

Could You Have Thought Differently?

An Argument Against Free Will

Nicolas Alzetta¹

Abstract: *This paper develops a new argument against free will, understood as the principle of alternate possibilities (PAP). This principle has been central in debates around free will and moral responsibility; however, it is almost always stated in terms of bodily, rather than mental action, and it is therefore mainly understood as the possibility to physically act differently, rather than to think differently. The argument presented here is aimed at the latter, which is termed the possibility of alternative thought (PAT). It argues on psychological grounds that it is impossible for a subject to think differently than it does in a given situation. First, I make the possibility of alternative thought explicit, explain what it means for a self to entertain a thought (or conscious content), and I define a notion of conscious control that is required for the argument. I then offer a taxonomy of content types, based on how they are being controlled. In the third section, I analyze the scenarios of content generation for each type of content, to determine whether they can satisfy the PAT. It will be shown that this cannot be the case.*

Correspondence:

Email: nicolas.alzetta@ucll.be

1. Introduction: The possibility of alternative thought

Perhaps the understanding of the term ‘free will’ which has been the most influential up until this day, is the idea that one has the possibility to do otherwise, to act differently than one did. The centrality of this particular understanding, also known as the *principle of alternative possibilities* (van Inwagen 1975, Ginet 1996, Kane 1996, Copp 1997), should not be surprising, as it captures the commonsensical intuition that if one cannot but act in a certain way, it becomes problematic to hold that person responsible. Traditionally, this principle has been

¹ UCLL University of Applied Sciences, Campus Hasselt, Oude Luikerbaan 79, 3500 Hasselt, Belgium.

refuted on metaphysical grounds, as being incompatible with a deterministic world (Lewis 1979). While I do not oppose such arguments, I think the principle can also be refuted on psychological grounds. I will therefore offer a new argument that does just this, while limiting myself to the mental domain. Hence, I will not ask whether it would be possible to act differently in a given situation, but rather ask whether it would be possible to *think* differently. To demarcate this from the principle of alternate possibilities, I will call this the *possibility of alternative thought (PAT)*. I will argue that this possibility cannot be satisfied, by showing that *no* conscious content could have been different than it is, at least not as a result of our own control. In section four, it will be shown as a corollary that if the PAT cannot be satisfied, then this must also be the case for the principle of alternative possibilities.

I will begin by defining the possibility of alternative thought and explain what it means for a *self* to entertain a thought (which will be henceforth be referred to as a conscious content), along with what it means for that self to entertain a different content as a result of its own control. In section 2, I will develop a taxonomy which categorizes contents by how they are being controlled. This will give us an overview of all the different ways in which contents can be generated, from a control perspective. This taxonomy is exhaustive in the sense that each possible content necessarily falls within one of the ten types in the taxonomy. In section 3, for each type of content, I will then analyze the different ways it can be generated, which I call ‘scenarios’, to determine whether they can satisfy the PAT. It will be shown that this cannot be the case. Since the taxonomy is exhaustive, it therefore follows that the PAT cannot possibly be satisfied. Before starting, I must add that my argument presupposes an event-causal view of action. In other words, I work under the assumption that every mental event is caused by some prior mental event; there are no uncaused (mental) events. It would not possible to transpose my argument to the two main alternative frameworks to the metaphysics of action, the non-causal and agent-causal views. Both seem problematic to me for a number of reasons, although I cannot go into them within the scope of the present article.

As a first point, we must state the possibility of alternative thought in a more exact form:

Possibility of alternative thought (PAT): for a particular mental content entertained by a self, that self could have entertained a different mental content than it did, as a result of its own control.

To give this meaning, a few things need to be clarified. First off, we need to be clear about what it means for a self to entertain a mental content. I use the term ‘content’ in a somewhat more liberal way than is typically the case in philosophy of mind. Here, the term will be used to indicate a *part of our total conscious experience, which can be used or referred to as such*. Thus, if you successfully imagine a green dog, there is a more or less unified complex of shape and color which can (but does not have to) be used for a variety of purposes: making it bark, having it recite some Rimbaud, etc. Also, you can refer to the dog as an imagining of yours. Somewhat less straightforwardly, a random green patch in the middle of this unified complex also counts as a content on this definition since you can alter it to red, and, in doing so, you have *de facto* referred to it, albeit implicitly since the alteration renders you in a position to retrospectively assert that you have altered the color of that patch. To be clear, the fact that you refer to a content does not mean that you have to be able to ‘fully’ grasp it by giving an accurate description of every experienced property. It simply means that one is capable of referring to the fact *that* it was experienced.

We still need to be more explicit about what it means for a *self* to entertain a content. This can seem difficult since there are many notions of self,² although for our purposes the notion which is at stake is that of the *minimal* self. The minimal self, also referred to as the *phenomenal* self, is engendered by our conscious experience and the possibility it allows to refer to itself, thus making us self-aware subjects (for detailed accounts, see Metzinger 2004; Dainton 2008; Zahavi 2008). Fully specifying the notion of the phenomenal self is not needed to make the argument, however. All we need is a way to specify what it means for a self to entertain a content. If I am able to identify a content as mine, as being ‘in’ *my* consciousness, this is sufficient to say that I (a self) entertain the content. I might be wrong about its origin, I might identify and describe it wrongly, but it seems I cannot be wrong about the fact *that* there is a content in my conscious experience. For example, I might think it stems from perception when, in fact, I am hallucinating. This does not pose a problem for our purpose since we do not need to infer anything extra (like knowledge about the world) from the fact that I identify a content as mine, other than the fact that I am the one who is entertaining that content. Adding this stipulation to what we know about what it means for a content to be conscious, we get the following definition:

² Indeed, Galen Strawson (1999) identified no less than twenty-one distinct types of selves.

Self entertaining a content: A self entertains a mental content if this content is conscious, it (the self) is capable to refer to the content and can identify it as being in its consciousness.

The last part we need to clarify is what it would mean for a self to have entertained a *different* content, *as a result of its own control*. First of all, at least some element of the control needs to be conscious. This is to ensure that our definition of control captures the types of decision-making and (rational) deliberation that we typically regard as central to our agency and moral practices. I would argue that *if* it was true that we could have acted differently than we did, then a case where a past action was altered or avoided as a result of a conscious decision would have a different moral status than a case where this was a result of merely unconscious processes or randomness. This idea matches a deep-seated folk intuition that consciousness is crucial for free will and moral responsibility (for experimental evidence of this intuition, see Shepherd 2015). This intuition is not merely a folk intuition, however, as reflected by the magnitude of the debate surrounding the conscious will, following Libet's experiments (Libet et al. 1993; Wegner 2002; Mele & William 2009).

Second, in order to control a mental process, you need to be aware to some degree what the desired outcome, or *goal*³, of the process will be. When asked, you should be able to say what a/the desired outcome of the process would be. In other words, you need to be able to stipulate a satisfaction condition of the goal of the process. Note that this doesn't mean that the goal is necessarily conscious before the initiation of the process, or that it must be conscious during the whole process: all that is needed is that the goal is conscious to *some* degree, because otherwise it would not be possible to stipulate a satisfaction condition, since you cannot specify how to satisfy a goal you are not aware of. Another condition for conscious control is that the

³ A goal (or *task* goal as it is often referred to in the psychological literature) encodes information about (a)(mental) state(s) that need to obtain at the end of an action task (or at certain points during the action, in case it consists of sub-actions). This doesn't mean that the end state of an action is fully specified by its goal. This is clearly not the case because if it was, the goal to recall the name of your second-grade teacher would already contain the name of that teacher. Nevertheless, *something* in the goal must specify that the action in question is a recall task and that the object of the recall is a name of a particular teacher, for otherwise, it wouldn't be clear why this specific recall task was executed (say, after an instruction to do so) and not any other random mental operation. These aspects of the goal (that it is a recall task, that the object of the recall is a name, etc.) can also be viewed as its *satisfaction conditions*. It is presently unclear how these satisfaction conditions are encoded by the task goal and how rich this encoding is (i.e., how much of the end state of the action is actually specified by the goal) (Vandierendonck 2016). One thing that seems necessary, though, is that these conditions *can be* encoded in a non-linguistic or non-conceptual manner, since animals and small children are capable of performing (complex) goal-directed behavior. The fact that the notion of a goal is not fully specified at this stage of inquiry, is not a problem for the purposes of the argument presented here. The fact *that* there are (task) goals will be sufficient to inform the argument.

process which is being controlled can be altered as a result of the awareness of that process. If you are not able to change the course of events and you have no choice but to pursue the goal you have set yourself, then this process is ballistic and not controlled, despite there being a goal. Together, these stipulations give rise to what I call the *minimal conscious control requirement*:

Minimal conscious control requirement (MCCR): 1) when asked, the subject is capable of stipulating a satisfaction condition of the goal of the process 2) the subject is capable of altering that process as a result of her awareness of it.

If both conditions are met for a particular process, then this process satisfies the minimal conscious control requirement. Another way of putting this, would be to say that the process was consciously controlled in the minimally required way. By extension, a content which is the result of such a process will also be said to satisfy the MCCR. Everything is now in place to develop the argument. I will proceed as follows: first, I will develop an exhaustive taxonomy of control types, based on their antecedent control structure (e.g., whether they result from a conscious or an unconscious goal, from a non goal-containing state like a desire, etc.). In section 3, I will analyze the different ways (scenarios) in which each type of content from the taxonomy can be generated. It will be shown that none of them satisfy the PAT. Since the taxonomy is exhaustive (in the sense that it captures every type of content), this implies that the PAT cannot possibly be satisfied.

2. A control-based taxonomy of content types

First, a goal (conscious or not) can embed or be embedded (by) other goals. For example, the goal to remember the name of your second-grade teacher might be embedded in a higher-level goal to recall the names of all your schoolteachers. Similarly, the goal to multiply 12 by 17 embeds the goal of multiplying 12 by 10 if that latter calculation is a step you take to arrive at the result of the former. Whether a goal is embedded or not will not change anything to the argument, as will be made clear in the next section. I therefore did not include it in the taxonomy below. Once a content is conscious or if it was previously conscious, one can have a certain attitude towards it. This can take many forms: believing the content to be true, wanting it, rejecting it, endorsing it, having an emotional response toward it, etc., or a combination of any of those. I will call any one of such attitudes (or their combination) a *valuation*. In certain

cases (beliefs, desires/wants, imaginings, and the like), the combination of content and valuation is simply a propositional attitude. A propositional attitude can be considered as a content in its own right, which can be the object of a valuation. This will be developed more fully below in the exposition of type VII contents. For our purposes, the valuations which will primarily be considered are those which are either ‘in favor of’ or ‘against’ executing a certain goal, although such valuations can themselves take on a number of different (phenomenological) guises. Note that a valuation is itself a content, so the taxonomy is also applicable to the generation of valuations. In total, there are ten types of contents, some of which have sub-types, as listed in the table below:

I	A) Directly resulting from a conscious, sufficiently temporally separated goal with valuation B) Directly resulting from a conscious, sufficiently temporally separated goal without valuation C) Directly resulting from a conscious, non-sufficiently temporally separated goal
II.	A) Semi-directly resulting from a conscious, sufficiently temporally separated goal with valuation B) Semi-directly resulting from a conscious, sufficiently temporally separated goal without valuation C) Semi-directly resulting from a conscious, non-sufficiently temporally separated goal
III.	A) Indirectly resulting from a conscious, sufficiently temporally separated goal with valuation B) Indirectly resulting from a conscious, sufficiently temporally separated goal without valuation C) Indirectly resulting from a conscious, non-sufficiently temporally separated goal
IV.	Directly resulting from an unconscious goal
V.	Semi-directly resulting from an unconscious goal
VI.	Indirectly resulting from an unconscious goal
VII.	A) Resulting from a conscious, non goal-containing mental state with valuation B) Resulting from a conscious, non goal-containing mental state without valuation
VIII.	A) Resulting from an unconscious, but previously conscious, non goal-containing mental state with valuation B) Resulting from an unconscious, but previously conscious, non goal-containing

	mental state without valuation
IX.	Resulting from an unconscious and not previously conscious, identifiable, non goal-containing mental state
X.	Not resulting from an identifiable mental state (goal containing or otherwise)

Table 1: a control-based taxonomy of content types

Type I contents result from a goal which ‘directly’ targets a piece of content. For example: “I’m going to remember the name of my second-grade teacher,” followed by the name of the teacher. ‘Directly’ means that the goal stipulates a single satisfaction condition, namely, remembering the name of the teacher in question. There is a one-to-one mapping between the goal and the outcome, so to speak. Either the goal can be temporally separated from the content, or it can be ‘attached’ to it. This distinction is somewhat similar but not identical to the one between F- and P-intentions. Basically, if the goal is not immediately followed by the content targeted by it, there is time to perform an attentional switch away from the goal and onto a content which evaluates it, hence assigning a valuation to the goal. I call this type of goal *sufficiently temporally separated*. Note that the length of this temporal separation is not fixed and can be very short; it only needs to allow for one attentional switch away from the goal and onto (an) evaluative content. Of course, it is not necessarily the case that such a temporal separation engenders a valuation, hence type I-B. In the third case, the goal is immediately followed by the content it targets, so there is no valuation. By way of illustration, in its simplest form, a I-A scenario can be depicted as follows:

$$1) \quad G_c \rightarrow V_c \rightarrow C_c,$$

where G_c is the goal (the subscript c indicating that it is conscious), V_c is the valuation of that goal – the attitude one has towards it – and C_c is the generated content. C_c becomes conscious at time t, V_c at t-1 and G_c at t-2. These time indexes (t-1, ..., t-n) do not represent actual time units but simply indicate where in the sequence of events a certain step is located. Note that a I-A content can be generated in multiple ways. For example, there can be many valuations, each of which can be the result of a goal in itself. We will be looking at some of these more elaborate scenarios in the next section.

Type II contents are almost identical to type I contents, except that they semi-directly result from a conscious goal. This means that the goal from which they result has multiple

satisfaction conditions. For instance, when you set yourself the task to remember the name of one of your primary school teachers, multiple contents can result from that which satisfy the goal. There is a one-to-many mapping between the goal and the outcomes which satisfy it. Another example would be to imagine a green elephant. The imaginings which satisfy this goal can take many forms, as long as they satisfy the condition of being a green elephant.

A type III content *indirectly* results from a goal. It is goal-dependent, meaning that without the antecedent goal, the content would not be generated. Indirect means that the content does not satisfy any of the satisfaction conditions of the goal. As an example, imagine you are about to calculate $12 \times 7 + 5$ (which is written without parentheses on a test), and the first thought after seeing the problem is ‘multiplication first.’ This thought does not satisfy the satisfaction condition of the goal (in this case there is only one), but it nevertheless results from it. Although type III contents can be generated after valuation of the goal, the valuation here plays a somewhat different role than in type I and II cases. As we know, the valuation pertains to whether it is desirable, correct, feasible, etc., to generate *the* or *a* content stipulated by the goal. If sufficient, it initiates the process which makes this content conscious. In a type III case, although the valuation might also have caused the initiation of the process generating the content, this did not happen in a way that was ‘prescribed’ by the goal-valuation pair.

Types IV-VI are the unconscious variants of the first three types, meaning that they are conscious contents which result from unconscious goals. It is well known that only part of the control we exert over our actions is conscious (Lau & Passingham 2007; Van Gaal et al. 2011; Hommel 2017). When we engage in actions, especially in routine or skilled ones, we are typically only conscious of high-level goals, while the sub-goals required to accomplish them are unconscious. As an example of a type IV content, suppose you are calculating 12×17 and you become conscious of the number ‘120.’ As it turns out, this content results from the unconscious sub-goal to calculate 12×10 first and then add it to 12×17 . In this example, the strategy of breaking up the calculation in two simpler calculations has been sufficiently practiced to render it ‘automatic,’ meaning that no conscious control is needed to execute it. For a type V content, consider a chess game where an expert player becomes conscious of the content ‘rook B5,’ without prior conscious goal. As it turns out, this move is required to shield the queen from an impending attempt to capture it. Moving the rook to B5 is one of the moves which satisfy the unconscious goal of shielding the queen, although others were also possible. The expert did not have to formulate a goal to shield the queen; it is unconsciously ‘running along’ during the game and might only become conscious if the decision to shield the queen needs to be revoked in order to gain a strategic advantage by sacrificing it. Staying with this

example, suppose the content ‘don’t expose your castle’ becomes conscious. As it turns out, both rooks can perform the task to shield the queen, but in doing so, moving one of them would leave your castle exposed. This is an example of a content which is indirectly controlled by a subconscious goal (type VI). Although type IV-VI contents might also result from unconscious valuations, I have omitted these distinctions in the taxonomy since they are of no importance to the present argument, which is only concerned with cases where there is at least some conscious element of control.

Type VII contents result from a conscious, non goal-containing state (a mental state which is devoid of any goals) In this case, the state is itself a content or a complex of contents, which can come in two forms: bound (VII-A) or unbound (VII-B). Examples of bound cases are propositional attitudes like beliefs, desires, judgments, etc., where you have a target content P which is valued in some respect (desired, believed, etc., or combinations thereof). A bound content can, of course, be the subject of another valuation, but this simply results in a bound content, which still consists of a target P plus a valuation. Therefore, it is not necessary to posit a supplementary distinction like ‘double-bound.’ An example of a type VII-A content would be the imagery of a glass of water, which is generated after you had a desire for a glass of water. In the unbound cases, the non goal-containing state is without valuation. Examples of such states can include mental imagery, inner speech, mathematical objects, etc.⁴ An example of a type VII-B content generated in this way would be the imagery of a triangle after the number 3 was conscious. It might, of course, be that the imagery of the triangle is in no way related to the number 3 being conscious, but in that case, it is not a type VII content. Again, to say that this content *results* from a conscious non goal-containing state means that without it, the content would not have been generated, making it a necessary condition. The triangle could also have been the result of the number 3 being conscious, together with the fact that you saw a square before it. In that case, each content is a necessary condition, and together they form a complex, which constitutes the non goal-containing state. One could argue that a content is always being subjected to a valuation in some (minimal) way and therefore, that there are no unbound contents. This may be the case for the content to which the valuation is bound, but it is unlikely to be the case for the valuation itself, under normal circumstances. For example, we do not necessarily have an emotion about our emotion associated with the mental image of our cat. Nonetheless, even if contents were necessarily bound, this will not alter my argument or its validity, as will become clear in the next section.

⁴ This does not mean that the latter examples cannot be the target of a valuation.

A type VIII content is the unconscious variant of a type VII one. Here, the state from which the content results is not conscious prior to its generation, although it has been at some prior time. As with type VII, the state from which the conscious content results is itself either a bound (VIII-A) or an unbound (VIII-B) content. Using the examples from above, imagining a glass of water after having had a desire for a glass of water at some prior time is a type VIII-A content, provided that the desire was responsible for the generation of the imagining. Similarly, the imagery of a triangle counts as a type VIII-B content if it results from seeing/hearing (being conscious of) the number three at some prior moment; again, provided that the content of the number was responsible for the generation of the imagery.

A type IX content is structurally similar to type VII and VIII contents, except that the mental state responsible for its generation has not been conscious at any prior time. The content ‘pops’ into consciousness for no apparent reason, but after reflection, it can be linked to mental states which are not related to the task(s) at hand. Imagine you suddenly have the thought “I should read more Shakespeare,” while working on a paper which is in no way related to Shakespeare. As it turns out, you have a desire to develop your understanding of interpersonal psychology. You also have a belief that reading Shakespeare can contribute to that. Both the desire and the belief have not been conscious prior to having the thought in question. It is by reflecting upon the thought that both the desire and the belief become apparent. You might, of course, be wrong in judging that these states were responsible for generating the content. This poses no problem, but to qualify as type IX content, eventually, a correct judgement about the cause of the generated content must be identified. If this is not the case, we simply have a type X content. As with type VII and VIII contents, the mental state responsible for a type IX content can be bound or unbound, but this distinction is left out of the taxonomy since it has no relevance for the current argument.

A type X content is unrelated to activities you are undertaking and simply ‘pops’ in your head, as when a green elephant appears in your mind’s eye for no apparent reason while you are working on a paper. Not only is it not task related, but it is also not linked to any goals, desires, beliefs, fears, etc., you might have.⁵ This type of content appears to be random, although it most likely is not. As any conscious content, it most likely serves some purpose in the cognitive economy in light of current or future behavior. What this purpose amounts to is unknown, however, unlike in type IX cases, where the content can be linked to other mental states.

⁵ This does not mean that once it is conscious it is necessarily unbound (i.e., cannot be the subject of a valuation).

3. Scenario analysis

Each type of content from the taxonomy can be generated in a number of ways, which I will call scenarios. The example of a I-A content given above, $G_c \rightarrow V_c \rightarrow C_c$, is one such scenario. In this section, I will analyze the possible scenarios for each content type, to determine whether they can satisfy the possibility of alternative thought. In other words, I will determine whether it would make sense to say that, for a content resulting from a certain scenario, a different content could have been entertained as a result of the (conscious) control (by the self entertaining that content). For this to be the case, it will be shown that at least one content within the scenario needs to satisfy the MCCR. If no content within the scenario satisfies the MCCR, then no content could have been different than it was (as a result of an element of conscious control⁶) and as a result, the whole scenario could not have generated a different content at the end; therefore, the PAT is not satisfied for that scenario. To be clear: when I say that the PAT is satisfied by a scenario, this means that the content generated *at the end* of that scenario could have been different as a result of some element of conscious control. The MCCR, in turn, applies to *single* contents *within* a scenario (e.g. ‘ V_c ’ in the example above). The stipulation that the MCCR is only applicable to contents within a scenario implies that the first content of a given scenario will itself not be analyzed to see if it satisfies the MCCR, since that would extend the analysis outside of that scenario. The first content is a *given*, from which we try to determine whether the content at the *end* of the scenario could have been different. Appealing to a different causal history that precedes the scenario will therefore not be allowed in order to make the possibility of alternative thought true. This excludes counterfactuals like ‘had I been born in China, I would have entertained a different content when conjuring up the word for the four-legged animal English speakers refer to as dog.’

I will work backwards from the least controlled (type X) to the most controlled types of content (type I). For type X scenarios, in which contents are not resulting from an identifiable mental state (goal containing or otherwise), it is clear that the possibility of alternative thought

⁶ The content might have been different as a result of purely unconscious factors, like indeterministic element(s) or randomness in the processes which generated the content. This randomness *might* result from quantum indeterminacy effects which are somehow amplified by non-linear dynamics (chaos), but it is highly contentious that such effects (if they even exist) end up having any computational relevance and in what capacity the brain follows non-linear dynamics (Honderich 1993, ch. 6; Bishop 2002; Walter 2009, ch. 3). Nevertheless, in case these effects are indeed computationally significant, the content could indeed have been different than it is. This doesn’t affect the argument, however, as some element of conscious control is needed to satisfy the PAT.

cannot be satisfied. If nothing which anteceded the content was conscious (neither goal nor valuation), then surely, there was no conscious control⁷ over the content. The content therefore could not have been different than it is, at least not in a way that involves a self which brings about the alternative. We know that although the content did not result from an identifiable mental state, it nevertheless is conscious, which means that it either had some relevance for current or future behavior or it became conscious as a result of randomness. In case the content became conscious for its relevance in current or future behavior, this was determined unconsciously. Even if, as a result of randomness, a different content would have been ‘found’ relevant and become conscious, this would again not have been the product of any conscious control, and therefore the PAT would not be satisfied.

For type IX scenarios, in which contents result from an unconscious and not previously conscious, identifiable, non goal-containing mental state, although conscious control is absent in the generation of the content, one could ask whether the states which were responsible for its generation could warrant the conclusion that a different content could have been entertained. Take the Shakespeare example, where the content ‘I should read more Shakespeare’ results from a previously unconscious desire to develop your understanding of interpersonal psychology and a previously unconscious belief that reading Shakespeare can contribute to that. Unconscious factors like randomness aside, the only way the content could have been different is when the desire and/or the belief had been different. However, since there is no conscious control over either of them, this could have only been the case if at some prior moment, the history of mental states had been altered. That would alter the scenario we are considering, though, since we are only concerned with the generation of content, given a fixed history of mental states until $t-n$. In this case, $t-n$ is the moment when the belief and/or the desire initiate(s) the process which generates the content (depending on which comes first and on whether they occur simultaneously). Although this example uses beliefs and desires, the same line of reasoning holds for all type IX scenarios, regardless of the type(s) of mental state(s) involved in the generation of the content. Type IX scenarios can therefore not satisfy the possibility of alternative thought.

In a type VIII scenario, a (non goal-containing) mental state responsible for the content was conscious at a previous time but is unconscious immediately prior to the generation of the content. What we have to consider is whether the fact that the state was conscious at some prior

⁷ In the following, when I say there was no conscious control or no conscious control was exerted, this will be synonymous for saying that the minimal conscious control requirement is not satisfied.

time would make a difference. Suppose you had a belief that the coming election will be rigged. Then, at a later date, for no apparent reason, you have the thought ‘I’m not going to vote.’ There are two options: either you assigned a valuation to the belief when it was conscious, or you did not. In the latter case (VIII-A), you did not try to deny this or convince yourself that, in fact, the election was not rigged. This way, the belief remained unchanged, and at a later time, it was responsible for the generation of the content, albeit unconsciously. No conscious control was exerted, so you could not have entertained a different content; again, not counting unconscious factors like randomness. If you *had* valued the belief (VIII-B), two possibilities present themselves: either the belief changed, or it remained unchanged despite the valuation. In the latter case, the situation is identical to that without valuation, where the belief ends up generating the original content. Suppose now the valuation did change the belief, which in turn generated a different content. Visually, this would look as follows⁸:

$$2) \quad S_c \rightarrow V_c \rightarrow S'_u \rightarrow C'_c$$

where S_c is the conscious belief, V_c is the valuation and S'_u is the new unconscious belief which is responsible for generating C'_c , the altered (conscious) content. For this scenario to satisfy the possibility of alternative thought, V_c must satisfy the MCCR, because if it doesn’t, no element of conscious control contributed to the generation of C'_c , since S'_u is unconscious. Suppose it became conscious right after S_c , without any intermediate conscious states. For example, after having had the thought that the election will be rigged, you deny it by thinking ‘this can’t be true.’ When it arises in this manner, V_c does not satisfy the MCCR since it is not the result of any goal. Assume then that V_c *was* the result of a goal, so you get:

$$3) \quad S_c \rightarrow G_c \rightarrow V_c \rightarrow S'_u \rightarrow C'_c,$$

where the goal (G_c) would be something like: ‘deny that the elections are rigged,’ which is then followed by the valuation V_c , ‘Belgium scores very high on various democracy indexes; therefore, the elections can’t be rigged’ (V_c satisfies a satisfaction condition of G_c). The $G_c \rightarrow V_c$ part of this sequence is itself a type I scenario. In order to determine whether (3) can satisfy

⁸ Note that the Arabic numbers used below do not refer to the types of content that are depicted in the scenarios. Content types are indicated with roman numerals. The Arabic numerals used below are simply there to number the scenarios and allow easier referencing.

the possibility of alternative thought, we therefore need to determine first whether the PAT can be satisfied by type I scenarios.

In type I scenarios, contents directly result from a conscious goal (temporally separable or not), with or without valuation. I will start with I-A (temporally separable goal with valuation) by using the example of the second-grade teacher (I-B and I-C will be analyzed along the way). There is a goal to recall the name of the teacher, followed by a valuation, followed by the name of the teacher. Visually the scenario is depicted as:

$$4) \quad G_c \rightarrow V_c \rightarrow C_c,$$

with G_c being the goal to recall the name, V_c is the valuation, and C_c being the (conscious) name of the teacher. Note that what we are asking is whether it would have been possible to recall something else than the name of the teacher *in this scenario*; in other words: given the goal to recall the name of the teacher at t-2. There are two options: either the valuation contributed to the process which made C_c conscious, or it did not. In case it did not, we are left with two options: in the first case, C_c was caused directly by G_c , meaning that the process P1 leading to the generation of G_c set in motion (caused) another process P2 which led to the generation of C_c . Since the valuation did not contribute, the process was ballistic: once the goal was set, there was no conscious way to avoid the content from being generated, and therefore, C_c does not satisfy the MCCR. In the second option, G_c (or the process underlying it) was in no way implicated in the generation of C_c . This is, of course, unlikely: why would C_c ‘randomly’ become conscious shortly after the goal to make exactly that content conscious was formulated? Nevertheless, we need to consider it as a theoretical possibility. It poses no problem, though, since if C_c is not the result of a prior goal, it does not satisfy the MCCR. Therefore, in both cases, since V_c does not contribute to the generation of C_c , and C_c itself does not satisfy the MCCR, the PAT is not satisfied. Since the valuation does not contribute to the generation of the content, both cases discussed here are de facto I-B or I-C scenarios, depending on the temporal distance between the goal and the content. Neither of these can satisfy the possibility of alternative thought since if there is no valuation involved, there is no way for C_c to satisfy the MCCR.

We are left with the first alternative, in which the valuation did contribute to C_c becoming conscious.⁹ Assuming the goal is fixed (G_c), there are four *basic* ways an alternative

⁹ Whether it was the sole cause is not important; what matters is that it was *necessary* to generate C_c .

content (C'_c) could have come about in a I-A scenario. There are, of course, infinite ways in which such an alternative could come about, but they constitute combinations of these basic scenarios. The only basic one that was left out is $G_c \rightarrow C'_c$ since it was shown in the previous paragraph that this cannot satisfy the PAT. The scenarios are the following:

- 5) $G_c \rightarrow V_c \rightarrow C'_c$
- 6) $G_c \rightarrow V'_c \rightarrow C'_c$
- 7) $G_c \rightarrow V'_c \rightarrow G'_c \rightarrow C'_c$
- 8) $G_c \rightarrow V'_c \rightarrow G'_c \rightarrow V''_c \rightarrow C'_c$.

In the first case (5), the valuation is ‘in favor’ of C_c , but still, another content C'_c became conscious. Assuming V_c was involved in the generation of C'_c ,¹⁰ we have a failure of the mechanism which under normal circumstances produces a content which matches the valuation. This is a failure of control, and in such a case it becomes impossible to think or act according to preference. Because of the mismatch between goal, valuation and content, the scenario doesn’t satisfy the PAT: there is simply no way to assert truthfully that C' was generated as the result of conscious control in this scenario. In the second case (6), the alternative valuation V' ‘discards’ G_c as a goal, which then leads to an alternative content C'_c . If this valuation became conscious right after G_c (without any intermediate steps), it does not satisfy the MCCR since it simply popped into consciousness, and there was no conscious control over it. For this scenario to have been obtained in a controlled way and not be the product of chance or randomness, V'_c must satisfy the MCCR, however. Since it does not, the scenario cannot satisfy the PAT.

One could object and say that V'_c could result from a new goal G' (with or without valuation) and therefore does satisfy the MCCR. In that case, we would have the following two possibilities:

- 9) $G_c \rightarrow G'_c \rightarrow V'_c \rightarrow C'_c$
- 10) $G_c \rightarrow G'_c \rightarrow V''_c \rightarrow V'_c \rightarrow C'_c$

¹⁰ As shown above, if V_c was not involved in the generation of the content, C'_c would not have been the product of conscious control. This option can therefore be discarded.

where in the first case (9), V'_c comes about without valuation, while in the second (10), it does result from one (V''_c). The problem is simply pushed back, however. In the first scenario (9), since there is no valuation involved in the generation of V'_c , it is a direct result of G'_c .¹¹ Here, the only way V'_c could have been different is if G'_c had been different. Again, this could have been due to unconscious factors like randomness, but it is not the result of any conscious control, since G'_c is not the result of a prior goal and therefore does not satisfy the MCCR. This being the case, V'_c could not have been different than it is (randomness aside), and therefore, there is simply no way to assert truthfully that C'_c could have been generated as the result of conscious control. The PAT is therefore not satisfied by this scenario. The second scenario (10) is no different: here, the second valuation V''_c acts as an intermediate between G'_c and V'_c . Again, this valuation is not the product of a goal (remember G'_c is the goal to generate V'_c), so it does not satisfy the MCCR. If neither G'_c nor V''_c satisfies the MCCR, V'_c could not have been different than it is, and we have the same conclusion as before: it is not true that C'_c could have been generated as the result of conscious control. The objection could be repeated, as one could insist that G'_c or V''_c are themselves the product of a goal (with valuation). Sure enough, but then we are led to the same conclusion for both cases since neither the goals nor the valuations responsible for generating G'_c or V''_c will satisfy the MCCR. This objection can also not be raised indefinitely since any sequence between G_c and C'_c will be composed of a finite number of steps. That this must be so is simply the result of the fact that the time between G_c (occurring at $t-n$) and G'_c (occurring at t) is finite, and each step requires time to unfold. In conclusion, we see that there is no way in which the possibility of alternative thought can be satisfied for this scenario.

Scenarios 7 & 8 are really variations on the same theme. In 7, V'_c gives rise to a new goal G'_c to generate content C'_c . As before, V'_c comes about without intermediate steps, so it does not satisfy the MCCR. G'_c is in the same boat – it does not result from any goal. C'_c is a direct result from G'_c , however, so it would be false to assert that it could have come about as a result of conscious control. The last scenario (8) leads to the same conclusion: The only difference is that the goal G'_c to generate C'_c receives a valuation V''_c . The latter does not satisfy any satisfaction condition(s) of any antecedent goals and therefore does not satisfy the MCCR. Since neither V'_c , G'_c nor V''_c satisfy the MCCR, there is no way to assert truthfully that C'_c could have been generated as the result of conscious control; thus, again, the PAT is

¹¹As above, we can ignore the case where V'_c does not result from G'_c since in that case, V'_c is accidental and not the result of any conscious control.

not satisfied. The same objections as before can be raised: insisting that G'_c and/or V''_c are themselves the result of prior goals, in an attempt to make them satisfy the MCCR. This would be to no avail and generate the same type of regress as before, which cannot go on indefinitely because of the finite time interval between G_c and C'_c . We therefore see that no possible type I scenario can satisfy the possibility of alternative thought.

In section 4.2, I said that it does not matter to the argument whether a goal is embedded in a higher-level goal or not. That this must be so is easy to show. First, a higher-level goal is simply an antecedent goal which ‘enslaves’ a lower-level one. In the hypothetical scenario $G_c \rightarrow V_c \rightarrow G'_c \rightarrow G''_c \rightarrow V''_c \rightarrow V'_c \rightarrow C'_c$, both G'_c and G''_c are embedded goals, while G_c is the highest-level goal that embeds the former two. This scenario is nothing but a combination of smaller scenarios: in this case, the sequence $G'_c \rightarrow G''_c \rightarrow V''_c \rightarrow V'_c \rightarrow C'_c$ is simply scenario 10 from above, while $G_c \rightarrow V'_c \rightarrow G'_c$ is scenario 6, both of which cannot satisfy the possibility of alternative thought. When considered as part of a larger sequence, scenario 10 still cannot satisfy the possibility of alternative thought: the only way for C'_c to have been different is if V'_c had been different and satisfied the MCCR, which is not the case. It therefore does not matter for the analysis whether the goal(s) of a scenario under consideration is/are embedded by (a) higher-level goal(s). More generally, since every possible scenario is a combination of the scenarios analyzed here, if none of the parts satisfy the possibility of alternative thought (which is necessarily the case), then their combination cannot either.

Coming back to type VIII contents (resulting from an unconscious, but previously conscious, non goal-containing mental state with or without valuation), the same conclusion has to apply: a type VIII scenario cannot satisfy the possibility of alternative thought. To see this, recall scenario 3, $S_c \rightarrow G_c \rightarrow V_c \rightarrow S'_u \rightarrow C'_c$, where in our example, S_c is the conscious belief that the elections will be rigged, G_c is the goal to produce a negative valuation (‘Deny that the elections are rigged’) and V_c is the valuation (‘Belgium scores very high on various democracy indexes; therefore, the elections can’t be rigged’), which results from G_c and satisfies a satisfaction condition stipulated by it. S'_u is the new unconscious belief (‘The elections won’t be rigged’), which is responsible for generating C'_c , the altered (conscious) content (‘I’m going to vote’). We have already established above that the only way for this scenario to satisfy the possibility of alternative thought is if V_c satisfies the MCCR. We know from our analysis of type I scenarios that this cannot be the case, however. V_c directly results from G_c , so it results from a ballistic process which does not satisfy the MCCR. Even if there had been a valuation V'_c between G_c and V_c , the analysis above has revealed that this scenario cannot possibly satisfy the possibility of alternative thought either, so V_c could not have been

different than it is (as the result of conscious control), and therefore, neither S'_u nor C'_c could have been different than they are. Consequently, no type VIII-A scenario can satisfy the possibility of alternative thought. This is necessarily also the case for VIII-B scenarios, which are without valuation: if no scenario *with* valuation(s) can satisfy the possibility of alternative thought because none of the necessary contents satisfy the MCCR, then this will a fortiori also be the case if no valuation is present.

The only difference between type VII and type VIII scenarios is that in the former, the state responsible for the generation of the content is conscious immediately prior to the content, whereas in the latter case, it was unconscious. We see that an alternative content C'_c can only be generated in ways that are structurally identical to those of type VII.

$$11) \quad S_c \rightarrow S'_c \rightarrow C'_c$$

$$12) \quad S_c \rightarrow V_c \rightarrow S'_c \rightarrow C'_c$$

$$13) \quad S_c \rightarrow G_c \rightarrow V_c \rightarrow S'_c \rightarrow C'_c.$$

In the first scenario (11), there is no conscious control. The starting state generates another conscious state, which in turn generates the alternative content. Whatever generated C'_c , it was not the product of any conscious control, so the scenario does not satisfy the possibility of alternative thought. In the second scenario (12), since there is no valuation involved in the generation of V_c , it does not satisfy the MCCR, and therefore, S'_c , being a direct product of V_c , could not have been different. However, since C'_c directly results from S'_c , there is no way it could have been generated as a result of conscious control, and therefore, the scenario cannot satisfy the possibility of alternative thought. The last case (13) is no different. Since neither G_c nor V_c satisfy the MCCR, S'_c could not have been different than it is, and therefore, this is also the case for C'_c , so the scenario cannot satisfy the possibility of alternative thought. Again, it is of no avail to add intermediate goals and valuations in order to make G_c or V_c satisfy the MCCR since this will lead to the same objections as in the analysis of type I scenarios. No scenario will satisfy the possibility of alternative thought, and the process of adding goals and valuations cannot go on indefinitely because of the finite time between the first and last step of the scenario.

Contents of type IV, V and VI do not have to be considered since there is nothing conscious in their generation. Even if there were unconscious valuations involved, no element of conscious control would be involved in the generation of an alternative content, so none of the scenarios can satisfy the possibility of alternative thought.

Content types II and III are essentially variations of type I; the only difference being that they don't directly result from a goal, but rather semi-directly (II) and indirectly (III). Because of this similarity, it is not needed to subject them to a full analysis in order to show that they cannot satisfy the possibility of alternative thought. Any type II or type III scenario has an analogue which is identical to a type I scenario. Therefore, the scenarios that could have led to an alternative content C'_c are identical to those we have considered in the analysis of type I contents. The only difference is that type II contents satisfy one of multiple possible satisfaction conditions, whereas type III contents are 'byproducts' of a goal, which do not satisfy any satisfaction conditions. It is not hard to see that this will not lead to a different conclusion regarding the possibility of alternative thought: for either type, no scenario will be able to satisfy it. For type II scenarios, the fact that there are multiple satisfaction conditions does not alter the fact that a specific content in the scenario does or does not satisfy the MCCR since the definition of the MCCR does not stipulate that there needs to be a *single* satisfaction condition. Consequently, each step in each type II scenario will lead to the same conclusion as its type I counterpart as to whether it could have been generated as the result of conscious control. Therefore, since no type I scenario could satisfy the possibility of alternative thought, this is also the case for type II contents. Type III contents, on the other hand, do not satisfy any satisfaction conditions. Therefore, they cannot satisfy the MCCR. As a result, the number of steps in type III scenarios that could have been generated as a result of conscious control will be necessarily equal or less than those in their type I and II analogues. If these latter scenarios could not satisfy the possibility of alternative thought, then a fortiori, a type III scenario will not be able to either. This concludes the scenario analysis for the ten types of contents.

4. Concluding remarks

It has been shown that no scenario can satisfy the possibility of alternative thought. Since the taxonomy is exhaustive in that it captures every type of content (viewed from its antecedent control structure), it follows that the PAT cannot possibly be satisfied: in no circumstance could a self have entertained a different mental content than it did, as a result of its own conscious control. Saying that we could have thought differently as a result of conscious control is therefore self-refuting. As a corollary, it is easy to show that any physical action suffers the same fate: in no case does it make sense to say that a self could have acted differently than it did, at least not as a result of any conscious control. The reason is simple: if a bodily action

resulted from one or more mental contents, then that action could not have been different than it was, because the contents could not have been different. In other words, any actions resulting from conscious deliberation, planning or decision-making, from beliefs, desires, intentions, etc. could not have been different than they were, since the contents that caused them could not have been different. The same is a fortiori true for actions that only resulted from unconscious processes, since there was no conscious control over them in the first place. Therefore, by showing that the PAT cannot be satisfied, we have also shown that the PAP (principle of alternative possibilities, understood as the possibility to act differently, as a result of one's own conscious control) cannot be satisfied.

The argument I have developed in this paper is new, but it is a variation on an old theme: it capitalizes on the fact that even though we have conscious control over our thoughts and actions, the factors that ultimately determine that control are unconscious and therefore escape our (direct) control.

Acknowledgments

Thanks to Bence Nanay and to the members his research group for their feedback on earlier drafts. Thanks also to two anonymous reviewers whose comments helped to clarify and strengthen the paper. This research was supported by a PhD fellowship fundamental research awarded by the Fund for Scientific Research Flanders (FWO).

References

- Bishop, R. C. (2002) Chaos, indeterminism, and free will, in Kane, R. (ed.) *The Oxford Handbook of Free Will*, pp. 111-124, Oxford: Oxford University Press. doi: [10.1093/oxfordhb/9780195178548.003.0005](https://doi.org/10.1093/oxfordhb/9780195178548.003.0005)
- Copp, D. (1997) Defending the principle of alternate possibilities: Blameworthiness and moral responsibility, *Nous*, 31(4), pp. 441-456. doi: [10.1111/0029-4624.00055](https://doi.org/10.1111/0029-4624.00055)
- Dainton, B. (2008) *The Phenomenal Self*, Oxford: Oxford University Press. doi: [10.1093/acprof:oso/9780199288847.001.0001](https://doi.org/10.1093/acprof:oso/9780199288847.001.0001)
- Gallagher, S. (2000) Philosophical conceptions of the self: Implications for cognitive Science, *Trends in Cognitive Sciences*, 4(1), pp. 14-21. Doi: [10.1016/S1364-6613\(99\)01417-5](https://doi.org/10.1016/S1364-6613(99)01417-5)
- Ginet, C. (1996) In defense of the principle of alternative possibilities: Why I don't find Frankfurt's argument convincing, *Philosophical Perspectives*, 10, pp. 403-417. doi: [10.2307/2216254](https://doi.org/10.2307/2216254)
- Honderich, T. (1993) *How Free Are You?: The Determinism Problem* (Vol. 2). Oxford:

Oxford University Press.

- Kane, R. (1998) *The Significance of Free Will*. Oxford: Oxford University Press. doi: [10.1093/0195126564.001.0001](https://doi.org/10.1093/0195126564.001.0001)
- Koch, C., Massimini, M., Boly, M., & Tononi, G. (2016) Neural correlates of consciousness: Progress and problems, *Nature Reviews Neuroscience*, 17(5), 307. Doi: [10.1038/nrn.2016.22](https://doi.org/10.1038/nrn.2016.22)
- Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1993) Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential), in *Neurophysiology of Consciousness*, Birkhäuser, pp. 249-268. doi: [10.1093/brain/106.3.623](https://doi.org/10.1093/brain/106.3.623)
- Lewis, D. (1979) Counterfactual dependence and time's arrow, *Noûs*, pp. 455-476. Doi: [10.2307/2215339](https://doi.org/10.2307/2215339)
- Mele, A. R., & William, H. (2009) *Effective intentions: The power of conscious will*, Oxford: Oxford University Press on Demand. doi: [10.1093/acprof:oso/9780195384260.001.0001](https://doi.org/10.1093/acprof:oso/9780195384260.001.0001)
- Metzinger, T. (2004) *Being No One: The Self-Model Theory of Subjectivity*, Cambridge, Ma: MIT Press. doi: [10.7551/mitpress/1551.001.0001](https://doi.org/10.7551/mitpress/1551.001.0001)
- Shepherd, J. (2015) Consciousness, free will, and moral responsibility: Taking the folk seriously, *Philosophical Psychology*, 28(7), pp. 929-946. doi: [10.1080/09515089.2014.962018](https://doi.org/10.1080/09515089.2014.962018)
- Strawson, G. (1999) The self and the SESMET, *Journal of Consciousness Studies*, 6, pp. 99-135. doi: [10.1093/acprof:oso/9780198777885.003.0003](https://doi.org/10.1093/acprof:oso/9780198777885.003.0003)
- Vandierendonck, A. (2016) A working memory system with distributed executive control, *Perspectives on Psychological Science*, 11(1), pp. 74-100. doi: [10.1177/1745691615596790](https://doi.org/10.1177/1745691615596790)
- Van Inwagen, P. (1975) The incompatibility of free will and determinism, *Philosophical Studies*, 27(3), pp. 185-199. doi: [10.1007/BF01624156](https://doi.org/10.1007/BF01624156)
- Walter, H. (2009) *Neurophilosophy of Free Will: From Libertarian Illusions to a Concept of Natural Autonomy*, Cambridge, Ma: MIT Press. doi: [10.7551/mitpress/4951.001.0001](https://doi.org/10.7551/mitpress/4951.001.0001)
- Wegner, D. (2002) *The Illusion of Conscious Will*, Cambridge, Ma: MIT Press. doi: [10.7551/mitpress/3650.001.0001](https://doi.org/10.7551/mitpress/3650.001.0001)
- Zahavi, D. (2008) *Subjectivity and Selfhood: Investigating the First-Person Perspective*, Cambridge, Ma: MIT Press. doi: [10.7551/mitpress/6541.001.0001](https://doi.org/10.7551/mitpress/6541.001.0001)