).

A) chance is not, by necessity, innate²⁵ in probabilism

or, alternatively,

B) probabilism is not the only alternative to determinism and randomness

1.3 An objection from partial responsibility

It seems hard to defend the claim that we have *absolute* freedom of the will (or freedom of action). By this I mean to say that it is evident that there are some external, or internal, influences which do set boundaries to our freedom. We do not choose our parents, the environment in which we are infants, or the laws of physics. So there clearly are some set of restrictions within which the *framework* for freedom of the will must take place. However, it does of course not follow from this alone that one cannot, given a certain framework, have freedom of choice between a certain number of possible actions.

To argue against Galen Strawson's argument for the impossibility of free will, one could take issue with the way in which he claims that *all* possible actions an agent does, are completely – and necessarily so – incompatible with any claims of responsibility or freedom. We can say that arguing for this point of view, is arguing for the idea that partial responsibility is possible. This of course grants the free will sceptic the idea that we do not always have a complete, libertarian form of free will, and thus seems to agree with the idea that we cannot be *ultimately* responsible.

²⁵ The idea behind the use of the term “innate” in A) is that the free will optimist must demonstrate that a probabilistic universe can exist without any chancy components. By chance being innate in probabilism, I mean to say that probabilism is necessarily characterized by chancy components; and that this is a fundamental characteristic of a probabilistic universe. In other words, the question I pose is how could probabilism possibly function, if not by way of chancy components? How could not chancy components be a necessary part of the very fabric of a probabilistic universe – which, after all, seems to make the distinction between probabilism, determinism and randomness meaningful in the first place. The chancy components I have described (with the values 1 to 9) seem to be what separates probabilism from determinism (in that determinism excludes the possibility of chancy components – as evidenced by its values of 0 or 10), and probabilism from randomness (where the latter is characterized by *pure* chance – as evidenced by its value of 5).

The question then, on this line of argumentation, is whether responsibility is something binary – meaning, that it is something we either fully have, or do not have at all. One who claims that the idea of partial responsibility is coherent, is thus suggesting that the idea of responsibility is *not* simply binary; and that there are *degrees of* responsibility.

I contend that *ultimate* responsibility is what must be meant by responsibility; and thus, that ultimate responsibility must be *equated* with “mere” responsibility. I will attempt to show that the idea of partial responsibility seems less meaningful than what one might initially think.

Are there degrees of free will?

The initial question is whether there can be degrees of free will, by which we are responsible.²⁶ By free will, I am (in short) supposing that at *t*, S could have chosen both A and B; and that if we were to rewind the universe – so to speak – S would *without any element of luck or deterministic influence* be able to choose the opposite of what he in fact did choose at *t*.²⁷

An agent, at any given time of an action undertaken, seems to - in this sense - either be free or not. There seems to only be degrees in free will in so far as certain actions (in a set of actions across time) are free, and others not. But for any *specific* action, the agent seems to either be free or not; and clearly so in the sense described above. So, by singling out one time slice, or one moment of “agent-choosing”, we must either be dealing with an action that is free, or not free (in the sense described above).

We can attempt to demonstrate why being “fully free” is necessary for responsibility by the following argument(s):

P1: Either we are partially free, or we are fully free

P2: If we are partially free, then it is not solely up to us whether we do, or *want* to do, A or B at *t*

²⁶ Here, I am presupposing that the libertarian account is the relevant one in question.

²⁷ In an incompatibilist fashion, I am here presupposing the falsity of determinism.

P3: But if the choice at t is not solely up to us, then it is not in our power to *ensure* what happens at t (both with respect to what we do, and what we want to do)

P4: To be responsible for what we do, or want to do, at t , we must be able to *ensure* what our action will be, or what we want to do

C1: But since we cannot *ensure* what our action at t will be, or what we want to do at t , we are not responsible

And, alternatively:

P5: Either we are partly free, or we are fully free

P6: If we are fully free, it is solely up to us whether to do, or to want to do, A or B at t

P7: If the choice at t is solely up to us, then it is in our power to *ensure* what happens at t (both with respect to what we do, and what we want to do)

P8: To be responsible for what happens at t , we must be able to *ensure* what our action will be, or what we want to do

C2: Since we can *ensure* what our action at t will be, and what we want to do, we are responsible

I think this gives us a compelling reason to argue that responsibility can only exist if someone is in fact genuinely free (that is, fully free) with respect to an action, i.e. possesses what one would normally call libertarian free will - in its absolute sense.

Are there degrees of responsibility?

The question, then, is whether an account of “full” free will can result in partial responsibility. If we thus grant that an agent - at t_1 , say - is either fully free or not, the main question is whether there can be some circumstances in which A does X freely, *without* being ultimately morally responsible. I do not see how this could be the case. For having free will, in the ultimate sense, *presupposes* that we are responsible for any given action in an ultimate way.

This ultimate way describes the most complete kind of freedom imaginable, which is a genuine choice between A and B (or A, B, C etc) at *t*; where it is up to us (and not deterministic or chancy influence) what we will in fact do.

We could, for the sake of argument, grant the idea of partial responsibility for a moment (let us suppose the “partiality” is due to circumstantial elements of upbringing and genes). Being only partly responsible for anything seems to still imply being ultimately responsible for the choice *itself*, or *that action which one is being held accountable for* (on the libertarian free will account). So even partial responsibility must have something to do with the action itself being free (if not, I do not see how responsibility of any form is possible).

An argument against the basic argument’s claim of responsibility as absolute

One way of arguing against the basic argument’s idea of absolute responsibility is claiming that it is demanding to the point of being unintuitive with respect to our general ideas of what constitutes moral responsibility (and free will). Robert Hartman suggests that responsibility can be conceived of in “some weaker sense of self-creation”, and that this is all which “is required for true moral responsible agency”.²⁸ He thus argues that demanding an agent to be *causa sui* (in order to be responsible) is either wrong, or that the term needs to be understood differently than how Strawson views it.

Are there degrees of self-creation?

With respect to the possibility of there being some weaker senses of self-creation (or degrees of self-creation²⁹), a question is whether it is true to say that one is either fully self-caused, or not. On the one hand it seems intuitive to say that there can be degrees of self-creation, in that self-creation can be gradual if we possess freedom of the will. We seemingly can create ourselves as we develop through time, in that we can develop our personhood and character throughout our lives. This idea seems to make sense if we have (libertarian) free will. However, I think we must clarify, and clearly define, what is meant by self-creation in order to understand why it cannot be gradual.

²⁸ Hartman 2018, 14.

²⁹ I will in this thesis not differentiate between *self-creation* and *self-causation* (unless otherwise specified).

Concerning the question of *causa sui* I think it is important to distinguish between what one could call *absolute* self-creation and what can be called *gradual* self-creation. By absolute self-creation I mean creation which entails that S did not exist at t_1 , but did exist at t_2 . In other words, that S created himself from nothing, without degrees or graduality. This kind of self-creation is seemingly impossible since there is no time slice for which S could have created himself without already existing. S could not have created himself at t_1 (since he did not exist); and at t_2 S was already existent. Thus, he could not create himself at that time slice either. The only conclusion must be that S can not create himself (at least not unless there is an infinite number of time slices between t_1 and t_2).³⁰

Gradual, or partial, self-creation would mean that one could create oneself while one existed (in *some shape or form*). In other words, that S did exist (in some form) at t_1 , and created himself at t_2 . But the problem with this view is that self-creation *is* a process by which one becomes actualized; i.e. that one *ceases to be non-existent*. It is the “ceasing” of being non-existent which defines what it means to be actualized or existent. Self-creation thus seems to entail the transition from non-existence to existence, and it seems to be this very distinction – and the underlying transition - which defines self-creation as a concept.

Thus, we arrive at the seeming impossibility of self-creation. We cannot create ourselves if we do *not* exist, and we cannot create ourselves if we do *already* exist. One might object that this conclusion would be to beg the question, since it presupposes that we either exist or not. But even if existence is gradual, we must *at some point* either exist (partly or fully), or not exist. The objection could be granted, but one would then have to concede that the gradual, or partial, self-creation that is being argued for, is in reality more akin to something like “self-change” or “self-forming”. But at that point, one has been driven away from the relevant argument in question, which is that of *causa sui*. From these claims, I will attempt to show that absolute self-creation is the only conception of self-creation which would grant us responsibility, since mere partial, “sliver-responsibility” must be equated with ultimate responsibility. In other words, I will argue that partial responsibility *entails* ultimate responsibility, and that the latter follows from the former.

³⁰ If not, there is certainly no time slice in which S neither does *not* exist or *does* exist. To be clear, an infinite number of time slices does not, *prima facie*, suggest that absolute responsibility is more plausible on this conception.

We seemingly must work from the assumption that to be the cause of oneself one must already exist; and surely one cannot exist to various degrees. One can exist in various *ways*, of course, but a statement concerning X' mere existence is something which must either be true or not. Whatever a "weaker sense of self-creation" could mean, it would have to imply that one can also exist to certain degrees. How does this not seem inconceivable for any finite being, let alone a human being? Not only would one have to be able to "partially" exist; one would also have to create oneself in the first place. Saying that this is easier on a weaker conception of *causa sui* does not seem more probable than saying that one could be one's own cause in a complete sense.

At best a weak, or partial, conception of self-creation is just as dubious as being one's own cause in a complete sense. The bottom line seems to be that if one is to be one's own creator, one must already exist. And existence, can only be actual in a complete sense. Additionally, there seemingly cannot be limitations to one's property of existence (or alternatively, of *having* any properties at all). Existence appears to only come in one form, and one either "has" or "is" it, or not. One can thus claim that the idea of there being degrees to self-creation, as Robert Hartman claims, is unconvincing.

We can present an argument against partial responsibility (and thus for absolute responsibility) – by way of *causa sui* - in the following way:

P1: Responsibility (if possible or actual) is either absolute or not³¹

P2: If it is not absolute, then there are (at least two) degrees of responsibility

P3: Degrees of responsibility would mean that one were responsible for some actions, but not others – or, that one were partly responsible for any specific action

P4: In order to be responsible for *any* action, or its component(s), one would need to be *causa sui*

P5: But if one were *causa sui*, one would be responsible for *every* action, and all of its components

³¹ Here it is important to emphasize that all that P1 claims is that it is either true, *or not*, that responsibility is only possible in an absolute form. P1 does not claim that responsibility *is* absolute. It merely points out the logical truth of the fact that actual responsibility must either be absolute or not absolute.

P6: Being responsible in an absolute sense would mean to be responsible for every action, and all of its components

C: Therefore, one would need to be responsible in an absolute sense, and for every action and its component(s) - if one is to be responsible at all

Premises one to three seem hard to deny. Premise four seems true due to the infinite regress its negation would imply. Premise five must be true in that if one were one's own self-cause then there is no reason why one would arbitrarily be able to choose some preferences, but not others. After all, one would be in a position to create oneself as one pleased. Premise six simply states what a natural interpretation of the terminology of "absolute responsibility" would be; and therefore strikes me as hard to question to any considerable degree. And lastly, the conclusion just follows from the premisses. In other words – if the argument is valid and sound - being responsible for a "sliver-mental constitution", or a part of a mental constitution, also implies being ultimately responsible for the totality of that mental constitution.

1.4 An objection anticipated – the claim of existence being gradual

One can of course argue against the notion of the binary claim of something either existing, or not existing; and thus simultaneously argue that there can be degrees of self-creation in this way. How, then, can one argue that existence comes in various forms? One might, for example, say that a human fetus can be "in a realm of existence which lies between something human, and something non-human". The problem with such a statement, I think, is that there is a confusion present between the property, or fact³², of X' existence; and the property of being human as something *additional* to the mere fact of existence (we can call the question of existence itself for existence*). One must note the difference between saying that:

I - something can exist to various degrees *as* a human or non-human, and that

II - something can *simply exist* (as anything) to various degrees.

³² One could of course argue that existence is not a "property".

The former is a statement relating to what it means to exist as one thing or the negation of that thing, *presupposing existence* (i.e., existing as *something* or *anything*). The mere fact of something existing is here taken as obviously true. The latter statement, on the other hand, is related to *existence itself*. In other words, one could grant that there is an arbitrary line between being human and non-human (as in the case of a fetus); without that position revealing anything about what one might think of existence* as absolute.

I would argue that existence* in itself is fundamentally different from existence (as in the example of something being human or non-human) in that the former cannot *not* be absolute. What I mean by that is that while one could argue that X neither exists as something fully human or fully non-human, it does not make sense to say that X does not fully exist *in any shape or form*³³. The only possible negation to X' complete existence, is X' complete *non-existence*. And thus we can identify a difference between existence (as something specific³⁴) and mere existence*. What this difference entails is that X' existence can be on any level. The important factor is to simply acknowledge that the property or fact of existence itself is one that either is true of X, or not. In other words, X either exists or not; and X' existence (if actual) is thus absolute – and “comes” in but one form.

Can responsibility or free will emerge gradually?

Another way Hartman argues that there can be degrees of responsibility, is by the idea that responsibility can *emerge* gradually over time, by way of an agent repeatedly engaging in actions which themselves emerge from yet other set of actions which the agent, to some degree, is responsible.³⁵ But this, again, fails to understand the necessity of *causa sui* being absolute – in the fundamental sense of only being conceivable in one way.³⁶ One cannot become morally responsible over time, when at t_1 , one was *not* responsible for one's nature. For this fails to account for the infinite regress, which is one of the main problems to begin with. The personal growth over time which Hartman describes, cannot be ultimately traced

³³ That is, exist as something – or indeed *anything*.

³⁴ As in belonging to a certain category of something, or someone, who is existent.

³⁵ Hartman 2018, 16-17.

³⁶ What is more, this is of course only the case if the free will sceptic grants the idea of *causa sui* to begin with.

back to the agent's control of his initial conditions of being; and is therefore fundamentally not within the agent's control.

One might object that it is obvious that an agent cannot have control of his initial state of being, since there are external factors present, which clearly limits and shapes the agent's repertoire for free action. A feasible position would be, one could argue, that what could still be true is that the agent can be in control of an action at t_{10} , for instance, even if events that took place between t_1 to t_9 were outside of his control. But while this might explain how someone could be free given the threat of external factors, and their limiting effect on free agency, it does not present a solution to the problem of the infinite regress.

For again, the challenge of the regress still holds in that the mental constitution of S at t_3 , say, is the result of S' mental constitution at t_2 , and S' mental constitution at t_2 is the result of S' mental constitution at t_1 . This is, of course, especially true with finite beings, who do not have an infinite number of time slices in which to become ultimately responsible.³⁷

In short, every action requires an explanation, and that explanation must fundamentally lie outside of the control of the agent, *irrespective* of whether that explanation could be found in external factors, or internal factors (such as mental constitutions³⁸).

Attempting to escape the infinite regress, it will in either case be true that the initial mental constitution (of any possible set of mental constitutions), or the initial external factor, originated in such a way as to be outside the agent's control. For the agent must in such a scenario have existed with a mental constitution with which he did not bring about, or "will" into existence. Hence, if an infinite regress is not required for ultimate freedom and responsibility, there must have been an initial explanation for the agent's free action. And an initial mental constitution, without any prior mental constitution connected to it (as via a set of mental constitutions), would seem to require an explanation which would be in such a way that no agent could have brought it about; since the explanation and the initial mental constitution would be explained only by itself in an absolute or non-reducible sense (if not by something external).

³⁷ It would also seem like an infinite regress would do nothing to prove the possibility of responsibility emerging over time, since there is hardly anything "free" with not having caused oneself to come into existence (which is what would be true of a being who has existed infinitely).

³⁸ I say that mental constitutions are internal in that they relate to an individual's mind, where it can be supposed that free will originates.

The agent could not possibly be the author of his initial mental constitution, because – by rejecting the argument from the infinite regress - there can be no prior mental constitution for which the agent could be responsible for. The idea of there being a mental constitution which was (freely) brought about by an agent – *without* that agent having a prior mental constitution with which to determine the later mental constitution, must surely be incoherent. An agent cannot be said to freely choose a mental constitution without that decision itself being based on a prior mental constitution, and by way of this notion we are yet again faced with the problem of an infinite regress. For any mental constitution to be within the agent's control, it seems to follow that it must have been brought about by the agent³⁹; and the only way that could happen was if the agent *prior to that acted on a different* mental constitution. Thus, this attempt to escape the objection from the infinite regress itself faces the very same problem. Simply stated, a mental state is required to do an action (freely or otherwise), and there is no easy solution to how an agent could possibly have control over what inevitably would have to be an *initial* mental constitution.

1.5 Identifying general misconceptions about free will scepticism

The fallacy of presupposing indeterminism *itself* to be a basis for free will

A common argument against free will scepticism generally, as well as Galen Strawson's account more specifically, can be labelled under the umbrella of "arguments from indeterminism". These arguments claim that from the (possible) fact that the universe is indeterministic, one can assume that free will is possible; if not probable or even actual. Randolph Clarke presents an example which illustrates this view.

Clarke initially asks us to imagine that S' being – or that which encompasses S - at *t*, can bring about *different* actions performed by the agent. In the words of Clarke, "... it was nomologically possible that all these events (desires and beliefs - and mental constitutions⁴⁰) occur and the thief not decide to refrain ..." from committing a certain action.⁴¹ Clarke's point

³⁹ Both willingly, and with respect to the notion of it being possible that S could have done otherwise.

⁴⁰ For clarification Clarke does not himself specifically use the terminology of mental constitution, but I think it can be equated or added with "desires and beliefs" – without altering Clarke's own example.

⁴¹ Clarke 2005, 16.

here is that until S decides to either do action X, or not do action X, there is something indeterministic (and unknowable) about the result of the process from which known mental constitutions existed prior to an action being performed by the agent.

In other words, given perfect information about S' mental constitution at t_1 , it was *indetermined* what the agent would do at t_2 . As a result of this indeterministic factor, it would appear (or so the free will optimist would argue) that the fact that an agent could both do X and not-X at t_2 , from an identical mental constitution or desires and beliefs at t_1 , gives some *prima facie* reason to think that the agent would, or at least could, possess free will.

What is being said here is that it was possible that from the contributing factors at t_1 , S did not perform a given action at t_2 . But the question is this: how does the indeterminism underlying this process point to the agent in question possessing free will? Clarke's example only states that there is something indetermined involved in an agent having a certain mental constitution at t_1 , and it being possible that this "origin", or initial point of factors, resulted in different outcomes. For the agent to have free will, he surely would have to "be in control" of the indeterministic factors himself? My point here is that we ought to ask why it would follow from the fact that event B is not determined by event A, that the possibility of event A *not* leading to event B was within the agent's control. Mere indeterminism cannot, at least in itself, be a convincing explanation for free will.

Clarke counters (perhaps more in the spirit of the free will optimist than his own) by saying that until S decides to do X, it remained open to him *not* to do it.⁴² This would seemingly suggest that the agent is in fact free, since it is not until S' choice is actualized – by way of indeterministic processes – that one could know what S would do (or did in fact do). But again, this fails to account for the fundamental problem that one cannot claim that indeterminism, *by itself or its very existence*, is equated to – or points towards - free will. Just because event A does not necessarily lead to event B, we have no basis for assuming that the agent is in control. And similarly, just because of the fact that the agent's choice could have been different - had we run the universe back - it does not follow that the indeterministic factors themselves (by way of necessity) shows that the agent acted freely.

This only proves that the agent could have *acted* differently, given the same previous, causal events that led up to the event of the supposed choice. The essential fallacy here lies in

⁴² Clarke 2005, 16.

supposing that simply because an agent could have acted *differently*, at a previous time slice, he was thereby acting *freely* (at the previous time slice). This is an overlooked and crucial difference. In fact, if there is one thing which I would want to press with respect to Clarke's example, it is the falsity of equating S having different possible actions, with S being free with respect to these actions; and the indeterministic factors which govern the choices that are to become actualized. It simply does not follow that just because numerous possible actions were open to the agent at t_1 , given the same mental constitutions, he is therefore *in control* of the indeterministic processes which could bring about the different actualized actions. In short, having the ability *act* differently, does not mean that one has the ability to *choose* differently.

The fact that things could have been different; and the fact that the agent, under identical circumstances in universe₁ and universe₂, say, - could have acted differently does not necessarily imply free will. Granted, there is something intuitive about supposing that indeterminism *could* be a basis for free will – and perhaps more so that determinism.⁴³ But upon inspection one realizes that indeterminism – on the face of it – is no more necessarily allied, or non-hostile, to free will than determinism.

In essence, this means that from the fact of the agent's mental constitution (and desires and beliefs) at t_1 being the same in universe₁ and universe₂ - and t_2 (at the time slice of the agent's action) being different with respect to the two universes - it does not follow that this difference *itself* is necessarily the basis of a free choice. All Clarke's example could possibly accomplish is proving that determinism would be false. And in this way, it would be a black-and-white-fallacy to suppose that the falsity of determinism, itself, presupposes free will.⁴⁴

A challenge for the free will optimist, concerning indeterminism in this respect, lies in coming up with a solution to the problem of distinguishing "chancy" indeterminism from indeterminism which is not hostile or incompatible with free agency. To explain and present an account of free will, one would initially have to acknowledge that for S, at t , to have had free will, one cannot exclusively rely on the truth of S *acting in a way that was not*

⁴³ Of course, the compatibilist would disagree with this notion.

⁴⁴ The perhaps most obvious example of this would be quantum mechanics; where no one would argue that the observed randomness *itself* would constitute a basis for a conception of freedom of the will.

bound to necessarily happen. To “could have acted or done otherwise” might be a necessary condition for free will, but it certainly is not a sufficient one.⁴⁵

For it is the agent’s relationship to bringing about the action that is (at least equally so) the crucial part; not the mere fact that the agent could have done X, instead of the negation of X (or vice versa). To could have done otherwise might mean, for instance, that S could have been *forced* to do the negation of X instead of X, say. And therefore, we can see why the ability to could have acted otherwise in certain cases is not a sufficient condition for freedom of the will. One must clearly have had the ability to act *freely* – in addition to having had the ability to could have acted *differently* – to have had free will.

The basis for any claim of free will must thus lie within what led up to the action being undertaken; and not merely by way of the indeterministic event that laid the foundation for the agent’s ability to could have done differently. What a free agent in this sense would have to be able to do (in order to be free), is to not only “could have *acted or done* otherwise”, but also to could have genuinely *chosen* to have acted otherwise (in contrast to the case of S being forced to do X). For again, surely the alternative, counterfactual action must have been within the control of the agent for the agent to be free. In order for the notion of “to could have done otherwise” to be a sufficient way of explaining free will, the alternative action must thus not only have been a possible, contingent turn of events that could have happened. The explanation or reason for the action (as well as the decision) must have also been within the agent’s *control*.

Arguments from external factors – why the sceptic does not presuppose that S needs to control more than internal factors in order to have free will⁴⁶

One argument against the free will sceptic, as well as the basic argument, is that it is too demanding. For John Martin Fischer, the argument takes the form of claiming that the basic argument is asking too much with respect to its supposed claim of an agent only being

⁴⁵ That is, S’ action could for instance have been non-free in that it was purely random.

⁴⁶ The sceptic does not argue that we need “total control”, instead he holds that the only control we should be responsible for is what is internal to us (that is, our mental constitution(s)).

responsible if his control of an action is exclusively “internal”.⁴⁷ What Fischer is saying is that the basic argument claims we must be in control in an ultimate sense, in order to be truly free and responsible agents. And in a way this is of course true.

Strawson, for instance, argues that we must be *causa sui*. But the problem of Fischer’s critique of the free will sceptic is that he supposes that *every possible* factor must be within the agent’s control in order to be free. Fischer says, for instance, that “The sun’s continuing to shine is a necessary condition of the existence of human beings. The sun’s continuing to shine is a causal “background condition” that enables me to be the way I am and to act as I do”.⁴⁸ Obviously this is true; for without the sun, earth, the presence of an adequate amount of oxygen and so on, we would not be able to be free agents (or human agents at all). The crucial point, however, is that the free will sceptic – or at least a Strawsonian - does *not* claim that this would have to be false in order for free agency to be conceivable.

All the free will sceptic says is that one must be ultimately responsible for *some factor* regarding one’s mental constitution. The sceptic does not demand that a free agent must have complete control over every causal “upholding effect”, which negation would eliminate our possibility of being free agents (that is, if its counterfactual(s) were true). The point is not, as Fischer seems to think, that we cannot be ultimately responsible because of external factors which are outside of our control. It is S’ *internal* mental constitutions which is what is under dispute, *not* the external factors which causally upholds the possibility of free agency.

In other words, the sceptic’s argument is not that we need to be in control of every underlying cause which lays the foundation for our *possibility* to be responsible and free agents. Rather, the point of the free will sceptic is that we need to be in control of the *internal* factors which contribute to our actions, desires, beliefs and so on. On Strawson’s basic argument, for example, Kristin Mickelson states that it “draws our attention to antecedent factors beyond our control which are internal to the agent, factors which— despite being inside and, indeed, partly constitutive of the agent—still seem to be beyond the agent’s control in troubling ways”.⁴⁹ One must not confuse an argument about the *internal* being outside of the agent’s control, with the *external* being so. The latter clearly is outside of S’ control in certain key

⁴⁷ Fischer 2006, 108.

⁴⁸ Fischer 2006, 115.

⁴⁹ Mickelson 2019, 19.

aspects, and so the more pressing question is in what way the former could be within control of an agent.

This is the mistake that Fischer commits when he states that we cannot be ultimately responsible, because of "... the air's continuing to be (barely!) breathable, and so forth are causally sustaining conditions for my agency, over which I have absolutely no control."⁵⁰ Further, Fischer adds that "It is as if Strawson thinks of free and morally responsible agents as having "total control". Again, this exemplifies the point a free will sceptic like Strawson would want to make. Fischer has simply not been able to identify the argument which is actually being made, since he confuses "total control" (with respect to both internal and external factors) with "ultimate control over a sliver-mental constitution"; or something like "total control over *any* mental constitution".

As Michael Istvan notes, Fischer's fallacy is in failing to realize that in order to disprove Strawson's argument, "The agent need only be ultimately responsible for the merest sliver of whatever contributes to A."⁵¹ The argument stating Strawson is demanding too much, as in the agent needing to have "total control" in order to be free and (ultimately) responsible, thus appears to be a strawman. Strawson *could* grant that S is ultimately responsible for X – with respect to some aspects - if one could show that it is possible that S could be ultimately responsible for *any* component of his mental constitution(s). However, Strawson questions the metaphysical possibility of being ultimately responsible for *anything*. His argument is in this way not really that concerned with the binary distinction between external and internal factors; just as his argument is less interested – or at least not dependent upon – the falsity or truth of (in)determinism. It is the metaphysical puzzle of ultimate sourcehood for one's mental constitution, and thus one's actions, that is the underlying foundation of Strawson's case against the concept of freedom of the will being possible.

⁵⁰ Fischer 2006, 116.

⁵¹ Istvan 2011, 406.

- **Part 2:**

Why a probabilistic universe does not point to the existence of free will

2.1 A theory of probabilism

Having argued that free will appears unlikely to exist metaphysically (and in any possible world), I will now turn towards discussing probabilism, and prospects of the existence of free will in probabilistic universes.

If my arguments in favour of the universe being deterministic, random or probabilistic so far are valid and sound - and my relevant assumptions so far are true - it appears that the free will optimist must account for his theory of free will within a probabilistic universe. Thus, I will in this section of my thesis present an account of a theory of probabilism, which appears to exclude – or at the very least make improbable – that such a universe would be a place where free agents could exist (let alone human beings).

Below I will explain my reasoning as to why I think that the existence or reality of a concept of free will is implausible in a probabilistic universe (and especially so the universe which we in fact inhabit).

Having control over one's will - why probabilism⁵² is hostile⁵³ to free will

Presupposing that we live in a probabilistic universe – which seems reasonable⁵⁴ – we can present the following argument for why a sensible theory of probabilism seems to favour the viewpoint of the free will sceptic.

⁵² For clarification, probabilism is equated with, or refers to the same, as “a probabilistic universe”.

⁵³ I deliberately choose the word “hostile”, as opposed to “incompatible”, in order to not appear too bold or assertive.

⁵⁴ Initially, I would say that it is reasonable to believe that we live in a probabilistic universe for two reasons. Firstly, the universe does not appear to be fully deterministic, nor fully random. Determinism seems to be unlikely with respect to our current understanding of quantum mechanics. A random, “raw chance” universe appears unlikely due to the fact of how our universe, in many obvious ways, appears to be pattern-based in a way which would suggest that there are indeed some things which are more likely than others. Secondly, we have *prima facie* reasons to believe that the third reasonable alternative for describing the universe is that the actualization of possible truths operates by way of probabilistic means. This appears to be the most straightforward way of describing how things

P1: S cannot *have control over* S' will through free action in a probabilistic universe, if that universe contains chancy components involved in S' decision-making.

P2: A probabilistic universe contains chancy components, and these are universally and necessarily imbedded in the very fabric of such a universe – from which it follows that there are also chancy components involved in S' decision-making.

C: Thus, S cannot *have control over* S' will in a probabilistic universe

Clarifying the premisses

P1.

What comes to immediate attention would be the specific wording used in the first premise, where the words “have control over” are highlighted in italic. It seems reasonable for any incompatibilist - and thus a libertarian - to state that free will, at the very least, implies being able to control what one *wants*, if not what one *does*.

It seems clear, or so I argue, that for S to possess free will - which would be present at any time slice in which S existed (that is, total free will in an absolute sense) - S would have to be able to *have control over* S' will in some sense, and thus not find that S' will was outside of S' ability to make sure that X happened. By all reasonable standards, when the libertarian states that S has free will at *t*, he intends to say that S is free in the ultimate sense described.

However, both being able to have control over what one *wants*, and being able to have control over what one *does*, are impossible in a probabilistic universe, *if* (and perhaps *iff*) that universe contains chancy components⁵⁵. The existence of chancy components means that it is

happen if they are neither determined nor random. Events having a value, or chance, of happening between 1 and 9 (excluding the value of 5), is a logical and coherent alternative to something either having the value 0, 10 or 5 – that is, something which is either determined to happen or not to happen, or something which is perfectly and truly random.

Further, I am at the time of writing not aware of any promising theory which presents itself as a forth alternative to determinism, randomness, and probabilism as a coherent model for describing the universe.

⁵⁵ There are two ways to understand the implications of this claim. The first is that for any relevant factor involved in S' decision-making at *t*, there are probabilistic influences present

impossible for it not to be true that it *could be the case that S, at t, is not free (in the sense of having control over S' will) to choose X*. In the simplest sense, this is clearly true in that if S cannot determine, with absolute certainty, that he wills⁵⁶ X; then it could be possible that S cannot decide – in a way which *is solely up to S* – whether S wills X. For S to be free in the absolute sense, S would have to “determine that which cannot be determined”⁵⁷. Or said in another way, his freedom of will would have to coexist with, or inhabit, say, a universe where the act of *deciding that a certain future truth-value would be certain in its “yet to come” outcome, is fundamentally impossible* (in a probabilistic universe). For one cannot determine or decide, with absolute certainty, that which – by its very definition – is not itself determined (or cannot be determined by anything else, including actions by agents).

P2.

How can one argue for the claim that a probabilistic universe contains chancy components? Firstly, one must realize that the expected truth-value⁵⁸ of a probabilistic component, or event, is neither fully *certain* nor fully *uncertain* (that is, not deterministic or random).

Secondly, this (partial⁵⁹) uncertainty could be accounted for, as I see it, by two types of fundamental components. The first one can be called *chancy components*, and the second alternative to this is some other kind of fundamental component. One such second, fundamental component of a probabilistic universe, would be what the free will optimist could label a *free will component*.

I will argue that chancy components are the simpler, more “Occamian” answer to what would constitute a fundamental component in a probabilistic universe. Mere indeterminacy, in the form of chance, is fairly straightforward metaphysically, and it is also understandable or explainable to a certain extent (such as in quantum mechanics). And if not fully

which negates the possibility of S acting free in the absolute sense. The second is that the fact of *any component* being probabilistic in a universe, makes it impossible – theoretically speaking – for S to positively *control* that S can “will” (or at least do) X.

⁵⁶ Here understood as equivalent to “wants X”.

⁵⁷ I will shortly expand upon the reasoning behind the use of this specific expression.

⁵⁸ Meaning the probability of a certain *p* becoming true, or of a certain event becoming actualized (or coming into existence).

⁵⁹ That is, not *complete uncertainty*.

understandable or explainable, it is certainly *conceivable* to imagine the metaphysical possibility of a chancy event taking place spatiotemporally somewhere in the universe (or a counterfactual, possible universe).

One can initially make a *prima facie* case for chancy components being more suited to account for the indeterminacy in a probabilistic universe in the following way. The first assumption is that it is not difficult to imagine a metaphysically coherent possible world where a probabilistic universe is characterized by chancy components, and chancy components *alone*. The second assumption is that it *is* (at least far more) difficult to imagine a metaphysically coherent possible world where a probabilistic universe is characterized by free will components⁶⁰. In the same way, more generally, I would add that it is fairly straightforward to conceive of a possible, metaphysically coherent world (whether probabilistic, deterministic or random) where there did *not* exist agents with free will. On the other hand, it would be more difficult to conceive of a possible, metaphysically coherent world where there *were* agents who possessed free will.

Although this does not constitute any concrete argument against the idea of free will components, I think it does say something about what we *prima facie* ought to favour as an explanation for the indeterminacy of a probabilistic universe. Adding to this initial case for chancy components, one can with respect to ordinary language and context realize that something being *up to chance* is what we mean by claiming that an event has a given probability of occurring. And this makes sense in that if an event were not up to chance by way of probability, it would either seem to be determined, or seem to be subject to raw chance (that is, every single event at any possible *t* having a truth value of 5).

What, then, of the claim that chancy components are necessarily imbedded in the very fabric of a probabilistic universe? If we assume that a probabilistic universe contains chancy components, then it would seem more likely that not that any area of space, or any time slice (and thus, the totality of the spatiotemporal and physical universe as a whole – at least as we know it), would also necessarily be characterized by chancy components. For this to be false, one would have to argue for one, or both, of the following claims of chancy components:

⁶⁰ If not difficult, then it is certainly not obvious how free agents could exist in the same way as it is obvious how free agents could *not* exist (in the actual, or in a possible, world).

- a) they are not universally spread, and/or
- b) they are not imbedded in the universe (as in the sense of being part of the universe's constituting fabric or its fundamental properties).

There seems to be no compelling reason to favour the theory that the universe is only *partly* composed of chancy components, since if some spatiotemporal slices⁶¹ are characterized by chancy components; *all* would seem to be so. This seems likely, I argue, given that it is hard to imagine what sort of metaphysical explanation there can be for the sort of constant *arbitrariness* that would have to take place for just a select number of regions of spacetime to be characterized by chancy components. Again, the simpler explanation seems to be that the universe is uniform with respect to chancy components, in the same way that it is uniform concerning some of its fundamental forces - such as that of gravity. And if this is true, then it follows that an agent, who occupies any possible region in spacetime, is also bound by the implications of the probabilistic rules that govern the universe.

Is having control over one's actions and desires necessary for the existence of free will?

Perhaps the crucial question, with respect to free will in a probabilistic universe (and with respect to the probabilistic argument I present against the free will optimist) is whether S *having control over what S does or wants* is necessary for S to have free will⁶². Is this the case, then? Or need S, for example, only have a high probability of getting S' will realized through S' actions or desires? I think that *not* being able to have control over one's will means that one cannot reasonably be called a free agent (that is, an agent who possesses free will). In what follows, I will, among other things, try to explain why I think this is the case.

⁶¹ As in a time slice, with the added dimension of a region of space within which the time slice in question occupies or inhabits.

⁶² Indeed, the first premise of my initial argument from probabilism relies on this assumption.

How the free will optimists⁶³ seem to agree with the importance of having control over one's actions or will

The vast majority of compatibilists would presumably agree that chancy components, or randomness, are not compatible with free will. And the libertarian, as well, would certainly agree with this (at least to a significant degree). The compatibilist would seem to disagree with the claim that free will presupposes being able to control *what one wants*, but agree with the claim that free will presupposes being able to control *how one acts on account of what one wants*.

For the compatibilist does not claim that choosing one's desires and volitions is necessary for free will, but he does claim that one needs to be able to act on those desires and wants – without external constraints - in order to be free. The libertarian would seem to hold that both having control over what one *wants*, as well as having control over how one *acts on account of what one wants*, is necessary for an agent to possess free will. Due to the nature of this thesis⁶⁴, I will primarily focus on the libertarian account of freedom of the will.

Why the challenge from probabilism remains

The problem, though, is that both having control over one's desires, as well as having control over one's actions (based on one's desires), is necessarily impossible in a probabilistic universe, *granted* the assumption that such a universe contains chancy components (which again, I have argued it is reasonable to assume). As the only alternatives to a probabilistic universe seems to be either a deterministic one, or a random one, it appears that chancy components are a *necessary negation* to deterministic components⁶⁵.

By this I mean that either something is deterministic, or the indeterministic components is accounted for by chance⁶⁶. For instance, this would mean that when we state that an event in the universe was non-determined, we make a twofold claim. Firstly, we make a claim about the possibility that that event, counterfactually speaking, could have *not* happened in that

⁶³ I here refer to both libertarians and compatibilists.

⁶⁴ This will be noted again in this thesis' concluding remarks.

⁶⁵ One must notice, though, that chancy components are *not* a negation of indeterministic components, since chancy components - by definition - is themselves indeterministic.

⁶⁶ Note that this does not necessarily mean "raw chance".

world. Secondly, we state by the first claim that the counterfactual possibility in question could only have been actualized by way of chance.

The libertarian would obviously disagree with the second claim. He would here either counter by saying that a probabilistic universe is not only accounted for by chancy components, since it is possible for S to control S' actions. Or, alternatively, the libertarian would state that the universe is *not* probabilistic. The latter seems rather far-fetched, since there clearly seems to be at least *one single event, or one single component*, which is non-determined by way of chance in the entire universe⁶⁷. Thus, the most reasonable position the libertarian could take, would be to claim that:

P3: S' choice can be free in a probabilistic universe if there are no chancy components involved in S' decision-making

P4: It is possible for a localized spatiotemporal event in a probabilistic universe (i.e. an agent undertaking a choice), to exist, or happen, in which the value of that said event happening (1 to 9) is accounted for by S' free action – and *not* by chancy components

C2: Thus, S can be free in a probabilistic universe

The conclusion would here imply that S is able to act freely in a probabilistic universe without chancy components interfering in S' free choice, since S' free choice is what *in itself* accounts for the probabilistic components of an event's value being between 1 to 9. Thus, S is free in the sense that S' action being undertaken – freely by the agent - is what accounts for the explanation of why the value of an event happening is not 0 or 10 (in other words, between 1 and 9).

What the free will optimist here claims⁶⁸, is that a *free will component* is what accounts for the indeterminacy in a probabilistic universe, as opposed to a *chancy component*.

⁶⁷ One should here note how a universe has been defined as probabilistic if only *one single* component or event has a value of between 1 and 9 of happening, as discussed earlier in the thesis. This definition itself is of course subject for debate as well.

⁶⁸ Again, I am here speaking with respect to the libertarian.

2.2 A theory of probabilism continued

Free will components and chancy components

The *similarity* between the two component's indeterministic explanation⁶⁹ is of course that both describe that an action or event is not determined, which means that it is not certainly going to happen, nor certainly going to not happen (as described by its value *not* being 0 or 10).

The *difference* is that the free will component is accounted for by S' *determining or deciding* a specific event happening (the event of an action freely being undertaken, and its consequences). The chancy component describes a similar indeterministic event. But here there is no factor which *determines or decides* the indeterministic event, since it is, of course, up to chance.

The challenge, then, is that the free will component states that something which is not determined, by its very definition, can nonetheless be *determined*, in that an agent singlehandedly decides that an event will take place⁷⁰. And hence we arrive at the reasoning behind my claim that the free will optimist must state that *free will require us to determine that which cannot be determined*. And this is why chancy components more coherently, and certainly less confusingly, seem to better account for the indeterminacy which a probabilistic universe is characterized by. Accordingly, we can argue for P2, and hence C⁷¹. We can also, in addition, more boldly argue for P2* - where the only difference with respect to P2 is that the word "exclusively" has been added. Thus, we can argue that:

⁶⁹ By indeterministic explanation, I am here referring to how the two components account for a given universe not being determined. By this I mean that both chancy components and free will components are possible explanations as to why a given universe is not deterministic. In other words, *something* has to account for the indeterminacy, and that something could for instance be chancy components or free will components, respectively.

⁷⁰ Such an event can either be an action undertaken, or it can describe the event of an agent making a conscious decision, or an agent "making up his will".

⁷¹ Again, P2 states that "a probabilistic universe contains chancy components, and this is imbedded in the very fabric of such a universe – from which it follows that there are also chancy components involved in S' decision-making".

P2: A probabilistic universe contains **exclusively** chancy components, and these are universally and necessarily imbedded in the very fabric of such a universe – from which it follows that there are also chancy components involved in S' decision-making*

Elaborating on why chancy components alone seem to best account for the indeterminacy of a probabilistic universe

I further claim that chancy components are the more promising explanation of the indeterminacy of a probabilistic universe, since free will components appear not to be subject to chance, and hence are not possible in a probabilistic universe⁷². The free will optimist may say (and indeed, must so, if he grants that free will necessitates having control over one's will) that free will components *are not determined, but not subject to chance either*. But again, the problem is that it seems to follow from indeterminacy that there is *necessarily uncertainty* involved with respect to an indeterministic event.

We can illustrate this with respect to what appears to be a temporal truth about the relationship between time slices in indeterministic universes. If an indeterministic event e_1 at time t_1 were known, and event e_2 at time t_2 were not necessarily *unknowable* at t_1 , then it would have to follow that it was in fact *possibly* knowable at t_1 . And hence, the universe would be determined (granting that this same ability to know what would happen at a future time slice – from the perspective of a previous time slice - was spatiotemporally universal⁷³). For if the events at t_2 were *possibly* knowable (that is, in an epistemological sense) at t_1 , then it follows that the event at t_2 was in fact determined⁷⁴.

For if there is something that can be *known for certain* about t_2 at t_1 , then that truth value at t_2 must be what one can call *set*, or universal across time slices, as in that the truth value was certain *before* the given event actually happened. In other words, it was possible to know that

⁷² If one were to disagree with the claim that the free will component is not up to chance, then one would grant that there is no freedom to control one's will after all.

⁷³ That is, valid for all time slices whose relation, or outcome, was knowable from a previous point in time.

⁷⁴ I am of course aware that one could here object to my use of knowledge as something atomic and non-reducible (in that knowledge in the sense described here is understood as S being certain of p if S is to know that p . And further, that it is impossible for S to be wrong about p if S is to know that p . I thus equate absolute knowledge with what one might call "partial" knowledge, or rather, suggest that the latter is an impossibility on any reasonable understanding of knowledge as a term.

p would be true at t_2 , *prior to* the time slice of t_2 . And this, again, gives us a deterministic relationship with respect to the specific events of these time slices. Additionally, since the negation of indeterminacy, by definition, is determinacy - the free will optimist thus needs to account for how it is possible that an indeterministic event can be under control of an agent who inevitably attempts to freely operate and navigate a universe which is necessarily characterized by uncertainty with respect to what happens at any given moment.

The free will optimist may object that the indeterministic uncertainty of free will components are not knowable, and thus uncertain, *precisely because S does not know what S will do at a future time slice*. Thus, he may claim that even if there is uncertainty involved in S' decisions, this uncertainty does not necessarily imply chance.

I do, however, think there is a problem with this line of thought. I think the problem can be stated as follows. If S decides at t_1 that S will do X at t_2 , can S then be free to decide what S will do at t_2 , if there is *uncertainty* involved in whether S will do X at t_2 ? The crucial point here lies within the relationship between the two time slices in question, and S' relationship to them. If the counterfactual of S doing X at t_2 is possible at t_1 , then it would appear that S cannot choose at t_1 what S will do at t_2 .

If it is true at t_1 that the counterfactual of S doing X at t_2 is possible, then it follows that S cannot freely decide at t_1 to necessarily⁷⁵ do X at t_2 – if we grant that S' action at t_2 probabilistically is not *necessarily* going to happen. This holds even if S at t_1 , in all other relevant aspects, *decides* to do X at t_2 . But S is not, at t_1 , free to decide what S will do at t_2 , *if* S wants to control what S will necessarily do. Thus, in any sense that matters with respect to P1 (temporarily across time slices) – and S having control over S' action, specifically – S cannot be free in the absolute sense described.

The explanatory power of the two components in the present

However, even if S cannot be free in a probabilistic universe with respect to *future* time slices (for the reasons stated above), it does not necessarily follow that S cannot be free in the *present*, or in the “now”, as one might say. And this would also be a natural way for the free

⁷⁵ In the way described as «having control over X”.

will optimist to view the concept of free will⁷⁶. The free will optimist might argue that even if *S* is not free at t_1 with respect to *S*' action at t_2 , *S* is however free at t_2 with respect to the action at that time slice. What this position would amount to, would be an idea that every possible free will component (and thus every free agent) exclusively accounts for the uncertainty of a probabilistic event at a specific time slice at a time – and necessarily in the present. And this would appear to be the most natural, and certainly the most conventional position the free will optimist could take. All this position would add up to, is simply that *an action is free at the point that the action is being undertaken*.

How, then, do we evaluate chancy components and free will components with respect to their explanatory power of probabilistic indeterminism?⁷⁷ Even though chancy components are also of course bound by only directly being able to influence the present (as with anything else with respect to time⁷⁸), an advocate for chancy components constituting a probabilistic universe would not claim that they rely on an agent's *control* to make sense – or form a coherent theory of indeterminism. This is one reason why chancy components are the more “Occamian” solution to solving the question of what sort of metaphysical components most likely characterize a probabilistic universe.

I claim, then, that even if we only account for the possibility of *S* being free with respect to the present in a probabilistic universe (which again, to be clear, by all reasonable standards is what most free will optimists would argue), there is no convincing reason to favour the explanatory power of free will components over chancy components. By this I do *not* mean to say that free will components are impossible, since it is not unthinkable that there could be such an explanation for why the universe is non-deterministic.

I do think, however, that it would take a very convincing case for any true free will agnostic⁷⁹ to favour free will components over chancy components as a fundamental explanation for probabilistic indeterminacy. The existence of both chancy components and free will components would also seem quite arbitrary, since the alternation of chancy indeterminism and free will indeterminism would in a very profound way adjust to the tune of how a free

⁷⁶ After all, how many people mean to say that they are free with respect to t_2 at t_1 ? This does indeed, to me at least, sound quite unintuitive.

⁷⁷ Said differently, how does the two components compare with respect to how well they account for, or explain, the nature of non-determined events in a probabilistic universe?

⁷⁸ After all, every event is “trapped” in the present, and necessarily so.

⁷⁹ By free will agnostic I simply mean someone who has not made up his mind as to whether free will optimism or free will scepticism is the most reasonable view in the free will debate.

will optimist would *want* the universe to function. In this way, a probabilistic universe having these two types of indeterministic components, really does sound like a theory that would be popular with human beings who would want there to be such a thing as freedom of the will. It would certainly be very convenient for someone who would want it to be the case that free will components existed in a non-determined universe.

Considerations of the animate and the inanimate⁸⁰

In addition to these considerations, I think there is another factor that needs mentioning with respect to the evaluation of chancy components and free will components. The free will sceptic can argue that if chancy components already exist with respect to the *inanimate* in the universe, then we ought to favour this explanation with respect to the *animate* as well.

Initially this might sound like a strange claim, as the free will optimist certainly would state himself, since surely there are reasons to believe that the animate is indeed very distinct from the inanimate – and that this itself be a reason to believe that the presence of chancy components in only the inanimate could be what explains the very difference between that which is animate, and that which is inanimate. But I think a reasonable free will agnostic must ask himself which of the following claims should be deemed most convincing in evaluating free will components and chancy components, respectively.

- a) We should favour free will components, since chancy components cannot, or does not most convincingly, account for the animate performing actions
- b) We already observe *unfree*, inanimate probabilistic processes in the universe, and should therefore – until further arguments to the contrary – regard both the animate and inanimate as governed by probabilistic chancy components

⁸⁰ I more generally choose to use the term “animate”, as opposed to (human) “agent”, since the former is more broadly applicable – and thus allows for the possibility of a less restrictive understanding and applicability of what - or who - can possess free will. If anything, this is done to make the case of the free will optimist more intuitively compelling.

The relevant dispute between the free will optimist and the free will sceptic can in many ways be described by the contention between a) and b).

From b) we already have something concrete we can show to. We have every reason to believe that chancy components *do in fact exist*. And for this reason alone, it does not seem implausible that they should also be universal, and constituting one of the fundamental properties of the universe. And if they are universal, and an imbedded property of the fabric of the universe, then they would also be present with respect to an agent acting (and thus, an agent could not ensure his will or actions). It would not matter, on this view, whether something was animate or inanimate (or an agent) – since chancy components would account for the very underlying fabric of all that happened at every time slice of the spatiotemporal universe. And hence, acting and choosing freely would be impossible.

From a) I think it is in fact not very plausible that there is such a distinct, vastly different, component to the already established chancy component. Even if there was a reason to think that the claim that the animate is so different from the inanimate, had to do with the fact that the former was indetermined and free, would this reason really be sufficiently convincing to abandon the reasoning behind b)? I think b) is the least speculative, as well as being most grounded in what we have reasons to believe is true about the way in which the universe functions.

Having made all these claims, though, I want to stress an epistemological, human predicament, so to speak. We really are not capable of making robust, let alone bulletproof, predictions about the reality or illusion of free will. But the best we can do, I think, is to evaluate what would seem like the theory which is dependent upon the *least amount of conditions or assumptions*. This is especially true of those conditions which are not based on assumptions where there is considerable agreement present (granted, such conditions are perhaps not very easy to identify in philosophy at all). And at least in this way, the free will sceptic presents a more convincing case, with respect to the considerations that is here mentioned.

For the theory stating that chancy components constitutes the totality of the probabilistic indeterminism present in our universe, presupposes only the following: *chancy components are universally spread*. The theory of free will components, on the other hand, makes bolder claims. Firstly, this theory claims to have identified a distinct component to an already

established one, without it in any obvious way being clear how this component functions or operates in relation to chancy components (nor is it clear how it can be explained in a metaphysically coherent way).

Secondly, it further challenges, as a result of this, the notion that probabilistic indeterminacy is not sufficiently explained by chance alone. I object to these ideas partly by claiming that a probabilistic universe can be more easily understood through chancy components.

Additionally, it seems clear that a theory attempting to describe how a probabilistic universe functions, would seem to be more easily explainable if it were described by chancy components alone. Thus, I think it is largely fair to put the burden of proof upon the free will optimist in this aspect, and challenge him to give a satisfactory explanation to questions like how free will components function, and why, or by what means, they would seem to be existent in a probabilistic universe on par with the already established existence of chancy components.

- **Part 3:**

How could the free will optimist resurrect his case for the existence of probabilistic free will?

3.1 The “could have done otherwise”-condition and transworldly committedness

Having discussed free will both in terms of its metaphysical possibility, as well as its likelihood to exist in probabilistic universes, I will in this section now turn to some arguments which may help the free will optimist in his search for a convincing theory of free will.

Supposing that the ability to could have done otherwise is *not* necessary for free will, I claim that this does not help the free will optimist, since he is still left with the question of accounting for how any choice – with the ability to do otherwise, or not – is free. As I have explained, the ability to could have done otherwise is not sufficient for free will, but it may not even be necessary. The question is what we can make of the concept of free will, granted that a principle of alternative possibilities is not required for free will in the first place⁸¹. In

⁸¹ Kane 2009, 17.

other words, how can the free will optimist coherently present an account of free will if we grant him the claim that alternative possibilities are not necessary for free will?

The point is that even if the free will sceptic grants the free will optimist the claim that the ability to could have done otherwise is not necessary for free will, the optimist still is no closer to establishing an account of free will – by the simple fact that free will could obviously both be true and untrue if the ability to could have done otherwise is unnecessary for establishing the possibility of freedom of the will. By no closer to establishing an account of free will, I mean simply to say that the optimist still has all his work ahead of him to arrive at a coherent and pervasive theory of free will *even if* we grant him the assumption that the ability to could have done otherwise is not needed to establish the possibility or actuality of free will in our universe.

This is fairly straightforward, in the sense that whether or not someone believes that the “could have done otherwise”-condition is necessary for free will tells us nothing about whether free will actually exists (or is possible). Thus, free will existing and free will not existing are what constitutes the binary possibility involved if we grant the free will optimist the claim that the “could have done otherwise”-condition is not a necessary constituting part of a theory of free will.

Free will could both be true and untrue, no matter the truth or falsity of the “could have done otherwise”-condition. In other words, it does not follow from either the truth or falsity of the “could have done otherwise”-condition, that free will is either true or untrue. This point may appear to be something of a banality, but it is worth mentioning to clear up that this point of contention (the condition, that is) does not *singlehandedly* prove the case for either the sceptic or the optimist.

Introducing transworldly committedness as a defence against the sceptic's "could have done otherwise"-condition

As mentioned, I do not think the "could have done otherwise"-condition is necessary for free will. The reason for this is that I think it is possible to speak of free agents existing within a universe which can be described by what I call *transworldly committedness*⁸².

Transworldly committedness simply means that for any specific set of factors which is true of a time slice, an agent inhabiting that time slice will only act freely in one way; and only act in this one way because he will *always want* to act in this way (given the specific circumstances at *t*). Initially this notion might seem odd, given that it (on first glance at least) appears to be a contradiction between saying that S is free, on one hand, and that S will only act in one way, on the other.

All transworldly committedness claims is that the two following claims are not metaphysically impossible.

1. *For any, and all, alternative or parallel universes described by a specific set of characteristics (a given scenario where S will make a choice between X and not-X), S can always freely do X iff S always wants to do X, and*
2. *S can always want to do X iff⁸³ S is – as a matter of fact - free to choose between wanting X and wanting not-X.*

What, then, is implied by 1. and 2.? If we take for granted that S is – as a matter of fact – free, then this would have to mean that S, at every possible time slice, can make a free choice between X and not-X. And since S can always make a free choice, it would mean that S at all possible time slices can decide to make *either* choice. Thus, within the same scenario across time slices or parallel universes, S *must necessarily* (due to S' free will) be in a position to always *have the ability* to freely choose to always want X (or not-X). The freedom of S must

⁸² I purposely use this term because of its similarity with Alvin Plantinga's term of "transworld depravity", in the hope that it may appear less confusing; and perhaps more familiar.

⁸³ Notice that «if» is here substituted for "iff", since it is possible for S to always want X *without being free*.

thus necessarily include *the ability* to always choose X over not-X (and vice versa), if S so wishes.

In this way, we can see why S being a free agent necessitates the ability to choose to be transworldly committed. For if S could not always choose X (that is, have the ability to do so) – and thus not choose to be transworldly committed to X - then S would not be free with respect to all of S' choices across the relevant time slices.

From this it follows that it is possible for a free agent to always want X⁸⁴. Thus, we can further logically conclude from this that it is possible for S to never freely want not-X⁸⁵. Hence, transworldly committedness shows us how the “could have done otherwise”-condition could be flawed. On this account, it simply fails to account for the necessary fact that it is in the power of a free agent *to always choose one thing over another*. Importantly, this holds true in any scenario or time slice which S, as a free agent, inhabits.

Said in another way, the “could have done otherwise”-condition does not account for the fact that a free agent can, given the premiss that he is indeed free, *choose to never change his mind*. For a free agent it is always possible to not do X, but actually choosing not-X can only be just that – *possible*, but not *necessary*. The freedom of S, on the other hand, entails - by necessity - that S can *choose* to be transworldly committed to X across all possible time slices in which S is in fact free.

Harry Frankfurt has described a view similar to transworldly committedness in the following way: “Suppose that a person has done what he wanted to do, that he did it because he wanted to do it, and that the will by which he was moved when he did it was his will because it was the will he wanted. Then he did it freely and of his own free will. Even supposing that he could have done otherwise, he would not have done otherwise ...”.⁸⁶ The ability to do otherwise is here understood as possible in the sense that it is possible for a free agent to change his will, would he want to.

Of course, this brings up the question of what “to could have done otherwise” means. Does it entail that S, in some parallel universe or possible world, *would* have done differently? This does not seem to be the case, since there is clearly an obvious difference between having the

⁸⁴ It is similarly, in order for S to be free, also necessary that S has the *ability* to always choose S – as demonstrated.

⁸⁵ This is logical because X is a contradiction to not-X.

⁸⁶ Frankfurt 1971, 19.

ability to do something, and *actually* doing something. And if the “could have done otherwise”-condition does not entail that S would have done differently, then there is nothing contradictory with saying that S *could* have done differently to what he in fact did in universe₁, universe₂ and so on, *ad infinitum* – even if S in all of these parallel or possible universes did the exact same thing under identical circumstances.

Crucially, “could have done otherwise” does *not* imply “would have done otherwise”, or even “will do otherwise”. It does not even seem to imply “can do otherwise”, for if S cannot change S’ mind so as to want what S does not want – it could still be true that S *could* have done otherwise, in the sense that S in a possible, metaphysical world *could* have done otherwise. Even if it is impossible for S to not want what S does in fact want (or vice versa), it is not thereby false to say that S could have done otherwise, if we understand “could” as merely saying that there could exist a metaphysically coherent world where S *wanted* not-X instead of X, say.

Further, assuming something happens the exact same way an infinite amount of times, it does not mean that it must be so. A probabilistic singular event happening does obviously not entail, by necessity, that it had to happen in the specific way that it did. So why would anyone be inclined to think that a series of identical, probabilistic events would be any different with respect to the way they come about?

“Could have done differently” would on such an account mean nothing more, or less, than that “*had* S wanted to act in a different manner, then S would have acted differently.”

Assuming S is free, it is true that S could have done differently (since this follows from the fact that S is free). But it could simultaneously be true that S *in every possible scenario with specified circumstances*, would have done X, and never not-X.

Expanding on these assumptions, it strikes me as necessarily true that S under the exact same conditions (or even merely similar ones) across possible worlds or alternative universes *could freely* have done the same as S did in the actual world – or our world, say. Or, stated more simply, that the action of S at *t* in universe₁ could be the same as S’ action at *t* in universe₂, universe₃ and so on. In other words, it does not seem impossible that S would only want to do X, and *never* not-X; no matter the number of times world₁, or our universe, was restarted or “rerun” from its initial time slice.

If these assumptions are correct, it follows that S need not have the ability to *do* otherwise, because S does not *want* to do otherwise (in conjunction with the assumption that S is free). In a strange sense, we would perhaps look at S as *unfree* were S incapable of sticking to S' desires and beliefs across alternative universes resembling a specific scenario in universe₁. How can anyone really say that a free agent must be in a position to occasionally (let alone always) never have stable and consistent "wants"? Would it be surprising if an agent, supposing he indeed had free will, had at least one desire and belief which was constant across all time slices or all alternative universes? On the contrary, I think it would be odd if every single free agent *at no point whatsoever* had rock-solid wants, desires, or beliefs – which were universally spread across all possible time slices and universes those agents could inhabit.

So even if S could have done otherwise *if* S wanted to do otherwise, it would still be true to say that S "could not have done otherwise" – merely in the sense that this sentence conveys that it is nonsensical to say that *S is not free if S cannot do what S in no possible world (or no alternative universe) wants to do*.

Thus, I do not think that the "could have done otherwise"-condition is a *necessity* for free will, even if it may - in most cases - be an obvious constituting part of agenthood, and the undertaking of free actions. Again, this shows how free agents may display the ability to could have done otherwise, but that this is only a contingency, and not a necessity, of what constitutes a free agent. The ability to do otherwise might be no more than an "accidental attribute" of a free agent. If the notion entailed in transworldly committedness is a reasonable one, then S can indeed be free with respect to an action S would always want to do, in any specified set of circumstances.

To sum up, transworldly committedness states that S *would not have acted (freely) differently* in any parallel universe (that is, in any one scenario with specific circumstances). Thus, S could not have acted differently *only in the sense that S cannot do what S does not want to do*. Ironically, perhaps this is exactly what some people would claim that free will is. But at least to my understanding, for the reasons outlined above, this appears to be a somewhat counterintuitive notion.

An argument against the idea of transworldly committedness

There may be convincing counterarguments to the notion of transworldly committedness. For instance, one can say that “could have done otherwise” must mean that it is possible for S to change S’ mind about *every choice, at any possible time*. And perhaps this strict and absolute understanding is what a meaningful account of free will must entail. Or, alternatively, that for any singular choice, “could have done otherwise” necessarily means that S can override S’ wants in a fundamental and complete sense; in a way in which transworldly committedness would be false. Specifically, it could on this notion be true that the “could have done otherwise”-condition must specify that the agent, if free, is *not* transworldly committed, *precisely because he is free*.

The semantical debate would here be over what “committed” actually means. Does it mean that S makes up his mind freely at t_1 so as to not be able to choose against this “freely willed will” at t_2 ? Or does it mean that transworldly committedness entails the assumption that S has never committed himself, because S can change his mind? The ultimate question is *whether S is free to choose what S wants*. In the third section in this part, (“How does the free will optimist give a coherent account of internal control of the will?”), I will discuss the notion of internal control and S’ ability to choose S’ wants and motivations.

At this point, one might ask the following. What is the distinction - if any - between “S could do X”, and “there exists a possible world where S does X”? I do not think that “S could do X” means that “there exists a possible world where S *does* X”. S having the ability to do X obviously means that there exists a possible world where S *can* do X, but as I have implied (with stating that “could have done otherwise” does not imply either “would have done otherwise”, or “will do otherwise”), saying that S can do X is not untrue by virtue of S not doing whatever action is in S’ power to perform.

For example, in order for it to be true that I can choose to leave my house at t , does that mean that some twin of mine from a parallel universe or possible world *does* leave his house at t ? I do not see why it should. A possible course of action at t_1 is not made *impossible* by the fact that it does not come about at t_2 . We must here notice an underlying problem.

Stating that it is possible, at t_1 , for X to happen at t_2 , in conjunction with the statement that “X did not happen at t_2 – therefore X was not possible at t_1 ” implies a contradiction. For either it was possible for S at t_1 to choose X at t_2 , or it was not. Looking at the events after the fact, at

t_2 (and not t_1), cannot tell us anything about what was possible at t_1 . In other words, the future of what S actually ends up doing does not dictate S' past possibilities of choosing an action.

If S has free will, S can at any moment choose to leave S' house – and this possibility of action is not dependent on whether S actually does end up leaving his house. Simply because something is never chosen, it does not mean that that choice was not a *possibility*. And there is no clear reason why this point would not hold both in respect to a single world, as well as across possible worlds.

3.2 Predictability and patterns

How the free will optimist can argue that the free will sceptic begs the question of an action being unfree because it is either predictable or non-predictable

S being S, and S being free, seems to be related to the fact that S' character remains somewhat true to itself across time slices. S acting seemingly randomly would give the appearance that S was not free, in that S' actions would seem to be so unpredictable that they were not in control of S; whereby S is understood as possessing agenthood (a personality, or a coherent and predictable set of beliefs and values, say).

However, if S in future time-slices – or yet-to-come time-slices - is consistent (to the point of being obvious) with S' actions in past time-slices, it also seems that S would appear unfree. Thus, the free will sceptic argues that no matter the nature of an agent's actions, they seem to point to something which does not seem to originate from a mind *which is both free and which represents personhood* (as in a person's desires, personality traits, and points of view).

The free will optimist might counter by arguing that this line of reasoning is begging the question. But in which way? Well, the free will sceptic is assuming that both S acting seemingly randomly, as well as S acting in such a way that S' actions appear determined, both point to S' lack of free will. But where exactly is this line of argumentation going, the free will optimist might ask. For surely there is no other way that the observance of an agent's action could be characterized. Either S' actions appear more or less random, because they are diverse and change sufficiently over time. Or, alternatively, they appear more or less determined – or rule-bound – since they follow patterns which are predictable to a certain extent (though not fully predictable of course, as on determinism).

To counter this, the free will sceptic might say that this is indeed the very point he wishes to make. For he says that it does not matter if S' actions are predictable or not, for he equates non-predictability with chancy components, and predictability as something which by its very nature makes it impossible for S to be free (in the same way that an incompatibilist states that S is not free in a deterministic universe). This is a possible fallacy the free will sceptic must be aware of.

Were the free will sceptic to admit to this process of equation, the free will optimist ought to argue, I think, that the former has proved nothing by begging the question of non-predictability and predictability only representing chancy randomness. I think this sort of reasoning might be compelling from the point of the free will optimist. Despite my argumentation for the likelihood that chancy components account for the indeterminacy of probabilistic universes, one can of course argue that this is unconvincing. But by doing this, the free will optimist needs to identify that the sceptic is in fact making the claim that both predictability and non-predictability in the actions of agents are simply the results of processes of probabilistic, chancy components.⁸⁷

How the free will optimist can claim that patterns are not opposed to the idea of free will components

A relevant question A.J. Ayer raises is whether free will is impossible or unlikely given the assumption that we can predict the actions of an agent with considerable accuracy. As Ayer states, "... to say that my actions proceed from my character or, more colloquially, that I act in character, is to say that my behaviour is consistent and to that extent predictable ...".⁸⁸

What Ayer here is implicitly asking is whether it is peculiar to say of S that S is free at t_{10} , say, with respect to an action which he has willingly done in all prior time slices (t_1 to t_9). If S drinks, and enjoys, coffee every morning for 100 days straight, would we say that S appears to be free with respect to, yet again, choosing to drink coffee on day 101?

The claim one can make here is simply that patterns, by virtue of describing what is to be expected to happen – or what can accurately be predicted to happen - is threatening to the notion of free agency. Again, the free will sceptic can here argue that we either have no

⁸⁷ As specified previously, chancy components are in the context of probabilism *not* randomness.

⁸⁸ Ayer 1954, 376.

character, and are thus chancy creatures; or, that we do have character, but that we appear to act probabilistically simply due to the fact that there is some pattern with respect to which we can be expected to act. In either case, the sceptic can view both of these possibilities as incompatible with the existence of agents possessing free will.

I do not, however, think this line of reasoning is convincing for the free will sceptic. S repeating S' action or behaviour across time slices does not mean that S did not do every single action freely. This would only be true if the universe was determined, and S was unfree in the incompatibilistic fashion stating that S *could not* help but do X. Or, alternatively, that the universe was random, but that S still in a random fashion continued to do the same action.⁸⁹

As argued, it also seems possible for S to *only freely want to do X*, and never not-X. With respect to transworldly committedness, one can expect to see a free agent having a preference for a given thing across multiple time slices. And so, on the face of it, there is nothing which appears unfree about predictability *in and of itself*. Surely it is not impossible to do something freely simply because someone makes an educated guess about what you are going to do, even if that person is right in his or her guess.

The question is rather *who or what* it is that accounts for the predictability of events and actions. Is it the fact that someone freely *chooses* to act in a predictable manner, or is it some probabilistic, chancy component which accounts for the non-deterministic pattern? Since this is an open question, the possibility of accurately predicting behaviour does not automatically mean that S, who is accurately predicted in his actions, is thereby unfree by default.

Adding further problems to the free will sceptics case, S repeating S' actions across past time slices does not mean that S will necessarily do the same action in future time slices. For again, if this were not true it would follow that the universe was determined. So, the predictability we have in mind here is not such that it somehow proves that S will do the same action at both a present time slice, and a future time slice. One of the main problems of the free will debate remains, of course. What, precisely, account for this indeterminacy? This is the very question by which, if we had an answer, we could establish the existence (or lack thereof) of incompatibilistic free will in a probabilistic universe.

⁸⁹ As mentioned previously, pure randomness can theoretically have the appearance of predictability, or even determinism.

3.3 How does the free will optimist give a coherent account of internal control of the will?

A problem remaining for the free will optimist

As I do, with respect to the (proposed likelihood of) free action being impossible, Roderick Chisholm equates internal reasons or causes with external ones in their explanatory power of free will, *if internal reasons or causes originate in such a way that the agent is not able to control his desires and beliefs*.⁹⁰ If the agent cannot control what he wills, or “will what he wills”, he cannot be said to be a free agent. For he would then be a prisoner of his own inability to decide freely in a way which is just as hostile to the conception of free will as anything else one can think of.

What noteworthy difference does it make whether the explanation for S’ inability to choose S’ beliefs and desires are internal or external? If S is not free to choose, it does not matter if the obstacle for S achieving free will is S’ own inability to will what S does not presently will; or if it is some external limitation to the possibility of free action. There appears to be no difference between external shackles and internal shackles, if these shackles obstructs the way of the free will.⁹¹ The only difference – and a considerable one, that is - between external and internal grounds for free will, is that it is only the latter which *could possibly* account for there being such a thing as freedom of the will⁹².

In other words, the lack of physical shackles does not necessitate free will, but the lack of mental shackles would seemingly imply the existence of free will (were it to be metaphysically possible for agents to possess freedom of the will). External causes can be a necessity – and a necessary condition - for free will (in the sense that the universe must contain energy, atoms and the possibility of change through the passage of time, for instance). But apart from these foundational conditions, free will must obviously reside in the internal mind of agents. In this way, the right external conditions are necessary, but not sufficient, for there actually being agents with the capacity of experiencing free will. Their freedom clearly

⁹⁰ Chisholm 1964, 391-392.

⁹¹ By “shackles” I simply here mean something which hinders the way of the free will.

⁹² We can reasonably assume that if S is indeed free, then S is free because of what S *is*, or because S has certain mental faculties or processes of consciousness which produce free will. Clearly S - existing as an agent making his own choices - is what makes S free.

does not come, in any way but the most foundational and facilitating way described, from an external source allowing for the possibility of free agenthood.

Can one have control over one's will?

However, in order to make way for an internal account of free will, one must first explain how it is possible for S to be in control of S' desires and beliefs.

First, one must posit whether Q1 is true, and consequently, one must answer - if Q1 is answered in the affirmative - Q2.

Q1: S only does what S wants⁹³ to do

Q2: S can control what S wants to do

Q1 is crucial, since if one were to accept that this statement is true, then the next question is of course (as laid out above) whether S can actually control what S wants. And if one answers in the affirmative regarding Q1, then one must also answer accordingly with respect to Q2, in order to claim that free will is possible (since I regard it as obvious that only internal factors can be a source of free will, and that controlling what one *wants* is what is meant by having free will in the first place).⁹⁴

The Hobbist⁹⁵ approach and the Kantian approach

In order to answer Q1 we must determine whether the Hobbist approach or the Kantian approach to the (possibly) logical connection between wanting and doing, is correct. The Hobbist approach states roughly that S only does, or can do, what S wants. The Kantian approach argues that there is no necessary connection between S wanting something, and S

⁹³ By «wants» I here simply refer to something equivalent to “desires and beliefs”.

⁹⁴ After all, what we mean by an agent having free will is that that agent has control over what he does and what he wants.

⁹⁵ I am aware that “Hobbesian” is a more used term than “Hobbist”, but due to the source material used in this thesis (Chisholm), I will in a similar fashion as in these sources – to avoid unnecessary confusion – simply adopt the latter term when referring to any philosophical view originating from the ideas of Thomas Hobbes.

undertaking an action. The tension between the two approaches is thus whether a free action must follow logically from motivations which can only produce the given action.⁹⁶

Roderick Chisholm appears to favour the Kantian approach to Q1. He states that “No set of statements about a man’s desires, beliefs, and stimulus situation at any time implies any statement telling us what the man will try, set out, or undertake to do at that time”.⁹⁷ In addition, he quotes Reid, saying that “though we may “reason from men’s motives to their actions and, in many cases, with great probability,” we can never do so “with absolute certainty””.⁹⁸ In this way, Chisholm defends the Kantian approach for explaining the relationship between motivations and actions.

What Chisholm’s Kantian position would imply, is that S can act *without having control over S’ action*⁹⁹. Or, perhaps more precisely, this would imply that S cannot have ultimate control in the sense of deciding X without it being possible that S does not want to X. For if S cannot “with absolute certainty” *choose* X on account of what S actually *wants*, then how is S free? S would not appear to be free if it was not obvious that S would do X if S wanted to do X, in conjunction with S’ ability to do X. A lack of certainty with respect to S acting on S’ beliefs and desires seems to me to be the complete opposite of what S having free will would have to entail.

If S does not *exclusively act freely* on S’ beliefs and desires, by what other means can S meaningfully be said to act freely at all, in the sense that it is the agenthood and choice-making of S that makes a conscious decision? I think the Kantian approach would best describe what could be characterized as an *unfree* choice, since this approach entails that S could want one thing, and yet do something else which in no way originated from a previous desire directly connected to the action in question. Here, S acts not on S’ beliefs and desires, but on some other means – whereby S’ actions would thus, by necessity, be *unrelated* to what S actually wants to do. Exactly where in this action-making process is the supposed free will found? Where is the freedom in this failure to trace an action back to a conscious desire which would explain the rationale behind S’ said action(s)?

⁹⁶ Chisholm 1964, 396.

⁹⁷ Chisholm 1964, 396.

⁹⁸ Chisholm 1964, 396.

⁹⁹ What I mean by this is the apparent absurdity of saying that S is in control of X which does not come about as a result of S’ free will to *want to do X*.

In short, we would expect S, if S were to be free, to do as S wants; and in such a way that it makes sense with respect to S' motivations, personality traits and prior actions. Further, I think we would expect S, were S to be *unfree*, to behave in the way in which Chisholm presents; where S seemingly (and paradoxically in my mind) can act freely without the action being explained so that it logically follows from S' motivation(s). I think a sort of randomness, which by its very definition is incompatible with free will, is present if one cannot *in any way* (even theoretically speaking) deduce that such and such exact, specified motivations would result in such and such specific actions¹⁰⁰.

The bottom line is that there is nothing unfree about the logical connection between actions and motivations, *unless* S is not free to control what S wants. It is as if S can only be free with respect to the first step in action-taking, which is choosing what S wants to do. But at the second step, S will by necessity do what S *wants* to do. So S can be free, but S *cannot be free to act in contradiction to what S wants at the moment of action*.

Analogously, a bowling ball (supposing for a moment that it has agenthood) could be free to choose which pin it would make impact with. But after choosing its direction towards one pin, it cannot choose to want to hit another pin, and *still* end up hitting the initial pin. At a certain point, the bowling ball would have to make up its mind, and choose which pin to hit. But once it chooses a direction and a pin, it cannot hit that pin without it being true that *this is what the ball wants* (if the ball is free).

In short, if the motivation, or the choice, to do X is free, then there is nothing unfree about the fact that that specific motivation would result in a specific action. In other words, if S is free to will what S wants – and if S is thus free with respect to S' motivations – S is not made unfree by the Hobbist approach being true, per se. I think the Hobbist approach is promising in the sense that no matter the truth or falsity of free will, any agent would nonetheless seem to be bound by a law stating that action B must follow from motivation A, if B only can make sense if A is fulfilled at a prior time slice.

It would make little sense, I think, to say that a motivation to *not* do X is followed by S doing X, and especially so if S is supposedly free – because then S would appear to act freely on what S does not want.

¹⁰⁰ Notice that this does not imply determinism – since just because 2 must follow from 1 (in an indeterministic universe), it does not mean that 1 must exist to begin with. Determinism would imply that both 1 *and* 2 must exist, not only that 2 must follow from 1.

Similarly, I argue that for S to have motivation Y (or beliefs and desires¹⁰¹), it would be odd to state that S did Y*, where Y* is represented by an action slightly different from the motivation, Y. This is not immediately apparent, however. For instance, the Kantian might object that S could be motivated to read a given newspaper, paper₁ say – but only come across a slightly different paper (paper₂)¹⁰². By reading paper₂, S has acted on a motivation which is similar to the action S did – and causally connected to it – without that action being a direct result of the specific motivation.

I would argue that the initial motivation entails something like the following: “I want to read paper₁, but it would also be acceptable to read paper₂, if it would prove to be impossible to read paper₁”. Thus, S’ motivation would in reality not be an “all or nothing” motivation regarding reading paper₁, but it would be – if the motivation itself were understood, analysed and examined deeply enough – a motivation which was so specific so as to indeed *actually account for the way in which S acted by reading paper₂ when paper₁ was unavailable*.

In short, S’ motivation to read paper₁ entails a subtle “want” to read something similar, (in this context, paper₂) *were S not in a position to read paper₁*. If, however, S’ motivation truly was to read paper₁, and paper₁ *alone* (if paper₁ had some sort of sentimental value for S), then S would not read paper₂ if paper₁ was unavailable. Thus, S’ motivations would be true to the action S actually performed. Hence, either way, if S’ motivation was understood in a complete sense, we could see how S would act as a result of S’ motivations. And thus, one can at the very least make a coherent account of what a Hobbist approach may look like.

What if the Kantian approach is flawed?

For the Kantian approach to be correct, it would have to be true to say that it is impossible to formulate a motivation so precise (X) so that S “could not help” but act on action Y. That is, in such a way that Y necessarily followed from X - were S to be free in the sense of not having external constraints upon S’ will to act in a given way. Though not impossible, this notion of impossibility relating to S’ motivation(s) strikes me as improbable.

¹⁰¹ Unless otherwise specified, I equate “beliefs and desires” with “motivations”.

¹⁰² For simplicity, let us imagine that the two papers are the very same edition of the local paper of the day, just in two different physical copies.

First, let us state something uncontroversial (or so I hope). Either S' motivations are finite in depth, intricacy, and layers, or they are not. If it is infinitely deep and complex, I do not see how any agent could possibly be in control of such an infinity of motivations and motivational reasons, which would by its very definition be bottomless and impossible to grasp for finite minds (let alone human minds). Thus, I regard the position of S' motivations being infinite in this way as somewhat absurd, especially so if one is talking about any possibly free agent.

But if this is correct, it follows that S' motivations are *finite* in their describable nature. This means that there is, at the very least, theoretically a finite number of operations, say, which can be done in order to fully understand, or - if not understand - at least *formulate* a description of what S' motivations are. For instance, it is either true or not that the betting limit for S at the poker table one afternoon is 5000 kroner¹⁰³. Either S is willing to go up to that number, but no further; or S is willing to bet something above that amount (or, S could of course not want to bet anything more than 4999 kroner – or something much less than 5000 kroner).

Whatever else is true about S as an agent, it is either true or not - or *will* be true or not – that S either has the motivation to bet X, or not. And if his action is not the result of what S actually specifically and precisely believes, then how can S be free with respect to S' action? And how is S not then simply a slave to randomness or his own lacking knowledge of what he wants? *An agent acting, but not knowing what he wants, is surely not a free agent.* S would appear to be acting more on reflex, such as in a scenario where S were to automatically and unwillingly blink if someone clapped their hands right in front of his face, than anything else.

In line with what I have argued, I think that a convincing case can be made for the fact that motivations are ultimately reducible to a specific variable or truth, meaning that a precisely enough - and sufficiently specified - motivation would necessarily result in a given, specific action (or event). Crucially, this is not to say that free will, if this assessment is indeed correct, would be proven impossible. Because it does not follow, of course, from the fact that S has only one specific motivation by which to achieve one specific action, that S thereby is unfree with respect to *choosing S' motivations*.

The Hobbist approach, however, does narrow down the possibility for the free will optimist with which he can account for the possibility of free will in a probabilistic universe. For if the

¹⁰³ Notice that this would in fact be true irrespective of whether S was free.

Hobbist approach rings true, a total emphasis must be made on how a free agent can be in control of his motivations, desires and beliefs – since these, on the Hobbist approach, necessarily and single-handedly are connected to an agent's actions.

The limits at hand

A Kantian, such as Chisholm, might argue that the Hobbist approach does not adequately account for the complexity of the motivations behind an action. And that if examined carefully enough, we would see that there is no logical or strict relationship between an action and a motivation (such in the case of motivation X *both* having the possibility of resulting in action Y, as well as action Z, for instance). I am willing to grant that this could, of course, be the more accurate way of describing the issue.

But no matter the complexity of the issue at hand, I think we can establish that when S acts, S *does so because of a reason*. What I mean by that is that this reason (or reasons), *is the very explanation itself* for why, and how, S acts. *All things considered*, this must be the case. On the Kantian approach, or any coherent approach for that matter, one would presumably agree that all else being put aside, when S acts, S does so – in the specific way in which S actually does act – because of *some kind of explanation*. Surely it would be absurd to deny this, since even stating that “S did X because of S' free will”, is itself some kind of explanation for S' action.

It does not follow from the fact that there must be some explanation, that there must be a *single* explanation for an action. For even though S' action at *t* must be explained by something, it would not decisively prove the Hobbist approach even to identify what the explanation for S' action was at *t*. That would only prove that *in that one instance*, S acted because of motivation X. It would not imply that S necessarily did action Y because of X. Just like one can show that an individual went to bed because he was tired, say, it does not mean that he can *only* go to bed because he is tired (perhaps he sometimes goes to bed because he has stayed up late, and wants to get his eight hours of sleep before next day's work – even if he is not feeling sleepy at all).

Why the Hobbist approach is not a theory of determinism

The Kantian might argue that the Hobbist, virtually speaking, is a determinist; or that he is indistinguishable or akin to one¹⁰⁴. Or, perhaps, that the Hobbist at least defends something resembling a deterministic approach to the relationship between sets of events, because he claims that a specific action necessarily follows from a specific motivation. The Hobbist approach does not, however, imply determinism. For it does not state which motivations or actions that will come about or exist, or even which agents that will exist. Nor does it state anything about at which specific time slices both of these considerations will be true. And it certainly does not say that it all will come about in but *one set* of events, where every event is necessarily preceded and followed by a different, specific event.

The Hobbist approach only states that *given* motivation X, action Y will occur *if it is all up to the agent*. It does not even say that action Y will *necessarily* occur, it simply says that granted no external interference, S will do Y. Or, more precisely, that S will *attempt* to do Y if S has motivation X. Thus, the Hobbist approach does not imply determinism, even if it might seem more or less similar at first glance. For instance, a universe where the Hobbist approach is true, motivation X being present at t_1 both can, and cannot, be followed by action Y at t_2 . This possibility for a counterfactual explains the confusion of labelling the Hobbist approach as deterministic.

There may be some external incident which happens, and which interferes with S' motivation, even if it is true that at every instance the universe was "restarted", S would in fact do Y because of X (if possible). In a deterministic universe, there can be no such incident, if it were not determined to begin with. Thus, on determinism, either it is the case that S will do Y because of X, or S will *not* do Y because of X. There is no room for multiple possibilities in a deterministic universe.

A difference between the Hobbist approach and determinism is thus that the former deals only with necessity regarding the relationship between *motivations and actions*, while the latter deals with the necessitated relationship between *all causal events*. The free will sceptic is thus not committed to determinism by incorporating the Hobbist approach in his theory of the impossibility, or improbability, of the existence of free will. The sceptic has in this way

¹⁰⁴ I must here emphasize that although Thomas Hobbes himself is known as an advocate of determinism, the Hobbist approach as discussed with respect to the views of Roderick Chisholm, does *not* imply determinism (as I shall explain in this current section).

therefore not departed from the free will discussion which presupposes the falsity of determinism.

“Rising above one’s desires”

A way in which Chisholm states that the Hobbist approach is unconvincing, is in claiming that one cannot argue that “... It is causally necessary that, given such and such desires and beliefs, and being subject to such stimuli, the agent will do so and so”. In addition, Chisholm states that “For at times the agent, if he chooses, may rise above his desires, or step aside, and do something else instead”.¹⁰⁵ What Chisholm seems to think is that an agent can “override” an initial impulse to do something, and thus that the agent can act on something which is not really his desire.

Let us suppose that S desires to eat a cake, but knows that doing this is unhealthy, and that this act will be in strict opposition to a diet S is currently on. What Chisholm implies, at least to my understanding, is that it is possible for S to annihilate – or override – a desire either without thus having one’s actions (which results from this) explained by yet another motivation – or, alternatively, doing something else without that something being explained by its own motivation. Not having a motivation for deciding not to eat the cake does not seem to make much sense, since S clearly has some kind of reason for not wanting to eat a cake (as well as wanting to eat the cake). If so, it follows that S does indeed have some kind of motivation for his action, and I would think that this is the case with any conscious (and thereby potentially free) action.

If S then overrides S’ initial inclination to eat the cake, S would – according to Chisholm’s view – have risen above his desire to eat the cake. But here Chisholm has shifted what he would have to mean by “desire”. S may have risen above the desire to eat the cake (desire₁), but he has – simply by virtue of doing another action – acted on another, separate desire (desire₂). For the very act of deciding not to eat the cake is itself based on a desire. This desire could, for instance, be based on a wish to lose weight (or some other inclination). But it does not really matter what the desire is. The point is that there is *some reason* why S decided to not eat the cake (and there would similarly be *some reason* why S decided to eat the cake, if S in fact did so).

¹⁰⁵ Chisholm 1964, 396.

Let us represent eating the cake with action₁, and not eating the cake with action₂. Chisholm and myself would presumably agree that desire₁ would result in action₁. But here Chisholm would seem to imply that action₂ is *not* entailed by desire₂. What else could he possibly mean by the statement that an agent can rise “above his desires”? Of course, an agent can rise above *a* desire to do X, but either an agent acts on a specific desire, or he does not. It seems to me that choosing to act implies some kind of desire towards a “want” to do the action in question (all things considered – or with respect a principle, such as that commonly called “the least of two evils”).

Concisely put, *the agent cannot have risen above his desire to do what he in fact did*. S acted for a reason, and that reason – however intricate, complicated or seemingly stripped of any conceivable desire S had to undertake that action – must fundamentally be understood by S *wanting something to happen by way of said action*. In this way, it seems to me impossible to *not* want to do what you in fact do, and to not necessarily do what you do because of what your specific motivation was with regard to the very action you committed.

It might seem quite simple to come up with an example of an action someone does, where there is no possible “want” involved; or where the person in question clearly does not wish to bring about a certain action. Admittedly, this might initially, seem like a good argument against my notion of the relationship between actions, intentions, and wants. Consider for instance - Chisholm or his equivalent might say - an example where someone is robbed. For simplicity’s sake, let us imagine that the person can either hand over their money, or pay with their life for refusing to do so. If the victim pays the robber, it is not because the victim *wants* to give their money away to a criminal. One would say that it is because, *all things considered*, that is preferable to losing one’s life.

Or, on the contrary, if the victim stands their ground, and refuses to compromise (because of rage directed towards criminals, or for some other reason), this is not because the victim *wants* to die as a result of a failed robbery. It is simply because, *all things considered*, the victim would rather stand up to the robber instead of letting him have his way. As with these examples, S does not need to actively *wish* to do X¹⁰⁶, to *all things considered* prefer X to not-

¹⁰⁶ The action in question can for instance be an example of S choosing “the lesser of two evils”. That is, an action which – all things considered – at a specific time might be the least terrible option (and thus, S *wants* to choose this action in the sense that it is preferable to the other options). Hence, the action is not something that S would “actively” want to do in the

X, in a given scenario. What I simply mean by saying that S necessarily does what S wants, is that S will act on whatever motivation of S' is the strongest, even if all available choices are bad ones. In this sense, S would presumably "want" to be robbed – merely in the sense that this is the best possible course of conceivable actions.

But how could Chisholm explain this, if he is unwilling to accept that the final, responsible motivation for S' action is the cause of the action? Unless one adopts an *ad infinitum*, infinite regress approach to the sets of desires and actions – which is clearly absurd (and incompatible with free agency) - one must at a certain point admit that a final motivation for an action, at least theoretically speaking, can be found and identified. Even if S is free at t_1 with respect to choosing between motivation X and motivation X*, as to which reason S will have for doing action Y, it would still be true that Y at t_2 is *either the result of X or X** (or, I could grant, perhaps some combination of these). Thus, notions such as a desire's ability to "incline without necessitating"¹⁰⁷ can only be relevant at t_1 .

Free will is, with respect to these considerations, still possible, since S may be free at a prior time slice with respect to an action which takes place at a future time slice. But that does not mean that S was not bound by S' free choice of motivation. Rather, S was, if in possession of free will, in a position to *choose S' motivation*, but S was not – and cannot – be in a position to choose a motivation, *and after this*, act on that which S deemed to *not* be what S primarily wished to do. The crucial point here is that the act of attempting to do so, would necessitate *another choice* of S' motivation, which brings us back to the initial point of S again doing what S wants. Even by wanting to act on a motivation which could not explain an action, *that would itself* be a want which would account for why S did the action he would in fact go on to do.¹⁰⁸

S would then have *changed* S' motivation in order to do another action, and that is what is implied by the statement of S only being able to do what S wants. Attempting to disprove this would have S running circles around himself, and constantly ending up doing something because that was what he wanted to do. Disagreeing with this notion would be to be one step

vast majority of possible scenarios; but which nonetheless can be the most favourable one in a situation where all other possible actions are even worse.

¹⁰⁷ Chisholm 1964, 396.

¹⁰⁸ In the same way, one cannot – for instance – demonstrate one's free will by actively choosing to eat something which one despises. Because one would then have wanted to prove a point. And thus, one is again back to the initial point of merely having done what one wanted to do in the first place.

behind the chain of events of different motivations and their necessity in explaining a rational action; and it would fail to realize that S would prove my point by simply changing his mind and doing another action.

The point is this:

S cannot escape what is necessarily implied by the very fact of what must be true of an agent's motivations for doing a certain action.

In this sense, I would argue that it is in fact nonsensical to say that S wanted to do X more than anything else, but that S did *not* do X. As far as I can tell, one simply follows from the other. That is, S does what S wants to do – and it is impossible for S to do what S does *not* want to do.

What this amounts to is that I claim that even a free agent is limited to only act freely in light of having *chosen* a motivation. Possessing free will does not mean that one would be able to choose not to want X, but still do X. That seems, to me at least, to imply a contradiction. To go briefly back to the example of the cake, eating the cake must necessarily follow from a “victorious” – and ultimately dominant - motivation to want to eat the cake (freely chosen as that motivation may or may not be).

Contrary, to *not* eat the cake must similarly follow from the opposite motivation – freely chosen or not (to not eat the cake). The mechanic of a free choice would, I suspect, be in choosing one's wants – and thereby, by necessity (*after choosing what one wants*) – do what one wants to do. Interestingly, what this would imply is that one can indeed be free to decide what one wants, but *not* free to do that which one does not want (again, this is *after* choosing one's wants).

3.4. Elaborating on control over the will – the importance of desires

First order desires and second order desires

Harry Frankfurt famously distinguishes between two sets of desires - in the first order and in the second order, respectively. First order desires are by Frankfurt described as desires which are "... simply desires to do or not to do one thing or another"¹⁰⁹. Second order desires are desires which reflects an agent's wish to have a set of motives and desires. These desires of the second order are thus representative of a "... capacity for reflective self evaluation ...".¹¹⁰ Second order desires are thus described as a set of desires which are higher and more complicated than first order desires; a set of desires which come about by reflecting, understanding, or weighting the different desires one has with respect to the desires of the first order.

This implies that there is something more to an agent acting than that agent necessarily acting *because* of his desires – which come about as if they were the desires of someone else.¹¹¹ On Frankfurt's account it seems that an agent with second order desires can, to a greater extent, be in control of his wants, and his actions. Free will might, in this way, seem more plausible if we could offer a coherent account of the relation between first order desires and second order desires. I think that a satisfactory explanation to the existence of second order desires would indeed be quite beneficial for the free will optimist.

I do not think, however, that second order desires can be meaningfully said to exist.¹¹² Thus, I wish to argue why it strikes me as improbable that it is possible to do something you do not want to do. Let us, though, examine a possible line of argumentation against this notion. Let us call this "an argument from coercion". The argument would go something like this.

¹⁰⁹ Frankfurt 1971, 7.

¹¹⁰ Frankfurt 1971, 7.

¹¹¹ In the sense that the agent is not in control of what he wants, and so he might as well be programmed, say, to want to do X – and still be as chained to his desires as in any other explanation for his wants.

¹¹² At least not in the way I understand them with respect to the free will debate.

“S cannot perform an action that S wants to perform, if S is put in a position where all of the alternatives of possible actions are such that S dislikes them”.

One could argue that external limitations on one’s freedom (such as a robber demanding your money or your life), means one was “forced” to choose X. And that one hence was making a choice which one did not want to make. But in the same way that a robber can “force you” to choose to give up the money, the laws of physics, or truths of logical necessity, for instance, can “force you” to make a choice about something which restricts you in an equally unwelcoming way.

Ought we then to be inclined to say that a set of possibilities for an action at a certain, arbitrary point become so bad that they negate the concept of a person *wanting* to perform one of the actions over the others? Perhaps I am here begging the question, since I am presupposing that the notion of doing the least of two evils necessitates that S – all things considered – wants to undertake the action which is the lesser of two evils¹¹³. I will shortly, however, defend this idea more thoroughly.

What I mean by stating that it strikes me as improbable that it is possible to do something you do not want to do, is that a statement of Frankfurt’s, such as “A does not “really” want to X”, is nonsensical in the way that it cannot be accurately said to describe the process of actions or desires among *any metaphysically possible agents*. Simply stated, I am not convinced that it is possible to partly want something at *t*, and that if you actually want something, you also “really” want it. What my position amounts to, is in essence that I equate first order desires with second order desires. Or, alternatively, that I deny the existence of second order desires.

A case for there only being first order desires

My argument goes something like this. While S can seemingly will both X and not-X, S will at *t* either do X or not-X; and that can only be explained by S either preferring X or not-X (S preferring X if S does X, and S preferring not-X if S does not-X). For again, how else would

¹¹³ I will shortly expand upon the distinction – or lack thereof – between first order desires and second order desires.

we understand S' choice, if not that the choice came about by way of S' preferences at *t* - *all taken into account*? For if S did not do X because S wanted to do X, then S doing X was surely not free. And if S did X for a reason which would be explained or accounted for by free will components, then S surely did X because S *wanted to do X*?

This presupposes that there is nothing external which manipulates the will of S. But given this presupposition, the only possible reason S acts a given way, is because S wants to do that very thing. And S necessarily, simply by way of having done what S did, wanted to do X instead of not-X at *t* (or vice versa). This appears to be true for free agents, and non-free agents alike (granted a lack of external interference). For no matter if S is free or not, in an undisturbed space, say, S will act as S wants to act.

Let us take Frankfurt's example of a case where "A wants to X".¹¹⁴ Frankfurt states that in such a case, the agent could want something different to X *simultaneously* as wanting X. Specifically, Frankfurt states that "It may be true that A wants to X when he strongly prefers to do something else instead".¹¹⁵ But what exactly is Frankfurt here saying? Obviously, if S wants to receive 1000 kroner, S in *this first sense* also wants to receive 500 kroner. But given the choice, S would presumably choose the former. Only during a scenario where S could not receive 1000 kroner, would S want 500 kroner in *this second sense*.

In the latter scenario, S wants to do both something which S does, as well as something that S does not do. But here we can identify the problem with saying that someone can "want to do something which he does not in fact do". At the very least it is misleading to say this if one by such statements means to deny the fact that there is indeed a *primary desire* of S, which accounts for S' action(s). For S can, in the first sense, want to receive 500 kroner, as well as choosing to receive 1000 kroner (and wanting this as well).

But here S does not want to receive 500 kroner in the sense that this "want" cannot be discarded by any other want in the relevant circumstances (or example). S' want to receive 500 is cancelled, so to speak, by the fact that S would prefer to receive 1000 kroner. What a statement such as "S both wants to do receive 500 kroner, as well as 1000 kroner", really ought to amount to, is something like the following.

¹¹⁴ Frankfurt 1971, 8.

¹¹⁵ Frankfurt 1971, 8.

S would want to receive 500 kroner, were it not possible to receive 1000 kroner.

Given the two options, S no longer in any meaningful way “wants to receive 500 kroner”, because there is something that S wants instead; and this something is mutually exclusive to S’ want to receive 500 kroner. S would prefer receiving 500 kroner to receiving nothing, of course, but then again – by the same reasoning – S would in that case want to receive 500 kroner instead of receiving 10 kroner, say.

The same logic is true if we go down this path further. S wants to receive 10 kroner, but what does this mean if he prefers, and wants in a stronger sense, to receive 500 kroner. Can we really say that S in such a scenario *wants* Y, when S would *prefer* X? My point is simply this: there is something unintuitive about saying that “S wants Y” in a situation where both a) and b) hold true.

- a) Y is made impossible, and negated,¹¹⁶ by X,
- b) S does X, and not Y

What do we mean by saying that “S wants Y”, if S wants X (and hence, not-Y) - and if S does indeed do X? S cannot in a meaningful way “want Y and X” in this sense any more than a predator, for instance, in the animal kingdom can want to kill a single pray at the same time as it, by its own preference, wants to (and does) kill two pray. In this way we can say that S can only want to Y – or *want Y to be undertaken in place of X - iff X were not possible* (given that Y and X are mutually exclusive and S decides to do X instead of Y at *t*).

All of this is bordering on semantical nitpicking, and in an obvious “everyday-sense” everyone knows what it means, of course, for S to want something which S does not act on. And in a similar fashion we all know what one means by saying that one wanted to do something which one did not do, because of a different alternative which was in fact preferred at the moment of choosing. But I think Frankfurt, and others, ought to be careful with

¹¹⁶ By negated, I here mean that Y is *in opposition* to X, such that action Y cannot be undertaken at *t* if action X is to be undertaken at *t* (or vice versa).

suggesting that the truth claims of statements which essentially argues that “S can want both X and not-X at the same time”, are not inherently problematic.

For if Y and X are mutually exclusive, and only one action can be undertaken at a time, then saying that S wants to do both Y and X is in this sense equivalent to the notion of S wanting both X and not-X. And here lies the confusion present in saying that S wants something at t which S does not in fact do at t . In summary, S cannot in a fundamental sense want that which necessarily conflicts, and is mutually exclusive, with S’ main and preferred desire. S cannot meaningfully want something to happen, if that something makes S’ main desire impossible to fulfil.

The aim of this part of the thesis

While adding a cumulative case to the denial of the existence of free will which I present in this thesis (with respect to the arguments from probabilism), I have not wanted to demonstrate the seeming impossibility of free will in this section. Here I have identified, and attempted to demonstrate, some of the *limitations which restricts the free will optimist in constructing a coherent and convincing theory and account of free will*. And thus, this third part of the thesis further builds on my case against the theory of free will. Specifically, I have highlighted some of the concerns I think the free will optimist ought to have about any account of free will in a probabilistic universe.

● Conclusion and final thoughts

In summary, I have in this thesis presented arguments which, most importantly, supports the following claims. Firstly, I have argued that we live in a probabilistic universe. Secondly, I have argued that the indeterministic properties of probabilistic universes are characterized by chancy components. And thirdly, I have argued that chancy components are incompatible with free will. Following this, I have argued that *within the framework of a probabilistic universe*, where chancy components seem to best account for the indeterminacy of sets of events within time slices, there are no satisfactory arguments, explanations, or theories which

in a convincing or sufficient way constitutes a convincing account of how such a thing as freedom of the will could operate in said universe(s).

In short, I argue that there is no apparent solution to the free will sceptic's well-founded (or so I argue) claim that the universe is probabilistic and characterized by chancy components. In this respect, the most substantive part of the thesis is the second. There may, however, be reasons to view the other parts as equally important, or perhaps even more essential, to this thesis' defence of free will scepticism.

In this thesis I have not concerned myself with the debate regarding the conflict between compatibilism and incompatibilism. This is primarily because I view this debate as a semantical one, instead of an ontological one.¹¹⁷ Thus, by way of the arguments I have focused on, I have in essence argued against the libertarian, since I have assumed in this thesis that determinism is *incompatible* with free will – in very much the same way that there is widespread agreement that randomness (and hence, a random universe exclusively characterized by raw chance) is incompatible with free will.

Thus, this thesis does not address arguments for and against (in)compatibilism. What I have attempted to refute is the claim that it is likely that free will is possible in our universe (and that agents with free agency inhabit it) – or *a* universe in which the metaphysical nature of states of affairs, or occurring events, are neither purely random, nor entirely deterministic in nature.

¹¹⁷ By this I mean that there is a correct answer to the question of whether or not libertarian free will exists, but that there cannot be a definitive answer to the question of whether or not the compatibilist is correct. The compatibilist and incompatibilist merely disagree on the definition of the term “free will” - and for this reason there cannot be a fact about the universe which would either prove or disprove the compatibilist's definition of the free will term. For example, the truth or falsity of randomness (that is, not probabilism) or determinism would either prove or disprove the truth of libertarian free will. But, the truth of determinism, on the other hand, would neither prove nor disprove the claims of the (in)compatibilist.

- **Bibliography:**

Ayer, AJ. 1954. "Freedom and Necessity". *Determinism, Freedom, and Responsibility*, 374-379. London: MacMillan.

Chisholm, Roderick. 1964. "Human Freedom and the Self". *Determinism, Freedom, and Responsibility*, 391-397. London: MacMillan.

Clarke, Randolph. 2005. "On an Argument for the Impossibility of Moral Responsibility". *Midwest Studies in Philosophy* 2005 (29): 13-24.

Fischer, John Martin. 2006. "The Cards That Are Dealt You". *The Journal of Ethics* 10 (1/2): 107-129.

Frankfurt, Harry. 1971. "Freedom of the Will and the Concept of a Person". *The Journal of Philosophy* 68 (1): 5-20.

Hartman, Robert. 2018. "Constitutive Moral Luck and Strawson's Argument for the Impossibility of Moral Responsibility". *Journal of the American Philosophical Association* 4 (2): 1-27.

Istvan, Michael. 2011. "Concerning the resilience of Galen Strawson's Basic Argument". *Philosophical Studies* 155 (3). 399-420.

Kane, Robert. 2009. "Introduction: The Contours of Contemporary Free Will Debates". *The Oxford Handbook of Free Will*, edited by Robert Kane, 3-42. Oxford Handbooks Online: Oxford University Press.

Mickelson, Kristin. 2019. "Free Will, Self-Creation, and the Paradox of Moral Luck".
Midwest Studies in Philosophy 43 (1). 1-33.

Strawson, Galen. 1994. "The Impossibility of Moral Responsibility". *Philosophical Studies:
An International Journal for Philosophy in the Analytic Tradition* 75 (1/2). 5-24.