

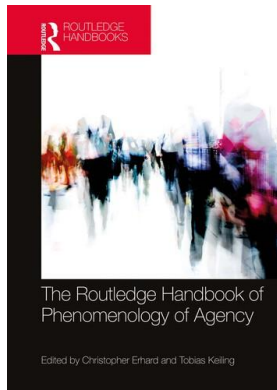
This article was downloaded by: 10.3.97.143

On: 20 Mar 2023

Access details: *subscription number*

Publisher: *Routledge*

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: 5 Howick Place, London SW1P 1WG, UK



The Routledge Handbook of Phenomenology of Agency

Christopher Erhard, Tobias Keiling

On the satisfaction conditions of agentive phenomenology

Publication details

<https://www.routledgehandbooks.com/doi/10.4324/9781315104249-23>

Terry Horgan, Martine Nida-Rümelin

Published online on: 30 Oct 2020

How to cite :- Terry Horgan, Martine Nida-Rümelin. 30 Oct 2020, *On the satisfaction conditions of agentive phenomenology from: The Routledge Handbook of Phenomenology of Agency* Routledge
Accessed on: 20 Mar 2023

<https://www.routledgehandbooks.com/doi/10.4324/9781315104249-23>

PLEASE SCROLL DOWN FOR DOCUMENT

Full terms and conditions of use: <https://www.routledgehandbooks.com/legal-notices/terms>

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

ON THE SATISFACTION CONDITIONS OF AGENTIVE PHENOMENOLOGY

A dialogue

Terry Horgan and Martine Nida-Rümelin

What is it like to experience one's own behavior as the exercising of one's agency? What are the satisfaction conditions of agentive phenomenology (the conditions that must be met in order for experiences of agency to be veridical)? Is such phenomenology typically veridical, or is it instead usually (or even always) illusory? In this dialogue we pursue a debate about these issues that began as a series of oral exchanges between us at recent conferences in Fribourg and in Munich. Horgan has been seeking to articulate and defend a position that (i) fully acknowledges the phenomenal character of agentive experience,¹ (ii) treats such experience as veridical, and (iii) renders agentive phenomenology compatible with familiar, broadly "materialist," tenets in metaphysics and philosophy of mind. Nida-Rümelin has been challenging Horgan; she has maintained that he does not successfully offer a position incorporating each of features (i)–(iii), and she has expressed persistent skepticism about the very possibility of such a package-deal position.

We share some important common ground. We agree that human action is suffused with rich phenomenal character of a specifically agentive kind, and we largely agree about the introspectable features of this phenomenology. We agree that agentive phenomenology is richly *intentional*, in the sense of having content with satisfaction conditions.² And we agree that such phenomenology typically is veridical rather than illusory.

We disagree sharply, however, about the nature of the satisfaction conditions that accrue to agentive experience. Horgan maintains that these conditions are satisfiable consistently with a broadly materialist metaphysics of mind (e.g., Horgan 2007a, 2007b, 2011, 2015), whereas Nida-Rümelin is doubtful (e.g., Nida-Rümelin 2007, 2016, 2018a). We will not resolve this disagreement in what follows.

We do agree, however, that the ensuing debate nonetheless manifests philosophical progress: viz., Horgan refines and improves his position through the course of his successive installments, in ways that are directly prompted by Nida-Rümelin's successive formulations and reformulations of her fundamental objection to Horgan's project.

We expect our debate to continue beyond what is said here. And we sincerely hope that others in philosophy will recognize the importance of the issues we address and will join the ongoing discussion.

Terry Horgan: proposal I

In this section I will first describe what I take to be some salient aspects of ordinary first-person agentive phenomenology, and I will then set forth some of my key claims, in prior writings and conference presentations, about the satisfaction conditions of such phenomenology.

What is it like, phenomenologically, to apparently perform an action? (Here and throughout, I deploy the modifiers ‘apparent’ and ‘apparently’ in order to maintain neutrality about whether or not such agentive phenomenology is veridical.) I will describe *seven* introspectively discernable aspects of such experiences.

First is the aspect of *self as source*. One’s apparent actions are not experienced passively, e.g., as bodily or mental events that are merely happening to oneself, or as the outcome of passively experienced state-causal processes one undergoes (e.g., a process consisting of an occurrent desire-belief combination state-causing an intention-formation or a bodily motion).³ Rather, one experiences one’s apparent actions as emanating from *oneself* as their source—as opposed to experiencing them, passively, as the end products of an experienced state-causal process. (When a process is experienced as transpiring state-causally, however, the phenomenology is as-of the process evolving *by itself*, without any agentive intervention either by oneself or by others.)

Second is the phenomenological aspect of *optionality*. One’s apparent actions are experienced as being under one’s voluntary control, and thus as being such that in the circumstances one can/could act otherwise. In advance of apparently performing an action that one is considering performing, one experiences the future as being “open” to oneself, with respect to whether or not one will perform it. And once an apparent act has occurred (either spontaneously or upon prior deliberation), one experiences both its potential performance and its potential non-performance as having been open to oneself.

Third is the phenomenological aspect of *purposiveness*. Normally, one experiences one’s actions as being done *on purpose*, and as being done for *specific* purposes both coarse-grained and fine-grained. For example, the coarse-grained purpose for which one enters one’s office in the morning involves further, more fine-grained aspects of purposiveness, such as pulling out one’s keys as one approaches one’s office, grasping one’s office key, extending it just so into the lock and then twisting it just so, etc.

Fourth is what Matjaž Potrč and I call *chromatic illumination* (cf. Horgan and Potrč 2010). Often, many of the purposes informing one’s actions affect the character of one’s agentive experience by being *implicitly appreciated* consciously without actually being *represented* in consciousness. These purposes are phenomenologically operative as reasons for one’s actions, coloring the specific character of one’s agentive experience without being overtly “before the mind.”

Fifth, the practical reasons (i.e., purposes) for which one (apparently) acts, and which figure in the character of one’s agentive experience, normally exhibit a means-ends teleological structure. In the familiar simplest case, one has a desire D for a particular outcome, and one has a belief B that performing action A is an available best-means to secure that outcome. This belief-desire combination constitutes a practical reason for performing action A. One acts in light of, and in response to, such a practical reason toward achieving an end. (Some aspects of the belief-desire combination might be implicitly appreciated experientially via chromatic illumination, without being overtly present in consciousness.)

Sixth, appreciation of a practical reason for performing an action A normally exerts a *motivational pull* toward A-ing. The appreciated reason is experienced as *authoritatively favoring* A-ing. In the case of ordinary practical reasons, this authoritativeness has a “hypothetical

imperative” character rather than a “categorical imperative” character, because its authority is desire-dependent. Such desire-dependence is intimately connected to the experienced voluntariness of one’s actions—the aspect of optionality. One’s experience is as-of being free to act on one’s desire or to refrain from doing so.

Finally, when one (apparently) acts on the basis of a practical reason, normally one experiences oneself as so acting *because* of that reason, *qua* reason. (Again, this can be implicit, via chromatic illumination.) This “rational because” aspect of agentive experience is *not* an aspect as-of passively experienced state causation—say, as-of one’s belief-state and desire-state jointly state-causing one’s body to move. On the contrary, the experience of performing an action *for* a reason has the phenomenological aspect as-of self-sourcehood—and, moreover, as-of *purposive* self-sourcehood, and, indeed, as-of self-sourcehood involving a *specific* purpose. That experienced purpose—the consideration that one experiences as the reason *for which* one now acts in the particular manner one does—is the “rational because” aspect. Phenomenologically, such experience resides squarely, indeed paradigmatically, in the “space of reasons,” rather than residing in the “space of causes,” i.e., the space of phenomena experienced passively as state-causal transactions.

Agentive phenomenology, in virtue of the several aspects I have been describing, is rife with *intentionality*—in the sense of ‘intentionality’ involving representational content. (It is also rife with intentionality in the sense of purposiveness, which is a part of its intentionality in the former sense.) The following two important, deeply intertwined, philosophical questions therefore arise. First, what are the satisfaction conditions of ordinary agentive phenomenology? Second, are these satisfaction conditions normally (or ever) actually met when human beings undergo agentive experience? I will now briefly rehearse some of my views pertaining to these two questions, which elsewhere I have articulated at greater length and have defended (e.g., Horgan 2007b, 2015).

First, it is introspectively obvious that agentive phenomenology is **not** as-of one’s (apparent) action being a state-causal process. It is not as-of one’s bodily motion or one’s intention-formation being state-causally generated by a mental state or a combination of mental states in oneself (e.g., a belief-desire combination). Nor is it as-of a physical state-causal process, *qua* physical (e.g., a state-causal process consisting of certain brain-states, *qua* brain states, state-causing one’s body to move in a specific manner).

Second, it is crucially important to distinguish between (1) one’s agentive phenomenology being **not** being **as-of** a state-causal process, and (2) one’s agentive phenomenology being **as-of not** a state-causal process. Feature 1 is a way that one’s apparent action is **not** experienced to be, and nothing more. However, if feature 2 were indeed an aspect of one’s agentive phenomenology then this would be a way that one’s apparent action **is** experienced to be—albeit a negative way-of-being, viz., being **not a state-causal process**.

Third, although there is presumably a determinate fact of the matter about whether or not ordinary agentive phenomenology has feature 2, introspection alone cannot reliably ascertain whether it does. Thus, those philosophers who think that they reliably detect the presence of feature 2 in their agentive phenomenology are mistaken; they are conflating feature 2 with feature 1. (This is not to say that introspection alone can reliably detect the *absence* of feature 2 in ordinary agentive phenomenology; on the contrary, introspection is not powerful enough by itself to yield a reliable verdict either way.)

Fourth, there is excellent reason to believe, based on well-documented scientific evidence, both (i) that all human behaviors that are experienced as actions are in fact state-causally generated by neuro-physical causal processes in the brain and central nervous system, and (ii) that all neuro-physical states and events in the brain and central nervous system are

themselves physically state-causally generated. (This is the hypothesis of “physical causal completeness,” specifically with respect to humans.)

Fifth, there is also strong reason to believe that the satisfaction conditions of agentive phenomenology include the following condition of *psychological state causation*:

PSC: In order for an instance of human behavior that is experienced as one’s own action to really be an action, this behavior must be state-causally produced by a psychological state that “rationalizes” it, in the sense of rendering it rationally appropriate—e.g., a state comprising both a desire for a certain outcome and a belief that such behavior would bring about (or probably would bring about) that outcome.

This is so even though, as I have already emphasized, experiences of agentively instigating one’s behavior are **not** experiences **as-of** such a psychological state-causal process. The PSC harks back, of course, to Donald Davidson’s hugely influential 1963 paper “Actions, Reasons, and Causes”—even though Davidson himself did not acknowledge agentive phenomenology at all, let alone frame his account of agency as being about the satisfaction conditions of such phenomenology.

Sixth, concerning the PSC itself, the state-causal efficacy of psychological states is compatible with the hypothesis of physical causal completeness—even though, because of the multiple realizability of psychological state types by neuro-physical state types, numerous psychological state types are distinct from the neuro-physical state types that realize them on specific occasions. (In response to Jaegwon Kim’s infamous problem of “causal exclusion,” I advocate a version of contextualism about causation and causal explanation; cf. Horgan 1989 1998, 2001; Horgan et al. 2009.)

Seventh, normally the satisfaction conditions of agentive phenomenology are indeed satisfied on occasions when people are undergoing such phenomenology. That is, normally this phenomenology is veridical, and hence normally people really do exercise genuine free agency of the kind they experience themselves to be exercising.

Finally, eighth (and as a consequence of the above), the contention that people normally exercise free agency of the kind they experience themselves to be exercising is compatible with the various other claims I embrace; in particular, it is compatible with the claim that the exercise of free agency is a psychological state-causal process (despite not being experienced as one), and it is compatible with the hypothesis of physical causal completeness. In addition, it is also compatible with state-causal determinism. (Although I regard it a seriously open question whether or not state-causal determinism is true at the level of fundamental physics, I believe that there is strong scientific evidence that neurobiological processes in humans and other organisms are *virtually* deterministic—with any quantum-level indeterminacies (if such there be) getting “washed out,” rather than getting amplified in such a way as to matter at the neurobiological level of temporal evolution.)

Martine Nida-Rümelin: rejoinder I

Experiencing oneself as active

Horgan distinguishes six phenomenally manifest aspects present in all or many cases of agentive experience: the aspect of self as source, of optionality, of purposiveness, of chromatic illumination, of apparently acting for certain practical reasons, of motivational pull, and finally the “rational-becauseal” aspect of agentive experience. I will focus on the first of these, on what Horgan calls the aspect of self as source.

If one wishes to discuss whether an aspect of human agentive experience is veridical without talking past each other, one must first establish a situation of shared reference: one needs to make sure that one is talking about the very same aspect of human experience. Based on numerous conversations I am confident that Horgan and I are in fact in such a situation of shared reference concerning the first aspect he mentions. What Horgan calls ‘the aspect self as source’ is what I have in mind when I use the locution “experiencing oneself as active” (see Nida-Rümelin 2007, 2016, 2018a). We are both confident that this specific aspect is present in all humans when they are acting. We agree on a stronger thesis as well: no bodily movement should count as a case of acting if that specific experiential aspect is absent.

I would like to make sure that we are in a situation of shared reference concerning that specific experiential aspect with the reader as well. In order to do so, let me first cite parts of a well-known passage occurring in several of Horgan’s previous writings about the topic. I would like to invite the reader to use that passage in order to individuate the specific aspect we are talking about in his or her own agentive experience:

What is behaving like phenomenologically, in cases where you experience your own behavior as action? Suppose that you deliberately perform an action—say, holding up your right hand and closing your fingers into a fist. [. . .] In order to help bring into focus this specifically actional phenomenological dimension of the experience, it will be helpful to approach it a negative/contrastive way, via some observations about what the experience is *not* like. For example, it is certainly not like this: first experiencing an occurrent wish for your right hand to rise and your fingers to move into clenched position, and then passively experiencing your hand and fingers moving in just that way. [. . .] It would be very strange indeed, and very alien. Nor is the actional phenomenological character of the experience like this: first experiencing an occurrent wish for your right hand to rise and your fingers to move into clenched position, and then passively experiencing a causal process consisting of this wish’s causing your hand to rise and your fingers to move into clenched position. [. . .] it seems patently clear that one does not normally experience one’s own actions in that way—as passively noticed, or passively introspected, causal processes consisting in the causal generation of bodily motion by occurrent mental states. That too would be a strange and alienating sort of experience.

How, then, should one characterize the actional phenomenal dimension of the act of raising one’s hand and clenching one’s fingers [. . .]? Well, it is the what-it’s-like of *self as source* of the motion. You experience your arm, hand, and fingers as being moved *by you yourself*—rather than experiencing their motion either as fortuitously moving just as you want them to move, or passively experiencing them as being caused by your own mental states. You experience the bodily motion as caused by *yourself*.

(Horgan et al. 2003: 327–329)

Here are various ways to describe the relevant aspect of agentive experience, two of them taken from the passage above:

- You experience the relevant part of your body as moved by yourself.
- You experience the bodily motion as caused by yourself.
- You experience your own movement as brought about by yourself.
- You experience yourself as active in what you do.
- You experience yourself as the causal origin of your movement.

These descriptions of what appears to be the case when one has an experience with that aspect should help attract the reader's attention to that specific phenomenally manifest aspect. They are used to fix reference to a specific part of how things appear to be (of how one's own behavior appears to be generated) while one is executing an action. One cannot refer to that specific phenomenal aspect without saying something about what appears to be the case in such experiences. Therefore, when one tries to establish a situation of shared reference, one seems to be saying something about the veridicality conditions of the kind of experience one has in mind. (Let us work, for the moment, with the following notion of veridicality conditions: the veridicality conditions of a given type of experience E are those conditions, which render, if satisfied, experiences of type E veridical where an experience is veridical if and only if it presents or represents the world as it is.) After all, on a natural understanding and in a first approximation (we will both say more on that issue below), the veridicality conditions of a given type of experience E are simply the conditions which *appear* to be fulfilled when one has an experience of type E. Therefore, each of the five ways chosen above (with the intention to fix reference to an aspect manifest in agentive experience) involves already an assumption about the veridicality conditions of experiences with that aspect.

It is however important to note that one can nonetheless remain neutral (at this point) about whether any of these assumptions is correct. The various descriptions above can successfully serve their communicative purpose (to establish a situation of shared reference to a specific experiential aspect we are all familiar with) even if the corresponding claims about veridicality conditions were, strictly speaking, all false. In what follows, I will call experiences of the relevant kind *experiences of oneself as being active in one's behavior*, or **experiences of the type OBA** for short and, analogously, I will call the phenomenal aspect which characterizes these experiences the **OBA-aspect**.

For various reasons, the OBA-aspect is of particular interest in the context of our discussion. First, it is phenomenologically fundamental in the sense that any other aspect of agentive experience includes it. For instance, experiencing oneself as being able to act otherwise is to experience the capacity to *actively bring about* a different movement. A parallel claim applies to all the other aspects listed above. Second, it is of all six aspects listed by Horgan the one which is metaphysically most relevant. To have an experience with that aspect is to experience oneself as active in a given movement, as bringing that movement about rather than something one passively undergoes. The line between actions and non-actions does not coincide with the line between active changes (doings) and passive changes (non-doings) since all actions are doings in that sense but not vice versa (see, e.g., Nida-Rümelin 2007). For reasons I develop in some detail in other papers, the metaphysically more significant line is the one between doings and non-doings (see, e.g., Nida-Rümelin, 2018a). Third, if we wonder about whether agentive experience is broadly veridical or rather illusionary, we should focus on the aspect of self as source (on experiencing oneself as active). If Horgan and the vast majority of philosophers and scientists today are right in assuming that human beings are what we may call biological robots (that is, if their behavior is causally determined by preceding conditions plus microphysical laws), then this particular aspect of agentive experience poses a serious problem. It is plausible to think that your experience that *you* bring about a given movement cannot be veridical if your movement is, in fact, the result of microphysically determined processes caused by preceding conditions prevailing in your brain. Horgan disagrees with that claim and I will defend it against his arguments. However, we can easily agree on the following. If the standard scientific view of human beings forces us to accept that agentive experience is non-veridical, then the aspect of self as source is the most plausible suspect. If agentive experience is in large part non-veridical then this is so in virtue

of that specific phenomenally fundamental aspect of agentive experience. Experiences with any of the other aspects are then non-veridical as well but only because they *include* the aspect of self as source. (This claim is however put into doubt by Baierlé 2016 who argues that the aspect of optionality is an additional ‘suspect’.)

In the passage cited above Horgan points out that it would be alienating to experience one’s own movements as caused by our own occurrent mental states such as wishes and desires. Let us call that type of experience (or that aspect of potential experiences) **experiences of type MC** (or aspect MC—“MC” for “mental causation”). According to Horgan, we never have MC-experiences with respect to our own behavior involved in action. Human agentive experience does not have the aspect MC. Yet, human behavior involved in acting actually *is* generated—according to that view—precisely in the way in which it would *appear* to be generated in MC-experiences. In short, MC-experiences would be alienating and yet veridical if we had them with respect to our own behavior in acting. In what follows, I will argue that such a result should not satisfy a philosopher who wishes to defend agentive experience against the charge of being illusionary.

A thought experiment: Alexa’s transformation

One morning shortly after waking up, Alexa realizes with surprise and with horror that something has radically changed. Until the evening before, Alexa was just like all of us. She experienced her own movements involved in her actions as brought about by herself. Her agentive experience had the OBA-aspect. Now that aspect is absent. Instead, Alexa is now under the impression that her own wishes and beliefs cause her movements. As every morning, she stands up, goes to the kitchen, and prepares a cappuccino. In the past, she always experienced herself as bringing all those movements about. Not so now. Her wish to have a cappuccino and her beliefs about how she can realize it appear to cause her legs to move in the way they do and her hands and fingers to move in the way required. Alexa is shocked. She experiences the *absence* of the familiar experience of being herself the source of what happens as deeply disturbing.

We can summarize Alexa’s transformation saying that up to the crucial moment (a) Alexa’s agentive experience has the OBA-aspect and does not have the aspect MC and that (b) after that moment Alexa’s agentive experience has the aspect MC but does not have the OBA-aspect.

According to the view Horgan presents in the first part of our paper, Alexa’s experience is veridical before *and* after that change. Her pre-change and post-change agentive experiences are both veridical in virtue of the fact that certain mental events cause her bodily behavior.

Before I use the example to formulate my objections, let us see if Horgan can object to anything in the above description of Alexa’s case. It does not seem so. Note that I did not claim that Alexa has the experience of *not* being actively involved or of *not* causing her own movements. I only state that, contrary to her previous experience, Alexa does *not* have the impression of bringing her movements about. One may safely assume that the mere *absence* of the OBA-aspect would be deeply disturbing. It would be alienating to realize that one’s body is moving as if one were acting without thereby having any impression of being the one who actively originates the relevant movements.

To see the problem Alexa poses for the view Horgan describes above, one will have to keep in mind that its main motivation is anti-illusionist. Horgan finds the idea unacceptable that agentive experience is generally and fundamentally illusionary. He intends to *avoid* the illusionist claim that agentive experience misrepresents the genesis of our own action and

yet to preserve both the phenomenological insight that we experience ourselves as actively bringing our movements about in every action and the theoretical claim that actions are caused by occurrent mental states. The argument I would like to develop based on Alexa's case is supposed to show that these claims are in serious tension. Horgan's proposal avoids the illusionist claim only in a superficial way, which cannot satisfy the anti-illusionist with respect to agentive experience.

First problem raised by Alexa's case: the appearance of change

Imagine how it is like for Alexa to become aware of the radical change in her agentive experience. Until yesterday, she had the impression that *she herself* brought her movements about in an active manner. Now she has no such impression anymore. Rather, her occurrent mental states appear to cause the way she moves. Vividly imagining Alexa's situation, one is pressed to accept a further claim. Not only does Alexa realize a radical change in the way her behavior appears to be generated. Rather, the way her behavior is generated will appear to her to have undergone a radical change. With the change in appearance, things will appear to have changed. However, according to Horgan's proposal, *that* impression is an illusion.

The view implies that no change has taken place in the way Alexa's movements in her actions come about. The very same occurrent mental states cause her movements before and after the radical change in her agentive phenomenology. This result tells us something not only about the imagined case involving Alexa. More importantly, it tells us something about ourselves. We all would suffer an illusion of change if we underwent an analogous modification in agentive experience.

After the change, Alexa has no experience with the OBA-aspect. She does not experience herself as active in her behavior. One must not jump to the conclusion that Alexa then experiences her movements as *not* being brought about by herself. Horgan stresses in various places and rightly so that we must not conflate the absence of an experience that *p* is the case with an experience of *p* not being the case. However, one may now use Alexa's illusion of change to argue for the stronger claim (that after the change Alexa experiences herself as *not* being actively involved in her movement). In general, if the change from experiencing something as having property *F* to experiencing that same thing as having property *G* necessarily involves an experience of change (the impression that the object itself has changed from having *F* to having *G*), then one may conclude that experiencing an object as having *F* includes experiencing it as not having *G*. One could defend that principle by application to concrete cases. One would have to show that it leads to intuitively correct results when applied to various examples. I am confident that this can be done in a convincing manner. However, for reasons of space, I will leave it to the reader to submit that principle to intuitive tests.

Second problem raised by Alexa's case: no genuine experiential access to the genesis of one's own behavior

The second problem I would like to raise against the version of Horgan's view presently under consideration is perhaps more serious; it also goes deeper and is a little more complicated to explain. To put it briefly, the view implies that human beings have no genuine experiential access to the way their own behavior actually comes about and that, even worse, having such genuine experiential access is impossible for humans since it is incompatible for them with acting.

According to the relevant version of the view, the veridicality conditions of OBA-experiences and MC-experiences are the same. Both kinds of experiences are veridical

(when they are) in virtue of certain occurrent mental states causing the relevant behavior. So far, these two kinds of experiences are on a par. However, there is an important disanalogy between these two kinds of experiences. Having an OBA-experience does not reveal to the subject how her own behavior comes about while MC-experiences do. Having an OBA-experience does *not* involve the impression that occurrent mental states cause one's own behavior. In fact, when a person acts, causation by occurrent mental states *is* the real genesis of the relevant piece of behavior (according to the view here at issue). It follows that OBA-experiences, even if they are veridical according to the view, do *not* present to the experiencing subject how her own behavior comes about. Even though her experience of how it comes about is veridical, it does *not* give genuine experiential access to what is in fact going on. This is so since, in an intuitively clear sense, OBA-experiences are not transparent with respect to their own veridicality condition. In having an OBA-experience one is not aware of the fact that it is mental causation which renders these experiences veridical. Nor could any reflection on how things appear in that experience make it obvious that this is so. To be told that one's experience of being active is veridical just in case certain occurrent mental states cause one's own behavior would come as a surprise. Contrary to this, one is aware of what renders one's experience veridical in the case of MC-experiences.

Putting aside issues about the metaphysical nature of causation, one may say that having an MC-experience comes with a clear phenomenally manifest awareness of how one's own behavior must be generated for the experience to be veridical. A person who has an MC-experience is thereby phenomenally aware of the fact that the experience presents the genesis of her own behavior correctly if and only if occurrent mental states cause her own behavior. One may perhaps object that to say so is to over-intellectualize phenomenal content. I do not think it is, but for the purposes at hand one may weaken the thesis: in the case of MC-experiences, one can easily discover by a little phenomenological reflection that the experience one is having presents the genesis of one's behavior correctly if it is in fact caused by the relevant occurrent mental states. (To make the present point precise, one needs to develop an understanding of what it is for a veridicality condition to 'show up' at the phenomenal level. See for a proposed account Nida-Rümelin 2018b.)

Let me summarize the result as follows. According to the view proposed by Horgan in the first part of the present paper, OBA-experiences do not reveal the genesis of our own behavior when we act while MC-experiences, if only we could have them, *would* reveal how our behavior in acting comes about. This result is highly problematic for a philosopher who wishes to avoid illusionism about agentive experience. To see this one must fully appreciate a few further elements of the view proposed. Horgan acknowledges that it would be alienating to have MC-experiences. *It follows that it would be alienating to be experientially aware of the way our own behavior actually comes about.* It would be alienating because having MC-experiences excludes that one experiences oneself as actively involved in what one does.

I take it that a theory with such a consequence does not genuinely avoid illusionism. It avoids illusionism in its letter (agentive experience is veridical) but not in its spirit (only alienating agentive experience could reveal the truth about the genesis of our behavior in acting). The situation is even worse. Since the view implies that acting necessarily involves OBA-experiences and that MC-experiences cannot co-occur with OBA-experiences in human beings, it follows, furthermore, that experiential access to the way our behavior actually comes about when we act is even incompatible (at least for humans) with acting. I conclude that the first proposal presented by Horgan in the first part of this paper, contrary to his intention, does not do duty to the anti-illusionist intuition.

Terry Horgan: proposal II

Nida-Rümelin characterizes my views about the phenomenology of agentive experience, and about the satisfaction conditions of such phenomenology, in a way that accurately reflects both my prior published work on this topic and my presentations about it at conference venues at which we were both present. In the course of thinking about her challenge, however, I have found myself making refinements to my position that seem to me both independently plausible and helpful in seeking to respond to her. Let me now set forth these refinements, as a prelude to my subsequent replies to her multi-faceted challenge. (I label these refinements “R1,” “R2,” etc. to facilitate subsequent cross-references.)

Refinement R1: In previous presentations at conferences I often said that a key satisfaction condition of agentive phenomenology is the requirement that the bodily motion which one experiences as one’s action is state-causally generated by a mental state, or a combination of mental states, that “rationalizes” such an action—for instance, a desire for a certain outcome, together with a belief that the pertinent action will produce (or probably will produce) the desired outcome. I now maintain that a more fundamental satisfaction condition involving mental state causation is this: one’s experience of agentively initiating the action must *itself* state-cause the pertinent bodily motion. Hereinafter, I will refer to such an action-initiation experience as a *willing*. (When a specific belief-desire combination constitutes the reason for which one acts, it too is state-causally operative, but indirectly: it state-causally generates one’s willing, which itself is then the proximal psychological state-cause of the pertinent bodily motion.)

Refinement R2: My previous discussions gave the impression that I think a creature could exhibit full-fledged agency even without undergoing agentive phenomenology—for instance, by undergoing a belief-desire combination which state-causes a bodily motion that it “rationalizes,” without also undergoing an experience as-of *willing* the action. In fact, however, I maintain that a creature cannot be a full-fledged agent, and cannot perform full-fledged actions, without undergoing the pertinent agentive phenomenology. Thus, the *presence* of agentive phenomenology is among the satisfaction conditions of such phenomenology: you can’t exercise full-fledged agency without experiencing yourself as doing so. (I take it that this second refinement is entailed by the first: willings must be state-causally operative—and hence present—in exercises of genuine agency; and willings are experiences as-of agentive self-sourcehood.)

Refinement R3: I now distinguish between two kinds of satisfaction conditions. On one hand are *content-based* satisfaction conditions, which are constitutive (or partly constitutive) of the intentionality of a given, intentional, psychological state type. On the other hand are *implementational* satisfaction conditions; these are not constitutive of the psychological state-type’s intentionality, but nonetheless must be met in order for that state type to be realized or implemented within a given kind of creature (e.g., by a human being). In humans, I claim, mental state types must always be physically realized by certain brain states; being thus realized is therefore an implementational satisfaction condition of mental state types (for humans), although not a content-based satisfaction condition.

Refinement R4: Now, as I write this installment of the dialogue, I maintain that the requirement that one’s willing-experience state-causes one’s pertinent bodily motion is only an *implementational* satisfaction condition—and is not a content-based satisfaction condition—of one’s agentive phenomenology. (However, I leave open the possibility that this implementation condition is applicable to all possible creatures with mentality, and that this contention perhaps can even be established *a priori*.)

Refinement R5: I now believe that the kinds of experiences as-of state-causal processes that humans are capable of underdoing are always—and can only be—experiences as-of *non-agentive* state-causal processes. That is, such experiences are always—and can only be—as of state-causal processes that evolve *by themselves*, without any agentive intervention either by oneself or by other creatures who one perceives as exercising agency. Since this “by itself” aspect is constitutive of such experience, the experiential content must always be as-of *non-agentive* state causation. It is a vexed question whether this so because state causation *simpliciter* is inherently non-agentive, or instead is so even though—as I myself maintain—exercises of agency are themselves a species of state causation even though they cannot be experienced as such. (I return to this question in my third installment below.)

I turn now to the two challenges that MNR raises for my position. The first concerns the appearance of change that her thought-experimental experiencer, Alexa, would undergo. The heart of the objection is this:

Until yesterday, she [Alexa] had the impression that *she herself* brought her movements about in an active manner. Now she has no such impression anymore. . . . Rather, the way her behavior is generated will appear to her to have undergone a radical change. With the change in appearance, things will appear to have changed. However, according to Horgan’s proposal, *that* impression is an illusion.

The view implies that no change has taken place in the way Alexa’s movements in her actions come about. The very same occurrent mental states cause her movements before and after the radical change in her agentive phenomenology. This result tells us something not only about the imagined case involving Alexa. More importantly, it tells us something about ourselves. We all would suffer an illusion of change if we underwent an analogous modification in agentive experience.

My response is to appeal to my refinements R1 and R2. In fact, I contend, the genesis of Alexa’s movements *has* changed, according to my position—and hence I am not committed to the counterintuitive claim that the appearance of change is an illusion. The state-causal etiology has changed because the kind of conscious mental state which, in a case of genuine agency, is the proximal mental state-cause of the pertinent bodily movements—viz., a conscious willing of a specific action—is totally absent in Alexa’s case. Now her bodily movements are proximately state-caused by wishes, or by her belief-desire combination states—whereas before her wishes and beliefs were etiologically distal, giving rise to the willings that were the proximal state-causes of her motions back when she was still undergoing normal agentive experience. (She no longer experiences agentive phenomenology, and indeed she no longer exercises agency.)

The second challenge posed by Nida-Rümelin is one that she herself considers more serious, as do I. She formulates it this way:

According to the view proposed by Horgan in the first part of the present paper, OBA-experiences do not reveal the genesis of our own behavior when we act while MC-experiences, if only we could have them, *would* reveal how our behavior in acting comes about. This result is highly problematic for a philosopher who wishes to avoid illusionism about agentive experience. To see this one must fully appreciate a few further elements of the view proposed. Horgan acknowledges that it would be alienating to have MC-experiences. *It follows that it would be alienating to be experientially aware of the way our own behavior actually comes about.* It would be alienating because having MC-experiences

excludes that one experiences oneself as actively involved in what one does. I take it that a theory with such a consequence does not genuinely avoid illusionism. It avoids illusionism in its letter (agentive experience is veridical) but not in its spirit (only alienating agentive experience could reveal the truth about the genesis of our behavior in acting). The situation is even worse. Since the view implies that acting necessarily involves OBA-experiences and the MC-experiences cannot co-occur with OBA-experiences in human beings, it follows, furthermore, that experiential access to the way our behavior actually comes about when we act is even incompatible (at least for humans) with acting. I conclude that the first proposal presented by Horgan in the first part of this paper, contrary to Horgan's intention, does not do duty to the anti-illusionist intuition.

I take it that the heart of this objection—the source of her contention that my position, as she says, “does not do duty to the anti-illusionist intuition”—is the following *first putative consequence* (for short, “PC.1”) of the position:

PC.1: If we were to experience our own (apparent) actions as mentally state-caused, then (i) we thereby would be aware of their real nature and (ii) we would find the experience deeply alienating.

The transition from this to her claim that my position is too close to the illusionist position is via this *key contention* (“KC”):

KC: If experience of our own (apparent) actions as mentally state-caused would be *both* veridical and yet deeply alienating, then my position avoids the illusionist position only in a superficial sense.

KC, together with PC.1, entails the following *second putative consequence* of my position:

PC.2: My position avoids the illusionist position only in a superficial sense.

One way I might try to push back against this objection would be to embrace PC.1, but repudiate PC.2 by denying KC. But that is not the tack I wish to take. Rather, I concede KC, at least for argument's sake (although I also find KC quite plausible). That is, I concede that if my position really did entail PC.1, then the position thereby would be so close to illusionism that it would fail to honor satisfactorily the contention that agentive experience is veridical.

Instead I propose to argue that the version of my position that I now espouse, incorporating the five refinements R1–R5 set out above, does *not* have the untoward consequence PC.1. I will consider three different construals of what it would be like to experience our own actions as caused by mental states of ourselves, and I will take issue with PC.1 under each of these construals. (I will explicitly invoke the various refinements R1–R5, when applicable.)

First construal (which presumably is what Nida-Rümelin actually has in mind): One's experiences are like Alexa's experiences of the originations of her bodily motions, after her transformation.

Second construal: One's experiences are simultaneously (i) as-of willing one's action in the manner that includes the phenomenology of self-sourcehood and the phenomenology of optionality, and (ii) as-of this very willing state-causally generating the pertinent bodily

motion. Moreover, aspect (ii) is an experience as-of a purely passive process in which agency is not involved.

Third construal: One's experiences are simultaneously (i) as-of willing one's action in the manner that includes the phenomenology of self-sourcehood and the phenomenology of optionality, and (ii) as-of this very willing state-causally generating the pertinent bodily motion. Moreover, aspect (ii) is itself neutral as to whether or not agency is involved in this state-causal process.

Under the first construal, the envisioned experiences would *not* “reveal to us how our behavior in acting comes about”—for two important reasons. First, now one would entirely lack the phenomenologically agentive *willing-states* which, in cases of genuine action (given R1 and R2), are the proximal psychological state-causes of one's bodily motions. Instead, the proximal psychological state-causes of one's bodily motions would be different—viz., states like wishes, belief-desire combinations, and the like. (In cases of genuine action, however, such states are only distal psychological causes of one's bodily motions, with willing-states being the psychologically proximal ones). Second, now one's bodily motions would be experienced as *non-agentive*, since (given R5) humans can only experience processes as state-causal at all by experiencing them as non-agentively state-causal. So although Alexa-type experiences would be veridical vis-à-vis the bodily motions that they themselves state-causally generate *in herself after her transformation*, in the respects mentioned they would reveal ways that those bodily motions are now being brought about in her case that are *different* from how ordinary human actions are really brought about. Thus, PC.1 is false under this construal, because the experience in question is not veridical vis-à-vis the real nature of actual human actions.

Under the second construal (and given R5), aspect (ii) of the pertinent kind of experience would be as-of a *non-agentive* state-causal process. Thus, one's overall experience (if one could undergo such an experience at all) would be as-of an impossible situation, experienced *as* impossible—viz., an experience as-of a phenomenon that is both agentive by virtue of feature (i), and non-agentive by virtue of feature (ii). Now, it is fairly plausible that such a putative experience, as-of such an impossible state of affairs, is *itself* impossible—at least for humans. But perhaps it is actually possible for humans, at least in something like the manner in which it is possible to experience apparent “impossible objects” like a visually presented Penrose triangle—in that case, by an overall diachronic experience consisting largely of successive, focally presented and possible, object-parts that collectively cannot be parts of a single possible entire object. (Many famous Escher drawings feature Penrose triangles; see also the cover of my 2017 collection *Essays on Paradoxes*.) Either way, however, such a putative experience certainly would not “reveal to us how actions are really brought about,” since it would be an experience as-of a bodily motion being brought about in a manner in which it *could not possibly* be brought about—viz., both agentively and non-agentively. Thus, PC.1 is false under this second construal, because the experience in question—if possible at all (and it may not be)—is non-veridical not only vis-à-vis the real nature of actual human actions, but also vis-à-vis the real nature of one's behaviors in the envisioned counterfactual scenario itself.

Under the third construal, one's experiences would, I admit, be *more* revealing concerning “how our behavior in acting comes about” than are actual human agentive experiences. However (and given R1, R2, and R5), these are not experiences that creatures like us humans are capable of undergoing. Given R5, this is because the only kinds of experiences as-of state causation that humans are capable of are experiences as-of *non-agentive* state

causation. *Of course*, it would be alienating for us humans to experience our bodily motions as proximally mentally generated state-causally, because that would be to experience them as *non-agentively* generated (or as generated, impossibly, both agentively and non-agentively). But the super-humans now being envisioned, unlike ourselves, are capable of experiences as-of state causation that are neutral as to whether or not the experienced process is also agentive. Indeed, they are capable of experiencing their actions simultaneously both (i) as emanating from willing-states with the inherent phenomenological aspects of self-sourcehood and optionality, and (ii) as bodily motions that are proximally state-causally generated, mentally, by these very willing-states. For them, I contend, there is no reason for such dual-aspect experiences to be alienating. But we humans are not capable of such experience. Thus, PC.1 is false under this third construal, because the experience in question—one that humans are incapable of undergoing, or of even imagining what it would be like to undergo—would not, I claim, be alienating for creatures capable of undergoing it.

The upshot is that according to my position PC.1 is false under each way of construing a counterfactual scenario in which one's bodily motions are experienced as being proximally state-causally generated by certain mental states. With this as the dialectical situation, I deny that my position entails PC.1. I also deny that I am committed to PC.2—the claim that my view is only superficially different from illusionism about agency—even though I do not contest Nida-Rümelin's claim KC. This is my reply to the first part of her principal objection.

Two residual issues still remain, however, each of which involves a potential challenge to my position that is still broadly in the spirit of Nida-Rümelin's own worries.

First, one might think that it is somehow philosophically problematic that according to my view, there are satisfaction conditions of agentive experience that are not reliably ascertainable from within agentive experience itself—i.e., are not phenomenally manifest. I contend, however, that there are several reasons why this consequence of my position is not problematic. For one thing (and given R3 and R4), the requirement that one's willing-the-action be the proximal psychological state-cause of one's bodily motion is only an *implementational* satisfaction condition of one's agentive phenomenology, rather than a content-based satisfaction condition; and there is no evident reason to expect agentive experience itself to reveal either the nature of its merely implementational satisfaction conditions or the fact that they are satisfied on a given occasion of experienced action. (After all, agentive experience certainly does not reveal its own *neuro-physical* implementational satisfaction conditions, or the fact that they are satisfied on a given occasion of experienced action.) *Second*, human powers of introspection evidently are somewhat limited anyway, even vis-à-vis the specific *content-based* satisfaction conditions of agentive experience; in particular, they are limited with respect to the satisfaction conditions that accrue to the optionality aspect of such experience, the aspect as-of “could/can do otherwise, given my actual circumstances and myself as I actually am.” (This second point is one I have defended at some length elsewhere; cf. Horgan 2011, 2012, 2015; Horgan and Timmons 2011.) *Third* (and given R5), there is a principled reason why humans simply are not capable of experiencing their own behaviors both as actions and as proximately psychologically state-causally generated: viz., the only kind of experiences as-of state causation that humans can undergo are experiences as of *non-agentive* state causation.

This last point leads to a second residual issue, however—one which seems to me considerably more serious. The challenge is to *explain why* the only kinds of experiences as-of state causation that humans can undergo are experiences as of *non-agentive* state causation.

One natural-looking candidate-explanation, of course, is the following. State causation is non-agentive by its very nature; hence, experiences as-of non-agentive state causation reveal the nature of state causation *simpliciter*, and do so more fully than would experiences as-of state causation that do not include the aspect of being non-agentive.

I myself am committed to repudiating this candidate explanation. And I am committed not only to the claim that state causation *simpliciter* sometimes is agentive, but also to the claim that despite this fact, state causation that is agentive cannot ever be *experienced* as such by humans. The challenge I face is to explain why not—and to do so in a credible way. (Such an explanation is likely to be somewhat speculative, and to involve encroaching upon matters scientific and empirical. So be it.)

In terms of evolution and natural selection, presumably it was important for nature to instill in humans (and in other critters too) the capacity to distinguish very vividly and very saliently between matters susceptible to effective influence by one's potential bodily motions and matters not thus susceptible. Consider, for example, a bear on a mountainside, generating bodily motions that transfer berries from an adjacent bush to the bear's mouth and on to its stomach. (I am assuming that the bear undergoes agentive phenomenology.) The bear now notices a giant boulder rolling down the mountain, heading straight toward himself. For him, there should be a *very* salient distinction between (on one hand) the experience of that boulder's motion as something not effectively controllable by bodily motions the bear might engage in, and (on the other hand) the experience of agentively transferring berries from the bush to its mouth and stomach. (And in light of this salient difference, the bear should rightly realize that he had better move his bear butt to another location pronto, before the boulder arrives!)

Now, one way for evolution to guarantee this kind of distinction would be to do two things together: (i) instill agentive phenomenology—in humans, bears, and various other reasonably sophisticated creatures capable of locomotion—that tends to arise only with respect to bodily motions that are under a critter's internal state-causal influence; and (ii) instill only a limited capacity for experiences as-of state causation, viz., a capacity for experiences as-of *non-agentive* state causation.

My suggestion is that this is exactly what evolution did do—and that evolution did it because of the clear advantages, in terms of fitness and survival/reproductive potential, thereby imparted upon creatures with both agentive phenomenology and inherently limited, inherently non-agentive, state-causal phenomenology. Distinctions of the kind experienced by the bear on the mountainside would be *very* vivid experientially, as they should be from the perspective of the evolutionary-biological design of agents with mentality.

No doubt there is more to say about this matter, by various parties to the dispute. Some will contend that state causation is inherently non-agentive and that genuine agency is inherently non-state-causal (although they owe arguments for both claims). Among those who embrace this dual contention, some—like Nida-Rümelin, I take it—will contend that humans often do exhibit genuine agency—and that in doing so, they operate outside of the state-causal nexus altogether. Others will instead contend that humans and other critters never exhibit genuine agency at all, and that experiences of agency are systematically illusory and non-veridical.

Were I among those who embrace the dual contention, I would opt for the second variant of it. I would do so because I believe that there is excellent scientific evidence for the thesis of physical state-causal completeness vis-à-vis humans and other critters, and also because I contend that willing-experiences are proximal state-causes of the bodily motions one experiences as one's actions. I would claim too that experiences of agency are a very *useful*

systematic illusion, from an evolutionary perspective—that nature instilled them in order to render very salient the differences between phenomena that are susceptible to effective influence by bodily motions that are state-caused internally within oneself, and phenomena that are not thus susceptible. (In Horgan 2014, I argue that color experience is an evolutionarily useful illusion, and that color “presentational content” is systematically non-veridical even though color “judgmental content” is not.)

But I repudiate the dual contention lately mentioned, in favor of the position I have been defending here. I contend that genuine, full-blooded, agency is a state-causal phenomenon in which one’s willings, with their inherent aspects of self-sourcedness and optionality, are proximal mental state-causes of the bodily motions one experiences as one’s actions. (I also contend that such psychological state causation is implemented by neuro-physical state causation, and indeed that physical causal completeness obtains for humans and for all other critters.) Although I think it is possible to believe that color experience is systematically non-veridical even though one cannot help undergoing it, and although indeed I do believe this, to me it seems literally impossible to believe—really and thoroughly believe—that I am not an agent of the kind I experience myself to be. I would be loathe to embrace a philosophical position entailing that I ought rationally to believe what is psychologically impossible for me to believe.

Martine Nida-Rümelin: rejoinder II

Horgan’s response to the argument from alienating experience (1)

Horgan has an excellent reply to the argument from alienating experience. After the relevant change of her agentive experience, Alexa fulfills three conditions: (a) she is capable of acting, (b) in acting she experiences her movement involved in her own acting as state-caused by her mental states, and (c) that experience is alienating. Each of these conditions is necessary for the example to serve its role in the argument from alienating experience. According to Horgan’s refined view, condition (c) is fulfilled since—as a consequence of fulfilling (b)—Alexa does not have the experience of being active. It follows—according to the refined version—that Alexa violates condition (a) since the experience of being active is essential to acting. It follows, furthermore, that it is impossible to construct a similar example fulfilling both (a) and (c). If the relevant experience of one’s behavior as state-caused is alienating *because* that experience excludes the experience of being active, then both (c) and (a) cannot be satisfied.

My original argument thus has no bite against Horgan’s refined view. However, I remain convinced that his proposal does not do duty to the anti-illusionist motivation we share. I will argue that his theory, contrary to his own intention, implies that the genesis of our own behavior in acting is, in a sense I will explain, radically different from how it appears in agentive experience. This may be surprising since his theory also implies that agentive experience is largely veridical. Before I present my new argument for the claim that Horgan’s proposal is, as one may say, a hidden illusion theory, I would like to make a few preliminary remarks. They will help clarifying what follows.

Willings and appearances—a misleading terminological choice (2)

Horgan introduces the term “willings” as a name for the experience of being active where that experience has two aspects, the one of ‘self as source’ (the OBA-aspect) and the one of

'optionality'. Given that terminology, Horgan's claims about the experience of being active (EA1, EA2, and EA3 below) turn into claims about willings (W1, W2, and W3 respectively):

EA1: The presence of an experience of being active is essential to action.

EA2: The experience of being active causes the movements involved in acting.

EA3: The relevant movement in a given action being state-caused by the experience of being active is a veridicality condition of the experience of being active.

W1: The presence of a willing is essential to action.

W2: Willings cause the movements involved in action.

W3: A veridicality (or rather, satisfaction) condition of a willing (related to a given action) is that the willing causes the associated movement.

"Willing" is an introduced term. We already have a pre-theoretical understanding of what willings are. Therefore, the terminological choice here at issue is not a harmless stipulation. Turning EA1, EA2, and EA3 into W1, W2, and W3 invites the reader to test her intuitions regarding EA1–EA3 using her implicit understanding of willings. To do so is however inadequate and leads to unfounded intuitions if, in reality, what we normally refer to using the term "willings" is distinct from experiences of being active. In fact, it is quite clear that this is so. Let me explain why.

One may express an important difference between willings and experiences of being active using Searle's famous notion of *direction of fit*. Willings have a world-to-mind direction of fit. Appearances have a mind-to-world direction of fit. Appearances, but not willings, present things to the subject as being the case. In experiences of being active, one's movements appear to have a certain genesis; they appear to be brought about by oneself. Such appearances are not directed at something still to happen as willings are. These remarks should suffice to make evident that the agentive experience of being active is not a willing.

Given the difference between experiences of being active on one side and willings on the other, one should not confuse the first three claims EA1–EA3 with the corresponding claims about willings W1–W3. Testing one's intuitions with respect to the former by thinking about the latter is illegitimate. In fact, it will lead one astray.

For instance, EA3 is a surprising claim with little intuitive appeal. Not so W3. It is plausible to think that a willing is only satisfied if the willing itself plays a causal role in bringing about what the subject wills. Therefore, testing EA3 by considering W3 will lend credibility to the former claim, which it does not merit. Similar remarks apply to EA2. One implicitly thinks of willings as something the subject does in an active manner. If one thinks of willings as active, then W1 and W2 may appear to be well motivated. One might tend to assent to W2 having in mind that movements in acting are active due to being caused by active willings. Implicitly presupposing that this is the only sense in which a subject can bring about her movements in an active manner, one then might tend to assent to W1 as well.

Such intuitions may have their own merits but they surely cannot support the corresponding claims about experiences of being active. Intuitions about willings (in the already introduced sense) and intuitions about experiences of being active have distinct subject matter. I therefore will resist adopting Horgan's proposed terminology.

Veridicality of agentive experience and the nature of being active (3)

I interpret Terry Horgan's proposal as implying a claim about what it *is* to be active, about what actively bringing one's own movement about *consists in*. This will play a role in my

argument below. Horgan does not explicitly formulate any such ontological claim. Nonetheless, or so I claim, he is committed to the thesis that for a subject to be active in a given movement *consists* in that movement being caused by the subject's experience of being active.

My argument for this interpretation is simple but I need to clarify a small point before I go on. Horgan speaks of experiences of being active as having two aspects: the aspect of apparently bringing the movement about and the aspect of optionality. I would like to leave the second aspect apart for the present discussion. I will focus on the former. Furthermore, I reserve the term "experiences of being active"—as I did already in the discussion so far—to agentive experiences with the former aspect. I leave it open whether such experiences always include the aspect of optionality. If so, this does not change anything for what follows.

Here is my argument for my ontological interpretation mentioned above. We all have an intuitive understanding of an important and fundamental distinction between changes of our own body or mind which happen to us (changes we passively undergo) on the one hand and changes we bring about ourselves (changes in which we are active) on the other. Not all philosophers agree with this first observation but I believe that Horgan has no objection. Furthermore, for a philosopher who concedes that we experience ourselves as active it seems unescapable to admit that to be active in one's own movement is precisely what renders such experiences of being active veridical. One may add that such experiences are our primary access to what it is to be active but this further plausible assumption which, I suppose, is in accordance with Horgan's ideas is not required for my argument.

Let us simply suppose the following:

Thesis AV (relating being active to the veridicality of experiences of being active)

S is active in her movement M is the veridicality condition of S's experience of being active in M.

Horgan does not talk about the nature of being active. He only expresses himself with respect to the veridicality condition of experiences of being active. However, if one accepts Thesis AV, then proposing a view about the condition which renders experiences of being active veridical and proposing a view about the nature of being active amount to the same. This is why I take it to be legitimate to presuppose in what follows that Horgan is committed to the following ontological claim.

Thesis NA (about the nature of being active)

The condition expressed by the following sentence (S1) consists in the fulfillment of the condition expressed by the following sentence (S2):

(S1) S is active in her own movement M.

(S2) S's experience of being active in M state-causes M.

Veridicality and the phenomenally manifest (4)

I need to attract the reader's attention to a further feature of Horgan's view, which might go unnoticed but will be crucial to my main argument against his proposal. One might think

that the veridicality conditions of a given experience are (or should be) most closely related to what appears to be the case to the subject undergoing the experience. If a thing appears to be spherical to me in a visual experience I am undergoing, then this justifies the claim that this thing being spherical is among the veridicality conditions of my present visual experience. This direction (the conclusion from what appears to be the case in a given experience to its veridicality conditions) is quite trivial. To deny it would be to miss the very notion of veridicality conditions. My own view is that the opposite direction should be accepted as well but I will not insist on this opinion nor presuppose it here (see for an argument in its favor Nida-Rümelin 2018b). Horgan's view must be that the opposite direction does not hold. To block the conclusion from veridicality conditions to what appears to be the case is the only way to make sense of the following two central claims of his proposal which otherwise would be plainly in contradiction.

Thesis VCA (about the veridicality condition of human experiences of being active)

The veridicality condition of the experience of a subject S of being active in her movement M is that M is state-caused by that very experience.

Thesis IE (about incompatible experience)

Human subjects cannot experience themselves as active in a given movement M and, at the same time, experience that movement as state-caused by their own experience of being active.

The relevant causal relation is thus the veridicality condition of experiences of being active and yet, in having these experiences *it does not appear* to the subject, on the phenomenally manifest level, as if her own movement were caused by her own experience. Horgan's view thus presupposes a distinction between two kinds of veridicality conditions (related although different distinctions are developed in Chalmers 2006 and Nida-Rümelin 2006). Veridicality conditions in the first sense, I will call that sense "the strong sense," can serve to describe the phenomenology of the experience at issue. They are manifest at the phenomenal level. In other words, if p is a veridicality condition in the strong sense, then p appears to the subject to be the case. Being active in one's movement is a veridicality condition in the strong sense of experiences of being active. The perceived object's being red is a veridicality condition in the strong sense of the corresponding kind of color experiences. Veridicality conditions in the weak sense are (or are supposed to be) relevant to assess whether the experience presents the world as it is (whether it is veridical). However, talking about how the experience presents the world in that context must not be misread as telling us what appears to be the case to the subject at the phenomenally manifest level. Arguably, a surface having a certain reflectance profile is a veridicality condition in the weak sense of visual experiences in which an object appears to be red. A movement being caused by the experience of being active is, according to my interpretation of Horgan's view, a veridicality condition in the weak (and not in the strong sense) of experiences of being active.

I suppose that the distinction between veridicality conditions in the weak and in the strong sense corresponds roughly to the distinction drawn by Horgan between content-based and implementational veridicality conditions. I take content-based veridicality conditions in Horgan's sense to be determined by the phenomenology of the experience alone, while

implementational veridicality conditions are (at least in part) determined by facts about the experience's causal role. If so, then the class of content-based veridicality conditions for a given experience should coincide with what I call veridicality conditions in the strong sense and the class of implementational veridicality conditions should either coincide or be part of what I call veridicality conditions in the weak sense. However, I do not presuppose this close relation between the two distinctions in what follows.

A regress problem (5)

If my argument in section (3) of the present reply goes through, then Horgan is committed to the following claim about the nature of being active already stated before.

Claim NA (about the nature of being active)

The condition expressed by the following sentence (S1) consists in the fulfillment of the condition expressed by the following sentence (S2):

(S1) S is active in her own movement M.

(S2) S's experience of being active in M state-causes M.

NA explains what it is to be active in terms of experiences of being active. This raises immediately a worry of a well-known type. (For a similar argument applied to dispositionalism about color, see e.g. Levine 2006.) If the term is used in the same sense on both sides, then we can plug the explanation given by NA of what it is to be active into (S2) and we get the following result.

Claim NA' (about the nature of being active)

The condition expressed by the following sentence (S1) consists in the fulfillment of the condition expressed by the following sentence (S2'):

(S1) S is active in her own movement M.

(S2') S's experience [of M being state-caused by her own experience of being active] state-causes M.

We can apply the same procedure to the occurrence of "being active" in S2' and we realize that we landed in an infinite regress.

As far as I can see, there are only two ways to avoid the regress. The first is to deny that the occurrences of "active" in (S1) and (S2) are uses of the term "active" in the same sense. The second is to forbid the relevant replacement in the context of the attribution of an experience. The first way out is not available if one subscribes to thesis AV motivated in section 3. According to AV, for a movement to be actively brought about by the subject concerned just is for it to fulfill the veridicality condition of experiences of being active. It is hard to see how one can accept AV and yet deny that "active" is used in the same sense in (S1) and (S2) above. I therefore put the first reply aside.

The second way out looks more promising. One may argue that attributions of experiences create intensional contexts analogously to belief attributions. One may take the example of color experience to illustrate the claim. Suppose that to be red is to have a surface with a certain reflectance profile R.⁴ Arguably it is not an argument against such an objectivist

account of color that we cannot replace “appears red” by “appears to have reflectance profile R” in attributions like “The apple appears red to Maria.” One may argue that replacing the property referred to by “to be red” by the—supposedly—co-referring term “to have reflectance profile R” in such an attribution changes the truth value of the sentence because colors are not *experientially* given as reflectance profiles although they actually are such profiles (according to the ontological claim here presupposed). One may argue that an analogous point applies to experience attributions like the one formulated by (S3):

(S3) Maria experiences herself as active in M.

According to this view, Maria experiences herself as active but she does not experience her movement as state-caused by that very experience despite the fact that to be active in one’s movement *is* nothing but that movement being caused by that very experience. Such a replacement would change the truth value of the attribution because it is not experientially present to the subject undergoing an experience of being active what being active consists in. Even if (S3) is true of Maria, this does not imply that she is, in that experience, under the impression that her own experience of being active state-causes the movement.

Attributions of appearances (that is attributions of experiences of what appears to be the case to the subject concerned in undergoing a given experience) thus create intensional contexts in a certain sense. One may explain that sense as follows. If the propositions p and q are rendered true by the very same state of affairs (by the very same ‘worldly’ conditions), then the replacement of q by p in an attribution of an appearance may nonetheless change the truth value of the appearance attribution.

As far as I can see this is how Terry Horgan would have to respond to the above-formulated regress problem. To respond in this manner amounts to accepting the above distinction between veridicality conditions in the weak sense and in the strong sense. In a case where replacing p by q turns a true appearance attribution into a false appearance attribution, p is a veridicality condition of the experience in the strong sense and q is a veridicality condition of the experience only in the weak but not in the strong sense. If no other plausible way out of the regress problem can be provided within Horgan’s account, then this is a further reason to say that he should accept the proposed distinction between strong and weak veridicality conditions.

Incompatible experiences versus experiences of apparent incompatibles (6)

I need just one further element to prepare my main objection. Horgan concedes in his response to my argument from alienating experience that the experience of being active is, at least for humans and perhaps more generally, incompatible with the experience of one’s own movement being state-caused. I agree that this is most likely the case. However, I would like to propose a different claim in the neighborhood, which I find phenomenologically convincing. Experiences of being active in a given movement on the one hand and of the relevant movement being state-caused on the other are experiences of apparent incompatibles. By experiences of apparent incompatibles, I mean this: a pair of experiences belongs to that kind if the subject undergoing these experiences is under the impression that the two experiences present conditions to obtain which cannot both be satisfied. In other words, on the phenomenally manifest level, the contents of the two experiences appear to be incompatible.

To illustrate the phenomenological notion of experiences of apparent incompatibles, one may choose the example of visual shape perception. In certain phenomenal types of visual

experiences, the subject undergoing the experience is under the impression of being presented with a spherical object. In a different phenomenal type of visual experiences, the subject is visually under the impression of being presented with a cubical object. Here is my phenomenological proposal: a subject having an experience of an object as being spherical and of a different object as being cubical is thereby under the impression that no object could possibly have both shapes at once. The incompatibility of the two shape properties a subject is presented with in such a case is available at the phenomenal level. The incompatibility of the two shape properties one is presented with in such experiences is phenomenally manifest.

Before I go on, I would like to note that not all incompatible experiences (for a given species) are experiences of incompatibles. For instance, humans are unable (under normal circumstances) to experience an object as reddish and greenish at once and in the same place. (See Nida-Rümelin and Suarez 2009 for a detailed discussion of supposed real counterexamples in unusual circumstances.) However, there is no phenomenological ground for supposing (as far as I can see) that experiences of being greenish and being reddish are experiences of apparent incompatibles. Furthermore, experiences of apparently incompatible conditions may nonetheless occur. One can be presented, in perception, with a situation that appears to satisfy conditions, which, at the same time, appear to be incompatible.⁵

Reflecting on the example concerning visual shape perception, a further observation comes to mind. It is not a contingent fact about visual experiences of something being spherical and of something being cubical that they are experiences of apparent incompatibles. If for some reason we were to suppose that a subject had an experience of something being spherical and of something being cubical and yet did not thereby have the impression that nothing could have both properties at once, then we surely would have to conclude that we were about to mis-describe the subject's visual phenomenology. It is part of what it is to have experiences of these phenomenal types that the properties presented in those experiences appear to be incompatible. This observation motivates introducing a further phenomenological term: some pairs of experiences are *essentially* experiences of apparent incompatibles.

There is a further phenomenally salient aspect to which I would like to attract the reader's attention using the same example. When you see an object as spherical and another object as cubical, the incompatibility of the properties you are presented with appears to be *visible*. You simply seem to *see* that nothing can have both properties at once. One might put it in this way: you seem to have sufficient access to the nature of the properties just by seeing them in order for you to also have visual access to their incompatibility. Your experience of those properties as incompatible appears to be due to their very nature, where that nature is (or rather: appears to be) in part available at the phenomenal level. Let us say of such cases that the apparent incompatibility (of the relevant properties or of the relevant conditions) is *experientially evident* in having the experiences at issue.

More about what appears to be the case in agentive experience (7)

I would now like to propose a few phenomenological observations about the case of agency. I will have to leave it to the reader to test their plausibility based on his or her own phenomenological reflection:

- PO1: Experiences of oneself as active in a given movement and experiences of that movement as being state-caused are (for human beings) experiences of incompatibles.
- PO2: Experiences of oneself as active in a given movement and experiences of that movement as being state-caused are essentially experiences of incompatibles.

PO3: The incompatibility of oneself being active in a movement and one's own movement being state-caused is experientially evident in any relevant potential agentive pair of experiences.

A lot would have to be said to justify these phenomenological claims by careful reflection. An obvious difficulty resides in the fact that we actually never have experiences of our own movements in actions as state-caused. I presuppose that we can nonetheless imagine in a sufficiently clear manner how that would be like. I cannot go deeper into the phenomenology of agency in order to convince the reader of PO1–PO3. I hope that these claims have enough intuitive appeal for it to be of some interest to see what happens to Horgan's proposal if one takes them to be true. Before I discuss these consequences, let me explain in greater detail what these claims state.

One may wonder if experiences of oneself as active on the one hand and experiences of one's own behavior as state-caused on the other are analogous to the case of experiences of reddish and greenish or rather analogous to the case of visual impressions of different shapes. My claim is that they are analogous to the latter and not to the former. They may be incompatible for humans (like the color experiences under consideration) but this is not all. They are, furthermore, like the pair of shape perceptions mentioned. It is part of their phenomenal character that—in having an experience of being active and in having an experience of your movement being state-caused (perhaps in the next moment like in the case of Alexa)—you are under the impression that the same movement could not possibly fulfill both conditions. This is the content of PO1. However, we can go further. Reflecting on what it is like to be active in one's movement and on what it would be like to experience one's own movement as state-caused, one will be led to a further and stronger conclusion. Suppose that a creature has experiences of type E1, which we are inclined to consider as experiences of being active in her movement and that it also has experiences of type E2, which we are inclined to consider as experiences of her own movement as being state-caused. Suppose we learn, in addition, that the creature does not experience what appears to be the case in E1 as incompatible with what appear to the case in E2. Then, or so I claim, we have reason to resist the former inclinations. If the appearance of incompatibility is lacking, then either experiences of type E1 are very different from our experiences of being active and thus do not deserve to be classified with them or experiences of type E2 are very different from experiences of being state-caused. This is what PO2 affirms.

Suppose that PO1 is correct. Then PO3 is quite obvious, I would say. If the two conditions are given in your experience as incompatible, then you surely have the impression that these conditions are in fact incompatible by their nature. You surely are under the impression that you are aware, experientially, of their actual incompatibility. It seems to you that you have experiential access to what it *is* to be active in your movement and to what it *is* for a movement to be state-caused which allows you to experience their incompatibility. If you have doubts about that claim, I ask you to reconsider the shape case first in order to judge whether you can agree that this is a case of experientially evident incompatibility in the sense introduced. If you come to a positive result, you might be in a position to admit that the agentive case we are interested in is analogous to the shape case in that respect.⁶

Veridicality without genuinely veridical appearance (8)

We are now in a position to see why and how Horgan's proposal does not do duty to his own intention. We would like to be able to rationally endorse a theory about agency, which allows us to assume that our own behavior comes about in the way it appears to come about

in agentive experience. When a subject experiences her movements as actively brought about by herself then, in undergoing that experience, her movements appear to be generated in a particular way. We would like to develop a well-founded and scientifically acceptable theory about agency according to which the subject's movement is actually generated in the way it appears to be generated in those experiences. This, I believe, is a central motivation behind Horgan's work about agency, which I fully endorse myself. Horgan's view saves the claim that the experience of being active is veridical. However, given the special situation in which we are if (a) his refined view is correct, if (b) the claim above about the ontology of being active (which I attribute to Horgan) is correct, and if (c) my additional phenomenological observations are adequate, then such veridicality is not enough. In that particular situation, there is a huge gap between the experience being veridical and the experience being such that things appear as they are to the subject undergoing the experience.

Let us first leave my new phenomenological observations apart and let us focus on (a) and (b). We thus assume Horgan's view and we presuppose that it commits him to the ontological claim that to be active in a movement M consists in the movement being state-caused by one's experience of being active. In that case, our situation in agentive experience has the following features:

- F1 In agentive experience, we experience ourselves as actively bringing about the relevant movement M.
- F2 To actively bring about one's own movement consists in the satisfaction of a certain condition C (which is that the subject has an experience of being active which state-causes the movement).
- F3 We are experientially blinded to condition C in the following sense: we cannot act and at the same time undergo an experience, which is such that condition C appears to be satisfied to the one who has that experience.
- F4 Condition C is experientially accessible in this sense: there are potential experiences E, such that in undergoing E the subject is under the impression that condition C is satisfied.
- F5 Experiences of being active in one's own movement are veridical (in cases where they are) in virtue of the satisfaction of C.

The last feature, F5, is crucial to what I find unsatisfying about Horgan's view. It is a consequence of F2 (the ontological claim). Condition C is what really underlies the veridicality of veridical experiences of being active. However, *that* condition plays no role in what is phenomenally manifest to the subject who experiences herself as active. It is only a veridicality condition in the weak sense but it is no veridicality condition in the strong sense. Even worse, although one could in principle have an experience in which C appears to be satisfied (an experience such that C is a veridicality condition in the strong sense of that experience), human beings are unable to have such experiences. They are unable to be phenomenally aware of what makes their experience of being active veridical. This is already enough to see that veridicality is a very weak condition in the present context. The relevant experiences are veridical *in virtue* of the fulfillment of a condition, which does not appear (and could not appear in the case of humans) to be fulfilled to the subject concerned.

Things become worse if we assume, in addition, that the above-motivated phenomenological observations are adequate. If they are, then our incapacity of having experiences with C as their veridicality condition in the strong sense prevents us from undergoing a very disturbing experience. Suppose a subject were given that capacity perhaps by removing a

certain suppressing mechanism in the brain, which has evolved in members of the human species 'in order' to guarantee certain survival advantages. That person, call her Aliena, has the experience of being active when she moves in her actions and she also has the experience of that movement being state-caused by her experience of being active. If what I said before is correct, then Aliena cannot but have the impression that she is in an impossible situation. To have these two experiences necessarily involves that the conditions that appear to be satisfied appear to be incompatible. Like Alexa, Aliena would have an alienating experience but for different reasons. Her experience would be alienating, not in virtue of a lack of the experience of being active (as it is for Alexa); it would be alienating in virtue of the apparent satisfaction of a further condition which appears to be incompatible with how things appear in her experience of being active.

We now can say this. We are experientially blinded to what makes our experiences of being active veridical. If we were freed from that blindness then we would be, necessarily, under the illusion that we are in an impossible situation, in a situation in which our movements are generated in two incompatible ways. Furthermore, according to PO3, we would then have the impression that the two conditions are, by their experientially accessible nature, incompatible. We would suffer the illusion that our experience gives us sufficient access to their nature for us to be aware, on a phenomenally manifest level, of their real incompatibility.

One may even go further and argue for a stronger claim. When visually experiencing an object as spherical one is aware on a phenomenally manifest level of the object having a shape, which is incompatible with being cubical. So, not only are visual experiences of being spherical and of being cubical experiences of incompatibles in the above introduced sense, the relevant incompatibility is phenomenally present even when a subject has one of those experiences at a given moment and not the other. Arguably, a similar observation is in order with respect to experiences of being active. In experiences of being active, a condition appears to be satisfied which appears to be, due to its experientially accessible nature (accessible by that very experience), incompatible with being state-caused. If so, then experiences of being active are non-veridical after all. In having them we suffer the illusion that the way our behavior is generated is incompatible with a condition C which, in reality, renders the experience of bringing something actively about veridical. This is a radical illusion indeed.

My above argument is supposed to show that Horgan's proposal cannot satisfy its underlying motivation. The strength of the result obtained depends on how many premises one uses. Introducing only the additional ontological thesis about what being active consists in we reach the following result: the veridicality of veridical experiences of being active is due to the satisfaction of a condition C, which is experientially inaccessible to humans. Being active consists in C, but experiences of being active provide no access whatsoever to what being active consists in. C, and thus the condition which renders experiences of being active veridical, is totally absent at the level of experience. C is absent in human agentive experience despite the fact that it is, in principle, experientially accessible. Introducing the phenomenological claims PO1–PO3, we arrive at a more disturbing conclusion. A subject who is able to be phenomenally aware of being active and, at the same time, of the fulfillment of the condition C, which in fact constitutes being active, would be—in having both experiences—under the illusion that her behavior emerges in two incompatible ways. Adding a further phenomenologically plausible claim, we come to the most puzzling result: in experiences of being active, we suffer from a serious illusion about what it is to be active. In such experiences, our behavior appears to be generated in a way, which is incompatible with C while in reality C is even what constitutes being active and what therefore renders the experience of

being active veridical. Furthermore, we suffer from the illusion of being immediately aware of that incompatibility by having experiential access to what it is to be active.

An alternative approach (9)

In order to avoid the problems highlighted in what precedes and in order to do duty to the motivation behind Horgan's view, one may take a more radical decision and fully endorse what appears to be the case in agentive experiences with the specific aspect of being active in one's movements. The following conditions appear to be fulfilled in experiences of being active:

A1: The subject brings her movement about in an active manner.

A2: The experience of being active provides partial access to what it is to be active.

A3: That experiential access provides one with the capacity to be phenomenally aware of the real incompatibility between (a) a movement being brought about in an active manner by the subject and (b) the movement being state-caused.

Such an account thus includes the claim that these conditions are actually fulfilled. It is thereby in contradiction with the received view that all physical events have only physical events as causes. Many philosophers, including Terry Horgan, are convinced that, in virtue of that feature, an account of this kind is seriously undermined by scientific wisdom. I doubt that this is true but will not start discussing that further issue here. (Compare for a discussion of that issue Nida-Rümelin 2016.)

Horgan's proposal is meant to fulfill the following two constraints:

Constraint 1: The theory is compatible with the dogma that physical events have only physical events as causes.

Constraint 2: The theory is compatible with the claim that our behavior appears to be generated in our experiences of being active in the way it actually is generated. Agentive experience is not radically and systematically illusionary.

Based on the considerations presented in this part of our paper, I suspect that no account fulfilling both constraints is actually available. One must choose, I believe, between two options: respecting constraint 1 and abandoning constraint 2 or respecting constraint 2 and abandoning constraint 1. I would recommend the second option. I wonder what choice Terry Horgan would prefer if he agreed on my diagnosis that the choice is unavoidable.

Terry Horgan: proposal III

Martine Nida-Rümelin's latest installment contains a challenging new argument for her claim that my position "does not do duty to the anti-illusionist intuition." The new argument is in much the same spirit as her earlier argument that was keyed to her thought experiment about Alexa, but does not deploy that thought experiment.

Largely because of Nida-Rümelin's ongoing objections, I now believe that my position needs to be refined yet further, above and beyond my prior refinements earlier in this dialogue. I also believe that the needed additional refinements are independently plausible and therefore should be embraced anyway, quite apart from the role they will play in my reply to her new argument.

My earlier articulations of my position put heavy emphasis on the contention that although agentive phenomenology **is not as-of** one's action being a psychological state-causal process, agentive phenomenology also **is not as-of** one's action **not** being a psychological state-causal process. I stand by this contention, but I now realize that it can be refined in either of two incompatible ways. (Both ways are compatible with the contention itself, despite being incompatible with each other.) One way is to embrace the claim that agentive phenomenology has **no negative content at all** involving psychological state causation. But a second way is to acknowledge that agentive phenomenology is indeed **as-of** one's action **not** being **a certain kind** of psychological state-causal process, while also claiming nonetheless that agentive phenomenology is **not as-of** one's action **not at all** being a psychological state-causal process.

I believe that my position should be refined the second way, not the first. Doing so is independently plausible, because this alternative accords better with the introspectable character of agentive phenomenology itself. Doing so also will enable me to resist Nida-Rümelin's new argument against me.

I begin with some preliminary remarks concerning satisfaction conditions. I now think that in principle, there can be three kinds: content-based, noncontingently implementational, and contingently implementational. Content-based satisfaction conditions are determined directly by the intentional content of the phenomenology: either they are literally part of that content, or else they are entailed by it as a matter of logic and/or analytic truths. Noncontingently implementational satisfaction conditions are not content-based, but are entailed by the content of the phenomenology plus synthetic necessary truths. Contingently implementational satisfaction conditions are neither content-based nor noncontingently implementational, and instead are dependent in part on contingent truths about the nature of the given kind of cognitive agent.

I also embrace a distinction that is somewhat orthogonal to this three-way distinction: viz., between satisfaction conditions that are phenomenally manifest, and ones that are phenomenally non-manifest. This is Nida-Rümelin's terminology, which she also aligns with 'weak sense' and 'strong sense'.

Another preliminary matter is that she objects to my use of the expression 'willing' as an allegedly misleading terminological choice. I think this concern can be set aside for present purposes, by just changing my terminology. So I will instead use the expression 'agentive-instigation experience'.

There is one more preliminary matter, before I turn to the main business at hand. She wonders whether I face a regress problem. She suggests that I do not, as long as I say that attributions of appearances create intensional contents—but that this commits me to the distinction between the weak and strong senses of 'satisfaction condition'. I'm fully on board with all this.

I turn now to further elaboration of the position I wish to espouse. Doing so involves attending carefully to refinement R5 from my second installment above, and at the same time acknowledging and embracing certain claims about negative satisfaction conditions of agentive phenomenology concerning matters of psychological state causation of one's bodily motions—and all the while keeping clear about which claims are, and which claims, are not, phenomenally manifest.

Let a process be *experiential-acquaintance-wise state-causal* (for short, EA-wise state-causal), just in case it is state-causal in a way with which humans are experientially acquainted. This feature will prove quite important in what follows. In particular, the following consequence of R5 will be important: all EA-wise state-causal processes are *non-agentive* state-causal processes.

I hereby incorporate into my position the following additional claims, which I list under three categories: phenomenally manifest facts, facts about the limits of reliable introspectability, and phenomenally non-manifest facts.

Phenomenally manifest facts

- PM1. Any experience one can have as-of a process's being state-causal is as-of that process's being non-agentively state-causal.
- PM2. Any experience one can have as-of a process's being **not** state-causal is as-of that experience's being **not non-agentively** state-causal.
- PM3. A satisfaction condition of an experience as-of agentively moving one's body a certain way is that the phenomenon thus experienced includes an agentive-instigation experience vis-à-vis the bodily motion.
- PM4. In experiences of being active, a condition appears to be satisfied which appears to be **incompatible** with one's bodily motion being **EA-wise state-caused** by one's agentive-instigation experience.
- PM5. A (negative) satisfaction condition of an experience as-of agentively moving one's body a certain way is that the motion is **not EA-wise state-caused** by one's agentive-instigation experience.

Facts about the limits of reliable introspectability

- L1. One cannot reliably ascertain introspectively whether or not an experience as-of a process's being **not EA-wise state-causal** is an experience as-of that process's being **not state-causal at all**.
- L2. One cannot reliably ascertain introspectively whether or not an experience of agentively moving one's body a certain way is an experience as-of the process's being **not state-caused at all** by the agentive-instigation experience.

Phenomenally non-manifest facts

- PNM1. Not all state-causal processes are non-agentively state-causal; some are agentively state-causal.
- PNM2. A satisfaction condition of an experience of moving one's body a certain way is that the agentive-instigation aspect of the experience state-causes the appropriate bodily motion.
- PNM3. It is **not** the case that in experiences of being active, a condition appears to be satisfied which appears to be **incompatible** with one's bodily motion being state-caused by one's agentive-instigation experience.
- PNM4. It is **not** the case that a (negative) satisfaction condition of an experience as-of agentively moving one's body a certain way is that the motion is **not state-caused at all** by one's agentive-instigation experience.

I have several comments about these various claims, by way of further elaboration. Concerning PM3: Although an experience as-of moving one's body a certain way is trivially an occurrence condition of itself (because such an experience cannot occur without the presence of an instigation-experience as a component), it is not trivial to say that such an experience is a *satisfaction* condition of itself. (By contrast, consider for example a conscious belief such as the belief that Aristotle was a teacher of Alexander the Great; this state's satisfaction condition—viz., that Aristotle was a teacher of Alexander the Great—would be met even if the state itself did not exist.) The claim that genuine agency requires agentive phenomenology is substantive and non-trivial.

Concerning PM4 and PM5: These claims explicitly acknowledge that agentive experience includes *negative content* regarding state causation: agentive experience is phenomenologically as-of one's action being **not** state-causal in a certain way (viz., being not *EA-wise* state-causal). Nida-Rümelin's objections to my earlier articulations of my position have helped me appreciate that agentive experience does indeed have this kind of negative content, and that its having such content is introspectively manifest. I should have acknowledged such negative content all along, rather than emphasizing so heavily the fact that "not as-of Φ " is logically weaker than "as-of not Φ ."

Concerning PNM3: Although I said in an earlier installment that I think this is an implementational satisfaction condition, I am now not sure about this. For the present I can be neutral between the following three alternatives: (a) it is actually a content-based satisfaction condition, albeit one that is not phenomenally manifest; (b) it is a non-contingent implementational condition (a view one might embrace, for instance, because one holds that it is a synthetic necessary truth that the only fully intelligible account of agency will incorporate this satisfaction condition); or (c) it is a contingent implementational satisfaction condition that perhaps is specific to certain kinds of creatures (including humans).

Finally, concerning all of the theses PM1–PM5, L1–L2, PNM1–PNM4 and the question "Why should it be that these theses are true?": I would again reiterate my remarks at the end of my second installment above, about the bear and the berries. There are plausible *evolutionary-biological* reasons why the only kinds of state-causal processes that humans can experience *as* state-causal are non-agentive state-causal processes. Maybe there are further reasons too why all of these theses are true—a matter of much philosophical interest, in my view. Maybe no realistically possible kind of intelligent creature can be capable of appreciating practical reasons *qua* reasons, and of acting upon reasons *qua* reasons, unless such a creature resides phenomenologically within an experiential "space of reasons" that is (and must be) *experientially* outside the "space of causes"—even though the creature does not (and cannot) reside *metaphysically* outside the "space of causes." Sellars (1962) looms large here, and it is worth juxtaposing Sellars' own talk of the "manifest image" of human beings with Nida-Rümelin's talk of what is "phenomenally manifest" in experience.

I am now ready to reply directly to Nida-Rümelin's principal argument, in her most recent installment. As nicely summarized at the end of section 8 of that installment, the argument comes in three stages. Each stage involves an alleged bad consequence of my position. Also, each successive stage appeals to additional premises beyond those at the previous stage(s), and involves an alleged consequence even worse than the alleged bad consequences at the previous stage(s). I will address the three stages in turn.

The first stage focuses primarily on what she calls condition C, which I now have formulated as PNM2. She says this about what I am allegedly committed to by embracing this satisfaction condition:

Being active consists in C, but experiences of being active provide no access whatsoever to what being active consists in. C, and thus the condition which renders experiences of being active veridical, is totally absent at the level of experience. C is absent in human agentive experience despite the fact that it is, in principle, experientially accessible.

I have three points in response.

First, I deny that on my account, experiences of being active provide *no access whatsoever* to what being active consists in. An important *part* of what being active consists in

is the presence of agentive–instigation experience itself; and this part is phenomenally manifest.

Second, I deny that in the pertinent sense, C’s obtaining “is, in principle, experientially accessible.” The only kinds of state–causal experiences that are experientially accessible to humans, I contend, are *non-agentive* state–causal processes.

Third, although I do contend that C itself is an experientially non-manifest satisfaction condition of agentive experience, I deny that this makes my view objectionably close to an illusionist position. Agentive–instigation experience is itself phenomenally manifest, after all. So is the experiential aspect of one’s body’s moving the way it does *because* of one’s having instigated that motion. Admittedly, this “rational because” aspect is not experienced **as** state–causal, even though it **is** state–causal. But that is because the only kinds of state–causal processes that humans can experience **as** state–causal are non-agentive state–causal processes.

Stage 2 invokes Nida-Rümelin’s phenomenological claims PO1–PO3, involving the phenomenally manifest incompatibility between experiences of oneself as active in a given bodily movement and distinct experiences of such a bodily movement as being state–caused. She says:

Introducing the phenomenological claims PO1 to PO3, we arrive at a more disturbing conclusion. A subject who is able to be phenomenally aware of being active and, at the same time of the fulfillment of the condition C, which in fact constitutes being active, would be—in having both experiences—under the illusion that her behavior emerges in two incompatible ways.

She sets forth a new thought experiment involving a character named *Aliena*, who Nida-Rümelin describes as suffering just such an illusion. She writes:

Like *Alexa*, *Aliena* would have an alienating experience but for different reasons. Her experience would be alienating, not in virtue of a lack of the experience of being active (as it was for *Alexa*); it would be alienating in virtue of the apparent satisfaction of a further condition which appears to be incompatible with how things appear in her experience of being active.

I have the following response. Although I grant Nida-Rümelin’s phenomenological claims PO1–PO3, I also claim that for humans, experiences of a phenomenon as state–causal are always as-of the phenomenon’s being *EA-wise* state–causal (thesis PM1). Likewise, experiences of a phenomenon as *not* state–causal are always as-of the phenomenon’s being *not EA-wise* state–causal (thesis PM2). So although it is true that experiences of oneself as active in a given movement, and experiences of that movement as being *EA-wise* state–caused, are essentially experiences of incompatibles, and although being active in a given movement is indeed incompatible with that movement’s being *EA-wise* state–caused, it is *not* true—cf. thesis PNM4—that experiences of oneself as active in a given movement are experiences as-of that movement’s being *not at all* state–caused.

Thus, poor *Aliena* is not undergoing an experience which, according to me, accurately represents the psychological state–causal nature of human agency. On the contrary, she is having a (deeply alienating) experience that is simultaneously both (i) as-of being active in a given movement, and (ii) as-of that movement’s being *EA-wise* state–caused. No movement can be both.

Concerning a hypothetical subject who is able simultaneously to be phenomenally aware of being active in a given movement and also to be phenomenally aware of that movement's being state-caused by the subject's agentive-instigation experience, I have four things to say. First, it would not be like *Aliena's* experience. Rather, it would be both (i) as-of being active in a given movement, and (ii) as-of that movement's being *agentively* state-caused. Second, we humans are not capable of such an experience, and cannot envision what it would be like. We are only capable of experiencing a process as state-causal by experiencing it as *non-agentively* state-causal. Third, I do not see why a (super-human) creature who *could* have such an experience—if a creature like this is possible, in some sense of 'possible'—need necessarily find it alienating. Fourth, it is worth pondering, somewhat in the spirit of Kant-style reflection about transcendental 'conditions of possibility', whether any possible creature capable of appreciating and acting upon reasons, *qua* reasons, would necessarily need to be experientially blind to whatever processes within itself causally implement its rational agency. (Perhaps a condition of possibility for occupying "the space of reasons" is to be *experientially* outside of the "the space of causes.")

The upshot is the following, regarding stage 2 of Nida-Rümelin's argument. Although I am indeed committed to claiming that we humans are experientially blinded to *part* of what makes our experiences of being active veridical, I am not committed to claiming that if we were freed from that blindness then we would be, necessarily, under the illusion that we are in an impossible situation. (This latter putative commitment is the alleged bad consequence of my position that supposedly arises in stage 2 of Nida-Rümelin's argument.) Instead, if, *per impossibile*, we were freed from that blindness, then we would experience our bodily motions both as the products of our agency and as *agentively* state-caused by our agentive-instigation experiences—rather than experiencing them, in the manner of *Aliena*, both as the products of our agency and as *non-agentively* state-caused by our agentive-instigation experiences.

Stage 3 invokes the claim that the incompatibility between being agentively induced and being psychologically state-caused is phenomenologically present "even when a subject has one of those experiences at a given moment and not the other." Given this claim, and citing my contention that she calls "condition C" (*viz.*, that the subject has an experience of being active which state-causes the bodily movement), Nida-Rümelin says,

we come to the most puzzling result [to which Horgan is allegedly committed]: in experiences of being active, we suffer a serious illusion about what it is to be active. In such experiences, our behavior appears to be generated in a way which is incompatible with C while in reality C is even what constitutes being active and what therefore renders the experience of being active veridical. Furthermore, we suffer from the illusion of being immediately aware of that incompatibility by having experiential access to what it is to be active.

The key contention here, I take it, is that a phenomenally accessible part of the content of agentive experience is that one's motion is *not state-caused* by one's agentive-instigation experience. This being so, the argument goes, my own position commits me—in spite of my intentions—to the claim that agentive experience is illusory. Such experience presents one's behavior as generated in a way that is incompatible with C, despite C's constituting the real nature of agency.

My response is essentially the same as my response to Stage 2. Although the experience of being active does have significant negative content regarding matters of psychological state causation, it does not actually have content that is incompatible with C. Rather, the experience of being active vis-à-vis a given bodily motion is, inter alia, an experience as-of that motion's not being *EA-wise* state-caused by one's agentive-instigation experience. There is no illusion here. On the contrary, being active really is incompatible with such EA-wise state causation, because EA-wise state causation is *non-agentive* state causation.

I turn finally to some concluding remarks. By my lights, Nida-Rümelin thinks she can reliably introspect certain phenomenally manifest aspects of agentive experience which I claim are not present at all in agentive experience (and hence are not really phenomenally manifest). I have come to think, though, that the underlying mistake here is more subtle than I once thought it was. (And I have always thought that it's quite subtle!) The real mistake, I now think, is **not** the putative mistake of conflating (i) the phenomenally manifest fact that agentive phenomenology is **not as-of** one's action being a psychological state-causal process with (ii) the *putatively* phenomenally manifest fact that agentive phenomenology is **as-of** one's action being **not** a psychological state-causal process. Rather, it is the mistake of conflating (iii) the phenomenally manifest fact that agentive phenomenology is **as-of** one's action being **not** an *EA-wise* psychological state-causal process with (iv) the *putatively* phenomenally manifest fact that agentive phenomenology is **as-of** one's action being **not at all** a psychological state-causal process. And this mistake is all the easier to make because **one cannot reliably ascertain introspectively** whether or not an experience as-of one's action being not an *EA-wise* psychological state-causal process is also an experience as-of one's action being not *at all* a psychological state-causal process.

Suppose, though, that I am wrong about all this, and that she is right in claiming that we must choose between (on one hand) a broadly materialist metaphysics of mind and (on the other hand) the claim that our behavior really is generated in the way it appears in our experiences to be generated. She wonders, at the end of Section 9 of her recent installment, which of these options I would choose if I found myself accepting that this choice is unavoidable. My answer is this: I would choose the first option, embrace an error theory about agentive phenomenology, and then seek philosophical psychotherapy regarding my residual inability to believe—really *believe*—that I am not an agent of the kind I experience myself to be.⁷

Outlook: agreements, disagreements, and conflicting background assumptions

We will conclude by summarizing both the principal contentions about which we agree and the principal points of disagreement between us. We agree on each of the following claims (labeled with 'A' for 'Agreement'):

- A1 Human action typically has a distinctive phenomenology, including the aspect of action-instigation (self as source/of being active), that is richly intentional.
- A2 Because human agentive experience is intentional, it has satisfaction conditions.
- A3 In order for human agentive experience to be veridical on a given occasion of human behavior, its satisfaction conditions must be met on that occasion.
- A4 Human agentive experience normally is veridical; i.e., normally when such experience occurs, its satisfaction conditions are indeed satisfied.

- A5 At least some of the satisfaction conditions of agentive experience are phenomenally manifest—i.e., are manifest in the specific way that human behavior appears in human experience to be generated.
- A6 Humans are agents of the kind they experience themselves to be.
- A7 Any essential aspect of how human behavior appears in human experience to be generated is a phenomenally manifest satisfaction condition of human experience.
- A8 Condition C—the putative requirement that one’s experience of act-instigation state-causes one’s pertinent bodily motion(s)—is not a phenomenally manifest condition of human experience.
- A9 There is a *prima facie* incompatibility between (i) the way that human behavior appears in human experience to be generated and (ii) the contention that condition C is among the satisfaction conditions of human agentive experience.

The principal matters of disagreement between us involve the following claims (labeled ‘D’ for ‘disagreement’):

- D.1 Condition C is among the satisfaction conditions of human agentive phenomenology (although neither C itself nor its being a satisfaction condition is phenomenally manifest).
- D.2 There is no genuine incompatibility between the way that human behavior appears in human experience to be generated and condition C.

Terry Horgan affirms D.1 and D.2, whereas Martine Nida-Rümelin denies each of them.

Horgan has been making the following background assumption throughout this dialogue (labeled ‘B’ for ‘Background’):

- B1 If condition C does not conflict *logically* with the manifest satisfaction conditions of human agentive experience, then there is no genuine incompatibility between (i) the way that human behavior appears in human experience to be generated and (ii) the contention that condition C is among the satisfaction conditions of human agentive experience.

In light of B1 and A9, he therefore has been regarding the dialectical situation as imposing on him a two-part burden of proof: viz., the burden of arguing (1) that condition C does not conflict logically with the phenomenally manifest satisfaction conditions of agentive experience, and (2) that the *prima facie* incompatibility cited in A9—which Horgan has been regarding as a *prima facie* *logical* incompatibility—can plausibly be explained away as subtly mistaken.⁸ He has attempted to shoulder this burden above.

One key question, of course, is whether Horgan has successfully discharged this two-part dialectical burden. Nida-Rümelin is not persuaded that he has done so. Others too might not be persuaded, and might even contend that among the phenomenally manifest satisfaction conditions of agentive experience is the requirement that one’s pertinent bodily motion is *not* state-caused by one’s act-instigation experience (or by any other psychological state of oneself).

But here at the end-stage of this dialogue, as we two have sought to reach agreement about the nature of our residual disagreement(s), it has become clear to us that perhaps our most fundamental dispute all along has really been about B1 itself. Nida-Rümelin has been making the following two background assumptions throughout this dialogue:

- B2 In order for human behavior really to be generated in the way it appears in agentive experience to be generated, a further requirement must be met: the non-manifest satisfaction conditions which, if satisfied, make it the case that the manifest satisfaction conditions are fulfilled must *provide a phenomenologically adequate theoretical account* of the way that human behavior appears in agentive experience to be generated.
- B3 C does not provide a phenomenologically adequate theoretical account of what appears to be the case in agentive experience.

Nida-Rümelin is acutely aware of the fact that B2 and B3 remain quite obscure as long as no explication is provided of what it is for a condition to be ‘a phenomenologically adequate theoretical account of what appears to the case’ in a given type of experience. At present, she has no such explication to offer.⁹ But the guiding idea behind B2 is this: fulfilling A6 requires that C must be nothing other than a more precise and theoretically informed reformulation of the way our own behavior appears to be generated in experiences of being active.

Nida-Rümelin contends that condition C fails to meet this further requirement. Horgan agrees. Thus, if B2 is true then Horgan’s assumption B.1 is mistaken: Horgan’s claim D1 (together with the assumption that C is in general fulfilled when the relevant kind of agentive experience occurs) does not imply A6 *even if he is right that C does not conflict logically with the phenomenally manifest satisfaction conditions of agentive experience*.

So our most basic disagreement turns out to be about conflicting respective background assumptions that we have been making, respectively, throughout: Horgan’s B1 vs. Nida-Rümelin’s B.2. Our efforts in trying to agree about how we disagree have brought this fact into the open. The debate will continue, and we invite others to join in.

Notes

- 1 Fully acknowledging the phenomenal character of agentive experience, in my view, requires acknowledging that such experience has distinctively *agentive* phenomenal aspects—over and above whatever other aspects (e.g., aspects as-of one’s body moving a particular way) it might have in any specific agentive experience.
- 2 This use of ‘intentional’ and ‘intentionality’ is fairly standard in analytic philosophy of mind, with the term ‘intentionality’ often being glossed as ‘aboutness’. It is an interesting issue just how this usage comports with the notion of intentionality inherited from the Phenomenological tradition descended from figures like Brentano and Husserl, in which intentionality is understood as “directedness to,” or awareness of, “intentional objects.”
- 3 Here and throughout, I employ the expression ‘state causation’ rather than the expression ‘event causation’. I am using the term ‘state’ in a broad sense that is intended to include not only persisting “non-changes” but also (nominally governed) events and processes. My principal reason for eschewing ‘event causation’ is that there is something phenomenologically episodic—something temporally located, and thus “event-ish”—about experiences of self-as-source. Thus, the expression ‘state causation’ works better than ‘event causation’ as a way of expressing how one’s behavior is *not* presented to oneself in agentive experience: one’s behavior is not presented as occurring within a causal nexus that is evolving, by itself, apart from any agentive intervention.
- 4 I do not presuppose the truth of that objectivist theory about the nature of color. In fact, I believe that objectivism about color must be rejected (see Nida-Rümelin 2018: section 7).
- 5 Examples for this may be found in the artistic work of Escher.
- 6 Perhaps I should underline that I am not proposing a claim, which Horgan already discussed at some length in his writings. I am not just claiming that it is part of the veridicality conditions of experiences of being active that the behavior at issue is not state-caused. PO1, PO2, and PO3 do not imply that different claim without further premises and these claims are, as far as I can see, immune against Horgan’s arguments against the former (See Horgan 2007a: section 2).

- 7 As I remarked toward the end of my second installment above, in Horgan (2014) I argue that although color experience is an evolutionarily useful illusion and its content (which I call “presentational” content) is non-veridical, color-ascribing *judgments* nevertheless typically are veridical because they have a different kind of content, “judgmental” content, which is roughly a matter of an object’s being disposed to cause the relevant kind of color-experiences in normal perceivers under suitably good circumstances. Were I to embrace an error theory about the content of agentive experience, I would be inclined here too to assign a different kind of content to act-ascribing judgments—so that such judgments typically would turn out to be veridical themselves despite the non-veridicality of agentive experience. But as I also remarked above, I find myself unable to believe that I am not an agent of the kind I experience myself to be, even though I can—and do—believe that color experience is systematically non-veridical.
- 8 Horgan acknowledges too that he bears a further burden—not addressed in this dialogue—of credibly explaining why his position concerning the satisfaction conditions of agentive experience is different from his illusionist position, in Horgan (2014), concerning the satisfaction conditions of color experience. Briefly, for him the difference is this: a phenomenally manifest satisfaction condition of color experience is that external objects instantiate certain properties—call them *presented color-properties*—whose phenomenally manifest essence is such that there is excellent scientific reason to believe that no such properties are ever actually instantiated in our world.
- 9 Preliminary work for such an explication may be found in Nida-Rümelin (2018b: section 5).

References

- Baierlé, E. (2016) *The Phenomenology of Choice*, urn:nbn:ch:rero-002-116061, <http://doc.rero.ch/record/288492>
- Chalmers, D. (2006) “Perception and the Fall from Eden,” in T.S. Gendler and J. Hawthorne (eds.), *Perceptual Experience*, Oxford: Oxford University Press, pp. 49–125.
- Davidson, D. (1963) “Actions, Reasons, and Causes,” *Journal of Philosophy* 60, 675–700.
- Horgan, T. (1989) “Mental Quausation,” *Philosophical Perspectives* 3, 47–76.
- Horgan, T. (1998) “Kim on Mental Causation and Causal Exclusion,” *Philosophical Perspectives* 12, 503–509.
- Horgan, T. (2001) “Causal Compatibilism and the Exclusion Problem,” *Theoria* 16, 95–116.
- Horgan, T. (2007a) “Agentive Phenomenal Intentionality and the Limits of Introspection,” *Psyche* 13/2, 1–27.
- Horgan, T. (2007b) “Mental Causation and the Agent-Exclusion Problem,” *Erkenntnis* 67, 183–200.
- Horgan, T. (2011) “The Phenomenology of Agency and Freedom: Lessons from Introspection and Lessons from Its Limits,” *Humana Mentis* 15, 77–97.
- Horgan, T. (2012) “Introspection about Phenomenal Consciousness: Running the Gamut from Infallibility to Impotence,” in D. Smythies and D. Stoljar (eds.), *Introspection and Consciousness*, Oxford: Oxford University Press, pp. 404–422.
- Horgan, T. (2014) “Phenomenal Intentionality and Secondary Qualities: The Quixotic Case of Color,” in B. Brogaard (ed.), *Does Perception Have Content?* Oxford: Oxford University Press, pp. 329–350.
- Horgan, T. (2015) “Injecting the Phenomenology of Agency into the Free Will Debate,” in D. Shoemaker (ed.), *Oxford Studies in Agency and Responsibility* 3, Oxford: Oxford University Press, pp. 34–61.
- Horgan, T. (2017) *Essays on Paradoxes*, Oxford: Oxford University Press.
- Horgan, T., Maslen, C. and Daly, H. (2009) “Mental Causation,” in H. Beebe, C. Hitchcock and P. Menzies (eds.), *The Oxford Handbook of Causation*, Oxford: Oxford University Press, pp. 523–553.
- Horgan, T. and Potrč, M. (2010) “The Epistemic Relevance of Morphological Content,” *Acta Analytica* 25, 155–173.
- Horgan, T., Tienson, J., and Graham, G. (2003) “The Phenomenology of First Person Agency,” in S. Walter and H.-D. Heckmann (eds.), *Physicalism and Mental Causation. The Metaphysics of Mind and Action*, Exeter: Imprint Academics, pp. 323–340.
- Horgan, T. and Timmons, M. (2011) “Introspection and the Phenomenology of Free Will: Problems and Prospects,” *Journal of Consciousness Studies* 18/1, 180–205.
- Levine, J. (2006) “Color and Color Experience: Colors as Ways of Appearing,” *Dialectica* 60/3, 269–282.
- Nida-Rümelin, M. (2006) “A Puzzle about Colors,” *Dialectica* 60/3, 321–336.
- Nida-Rümelin, M. (2007) “Doings and Subject Causation,” *Erkenntnis* 67/2, 255–272.

- Nida-Rümelin, M. (2016) "Active Animals and Human Freedom," in W. Freitag, H. Rott, H. Sturm and A. Zinke (eds.), *Von Rang und Namen. Philosophical Essays in Honour of Wolfgang Spohn*, Paderborn: Mentis Verlag, pp. 339–378.
- Nida-Rümelin, M. (2018a) "Freedom and the Phenomenology of Agency," *Erkenntnis* 83/1, 61–87.
- Nida-Rümelin, M. (2018b) "Colors and Shapes," in F. Dorsch and F. MacPherson (eds.), *Phenomenal Presence*, Oxford: Oxford University Press, pp. 77–101.
- Nida-Rümelin, M. and Suarez, J. (2009) "Reddish Green. A Challenge for Modal Claims about Phenomenal Structure," *Philosophy and Phenomenological Research* 78/2, 346–391.
- Sellars, W. (1962) "Philosophy and the Scientific Image of Man," in R. Colodny (ed.), *Science, Perception, and Reality*, New York: Humanities Press and Ridgeview, pp. 35–78.