Sonderdruck aus:

Nietzscheforschung

Jahrbuch der Nietzsche-Gesellschaft

Band 20

Wirklich. Wirklichkeit. Wirklichkeiten

Herausgegeben von Renate Reschke

ISBN: 978-3-05-005742-2

Akademie Verlag

Manuel Dries

The Feeling of Doing – Nietzsche on Agent Causation*

That we are *efficacious* beings, forces, this is our fundamental belief.
(NL 34[250], KSA, 11, 505)

Introduction

The goal of this article is to re-examine Nietzsche's critique of agent causation. I contend that we first need to understand what, according to Nietzsche, our sense of agency refers to – what it ‚is' – before we can determine in what sense it is illusory. A number of recent commentators agree that agent causation is incompatible with Nietzsche's view on agents, actions, and free will. Agent-causal theories claim that actions are not produced by antecedent events but rather directly by means of a special kind of causal relation, namely agent causation. Nietzsche seems to rule out the first-person perspective: the feeling that I am the author of a particular action is not veridical. He holds that agents are best conceived as collections of drives and affects, and actions are caused by one, or a set, of such sub-personal drives that temporarily govern the drive self. My *feeling* of agent causation, of being causally efficacious, is merely a by-product, *Begleiterscheinung*, epiphenomenal.

Nietzsche's project „to translate humanity back into nature" (BGE, KSA 5, 169) rules out strong agent causation but it does so intentionally with the long-term goal being „to gain control [*Herr werden*] over the many vain and fanciful interpretations" (ibid.) (*strong* agent causation presumably being one of them). While he considers it unrealistic to attribute to agents *super*natural, special, causal powers, a naturalistic elimination of agent causation and reduction to event-causation that would rule out intentional action might be equally ‚vain and fanciful'. That there are other possible conceptions of action and agency has been argued some time ago by Bernard Williams. There might be room for a different conception of conscious control compatible with Nietzsche's view on agents, actions, and free will. In order to gain a better sense of what such a conception might consist of, I will re-examine the phenomenology of agent causation that Nietzsche provides. I consider this a first step on the way to a fuller appreciation of Nietzsche's reinterpretation of agency.

* I would like to thank audiences at King's College (London) and Oxford, as well as the members of the Philosophical Research Colloquium based at the Cambridge Faculty of Philosophy, for helpful discussion on earlier versions of this paper. I would also like to thank referees for Nietzscheforschung, as well as Paul Katsafanas, Richard Raatzsch, Marco Brusotti, Helmut Heit, Margaret Clare Churchill Ryan and the participants of the 20th Nietzsche-Werkstatt Schulpforta for invaluable comments and criticism.

This article is divided into 7 sections. In Section 1, I discuss two recent accounts that dismiss as epiphenomenal the conscious, first-person sense of agent causation. In Section 2, I look at *Gay Science* 127's ‚mechanism of willing' and consider Katsafanas' account that, while it allows for conscious mental states to be causally efficacious, rules out the Ego as epiphenomenal. In Section 3, I will re-examine Nietzsche's analysis of the phenomenology of agent causation, in particular his analysis of our conscious, first-person sense of agency. Nietzsche claims that a core self-mechanism (that experientially underpins our false concept of having ‚free will') tracks a self-system's successful effort in overcoming resistance, and leads to our conscious awareness that we are effective agents. In Section 4, I argue that Nietzsche uses this hypothesis to explain the emergence of ‚slave morality' in GM. In Section 5, I formulate what I wish to call Nietzsche's ‚resistance axiom', which he relies on to explain our fundamental belief in ourselves as efficacious forces. In Section 6, I briefly show that Nietzsche focuses on an embodied sense of agency that arises at the level of the organism and is sustained by our embodied, higher-order cognitive functions. In Section 7, I close with two brief comments that align Nietzsche's account of our sense of agency with Albert Bandura's self-efficacy studies and Antonio Damasio, who understand the self-system as a homeostatic system that knows of itself precisely because it ‚tracks' – with the help of primoridial feelings, affects, and emotions – its embodied and situated nature.

1. The Illusion of (Strong) Agent Causation

Brian Leiter's influential reading of key passages in Nietzsche that criticise the idea of free will rules out agent causation in any traditional sense.[1] His 2010 article on Nietzsche's philosophy of action summarises Nietzsche's position as follows: „The ‚person' is an arena in which the struggle of drives (type-facts) is played out; how they play out determines what he believes, what he values, what he becomes. But, *qua* conscious self or ‚agent', the person takes no active part in the process. As Nietzsche puts it elsewhere: ‚The will to overcome an affect is, in the end, itself only the will of another, or several other, affects' (BGE 117). The will, in other words, or the experience of willing (in self-mastery), is itself the product of various unconscious drives or affects."[2]

According to Leiter, Nietzsche's ‚higher types' cannot be said to display agent causation *qua* conscious self either. Higher types are distinguished by a drive coherence that is a „fortuitous natural fact about certain persons, not an achievement of autonomous agency".[3] Leiter does not say here what ‚autonomous agency' would look like but he is very clear on what it is not. It cannot be a causal process of the kind Nietzsche himself utilises, for example, when he speaks of ‚freedom': this is merely „to cause an affective response in some readers, which might lead to a transformation of their consciousness.

---

1 Brian Leiter, *Nietzsche on Morality*, London 2002; he also defends this position, for example, in his *Nietzsche's Theory of the Will*, in: *Philosopher's Imprint*, 7 (2007), 1–15.

2 Brian Leiter, *Nietzsche*, in: *A Companion to the Philosophy of Action*, Timothy O'Connor, Constantine Sandis (eds.), Chichester, UK, Malden, MA, 2010, 534.

3 Ibid.

Journal: Nietzscheforschung (ISSN: 1869-5604) vol. 20, 2013: 235–247
(Postprint, may differ from published version, for research only)

But such a transformation is, itself, a causal process in which free choice is irrelevant, but evaluative, i. e., emotional, excitation is key".[4] So understood, Nietzsche leaves very little room for our recalcitrant feeling and concomitant belief that we are agents capable of self-regulation through intentionality, forethought, self-reactivity, and self-reflection.[5] Our first-person feeling of being a causally effective agent is at most a sense of ownership.[6] Sense of authorship is a by-product, perhaps the ‚affect of command' that accompanies the drive that takes command over a set of drives, but it is no indication of an agent's conscious actions or causal powers. Leiter's account is correct but it assumes a very strong concept of agent causation and thus sets the bar very – perhaps too – high for a conception that would allow for both consciousness and the affective system to work in tandem.

Ken Gemes' account in *Nietzsche on Freedom and Autonomy* seems *prima facie* opposed to Leiter's rejection of agency free will. He proposes what he terms a naturalist-aesthetic account that allows for free acts. The latter are possible provided one first achieves the transition from a disorderly set of drives into a „coherent, ordered, hierarchy of drives".[7] Only then is a free act possible as „the expression of the character from which it originated".[8] Gemes seems undecided in regard to the extent character is open to conscious control and shaping.[9] The difference between Gemes and Leiter is significant. For both, however,

4 Ibid., 535. Cf. Brian Leiter, *Nietzsche on Morality*, 91–101; 157 f.

5 As is, for example, proposed by social cognitive theory, for example, in Albert Bandura, *Toward a Psychology of Human Agency*, in: *Perspectives on Psychological Science*, 2006 vol. 1 no. 2, 164–180.

6 Neuroscience and cognitive science distinguish between sense of agency (‚I initiated the action') and a sense of ownership (‚It is my body'). Sense of agency and sense of ownership can come apart in involuntary action (e. g. a push) but are of particular importance for the investigation of pathological disorders like schizophrenia, Tourette's, and alien hand syndrome. See, for example, Sean Gallagher, *Self-Reference and Schizophrenia: A Cognitive Model of Immunity to Error through Misidentification*, in: Dan Zahavi (ed.), *Exploring the Self: Philosophical and Psychopathological Perspectives on Self-Experience*, Amsterdam 2000, 203–239); Dan Zahavi, *Subjectivity and Selfhood. Investigating the First-Person Perspective*, Cambridge, MA 2005.

7 Ken Gemes, *Nietzsche on Free Will, Autonomy, and the Sovereign Individual*, in: *Nietzsche on Freedom and Autonomy*, Ken Gemes, Simon May (eds.), Oxford, New York 2009, 48.

8 Ibid., 47. I cannot do justice here to John Richardson's important account of Nietzsche's ‚power biology', in: *Nietzsche's New Darwinism*, Oxford 2004. Richardson also regards consciousness as largely ineffective: „Nietzsche thinks of values as creatures not of consciousness but of our drives and habits […] the selective forces behind them will inexorably work in us, and we'll continue to value and *care* about things in the ways they've designed that we do" (101); „Our conscious ‚values', […] either turn no wheels, or are merely tools […] in support of our social habits" (128). The maximum attainable is to „‚sculpt' my valuing drives and habits into a form and unity that I can regard as beautiful" (268 f.), based on a „Darwinian view of life as bearing a deep noncognitive design" (207).

9 Ken Gemes seems somewhat undecided on this point. On the one hand he uses Martin Luther's famous statement „Here I stand, I cannot do otherwise" (43) as an example of someone capable of exercising genuine agency. On the other hand, he argues that „Some individuals, due perhaps to conscious design but more likely due to fortuitous circumstances, actively collect, order and intensify some of those disparate forces and create a new direction for them […]" (42). Goethe himself seems to assume that one's character requires also conscious self-control and formation, e. g. when he famously says that „None proves a master than by limitation / And only law can give us freedom"; or „The formation of one's character ought to be one's chief aim".

Journal: Nietzscheforschung (ISSN: 1869-5604) vol. 20, 2013: 235–247
(Postprint, may differ from published version, for research only)

Nietzsche „does not seek to explain […] what it feels like when one is freely choosing to do rather A than B" since his account „does not place the subjective I at the core of the account" and „consciousness, on this view, is to be regarded as the by-product of various drives".[10] For Gemes then, as for Brian Leiter, Nietzsche's interest lies in a third-person account. The first person – though it feels so credible and reliable – is actually extremely unreliable, and in both Gemes' and Leiter's accounts our conscious awareness of being the author of our actions is treated as an epiphenomenon. While I am sympathetic to both Leiter's elimination of conscious agency and Gemes' reconceptualisation of agency free will, I think that the first person might play a more important role in Nietzsche's arguments.

2. The ‚Mechanism of Willing' and the Efficacy of Consciousness

I contend that most accounts that rule out the first person and conscious willing are based on Nietzsche's critique not of our sense of agency and willing in general but on his rejection of a specific, libertarian interpretation of our sense of agency. Likewise, many of Nietzsche's remarks on the subject of consciousness are critical of this particular interpretation, and they are often formulated as questions: „Are not all phenomena of consciousness merely terminal phenomena, final links in a chain?" (NL 7[1], KSA 12, 247). Are they really? All of them? And can we rule out that a final link can at some point function or contribute to the starting of a new chain? Crucially, in GS 127 Nietzsche suggests that willing is nothing mysterious but actually a well-practiced mechanism that functions both affectively, through the affects of pain and pleasure, *and* through the interpreting intellect: „willing is actually such a well-practised mechanism that it almost escapes the observing eye. […] first, in order for willing to come about, a representation (*Vorstellung*) of pleasure or displeasure is needed. Secondly, that a violent stimulus (*Reiz*) is experienced as pleasure or pain is a matter of the *interpreting* intellect, which, to be sure, most of the time (*meist*) works without our being conscious of it (*uns unbewusst*); and one and the same stimulus *can* be interpreted as pleasure or pain. Thirdly, only in intellectual beings do pleasure, pain, and will exist; the vast majority of organisms have nothing like it" (GS, KSA 3, 483).

What counts as pleasure, and what counts as pain, thus depends on (a) the ‚interpreting intellect', and (b) ‚one and the same stimulus *can* be interpreted as pleasure or pain.' Nietzsche states that the mechanism of willing ‚*almost* escapes the eye' and ‚*most of the time* works without our being conscious of it'. Crucially, as the caveats I have highlighted show, the mechanism can be spotted. Moreover, Nietzsche thinks he *has* spotted it, and he thinks that it can be made more transparent than it has been in the past. So, if it is true that (1) willing depends on rewarding and punishing affects such as pleasure and pain; and (2) what counts as pleasure or pain depends on an interpreting intellect; and (3) the intellect though working mostly unconsciously can *also* work reflectively, then it might be difficult but not in principle impossible to also use this mechanism consciously.

---

[10] Ibid., 48. Maudemary Clark and David Dudrick in *The Soul of Nietzsche's Beyond Good and Evil* (Cambridge 2012) allow for a normative political order but they, too, conceive of any possible ‚political order' as a ‚dominance hierarchy' among the drives. To what extent this normative order can be said to depend also on conscious control remains difficult to determine.

Journal: Nietzscheforschung (ISSN: 1869-5604) vol. 20, 2013: 235–247
(Postprint, may differ from published version, for research only)

Nietzsche's critical positions are usually much more developed than his positive proposals. And, as I intend to show, Nietzsche spends most of his time explaining how the mechanism of willing functions almost entirely unconsciously in those ‚ascetics' who have internalised a certain Moral (with a capital) interpretation, and in whom the mechanism thus works almost entirely unconsciously within their particular, deeply embedded affective-interpretive framework. But it is precisely this affective-interpretive framework that Nietzsche targets (e. g. in BGE). Note that three sections later Nietzsche writes, using clearly person-level, reflective terminology: „*A dangerous decision:* – The Christian *decision* [my emphasis] to find the world ugly and bad has made the world ugly and bad" (ibd., 485). Might we ‚decide' otherwise?

Paul Katsafanas has successfully made the case for a very specific understanding of what kind of mental states count as conscious for Nietzsche. He argues that mental states are conscious only if they are conceptually articulated. The proper target of Nietzsche's critique is better understood as an attack on consciousness as a substantive faculty and not as an attack on „consciousness as a property of mental states".[11] Conscious mental states, Katsafanas proposes, are causally efficient because they assign a specific, conceptually articulated content to an unconscious mental state, which, once it is conceptually articulated, causally interacts with other conscious states and can even create further unconscious and conscious states. Thus, for Katsafanas, conceptualising the unconscious feeling that Nietzsche refers to as bad consciousness as ‚guilt' has a profound effect – and it would be erroneous to see such a conscious mental state as epiphenomenal simply because it also had antecedent causes.[12]

For Paul Katsafanas, epiphenomenalism is not the mark of consciousness but is only the mark of the Ego, i. e. the assumed substantive faculty that we think we are: „We can capture this position by saying, as Nietzsche occasionally does, that the Ego is ‚epiphenomenal'."[13] It is puzzling why the Ego, a contender for a conceptually articulated mental state, should, in Katsafanas' account, be a mere epiphenomenon. It seems that to conceptualise unconscious mental states as ‚my Ego' would have profound consequences. And is it not precisely because of such profound consequences that Nietzsche wants to change our self-conception? The same problem arises with regard to the will as substantive faculty. Is it not precisely because our concept of will had such profound effects that

---

11 Paul Katsafanas, *Nietzsche's Theory of Mind: Consciousness and Conceptualization,* in: *European Journal of Philosophy*, 13:1 (2005), 1–31, 12.

12 Once a mental state is conceptually articulated it is linked conceptually and inferentially to other entities in a way that could be referred to as causally efficacious. For example, the non-conceptualized feeling of anxiety might not be efficacious in the above sense. On the other hand, once this state is conceptualized as an instance of 'acrophobia', it might 'cause' an agent to act in such a way as to, for example, avoid heights. For a defence of conscious, intentional action that takes into account recent challenges from neuroscience, see, for example, Alfred R. Mele, *Effective Intentions. The Power of the Conscious Will*, Oxford 2009.

13 Paul Katsafanas, *Nietzsche's Theory of Mind*, 13. In correspondence Paul Katsafanas clarified that the Ego cannot really be epiphenomenal on his account and that two claims should be distinguished: (1) Whether our conceptualization of various mental events in terms of an Ego has causal effects. And (2) Whether there is any mental entity or faculty that is (i) a seat of awareness, (ii) independent of the drives and affects, and (iii) exerts control over these drives and affects. Nietzsche answers claim (2) negatively; but he must indeed answer (1) affirmatively.

Nietzsche wants to change our conception of willing?[14] Katsafanas' account no longer allows us to simply dismiss Ego and will as inefficacious epiphenomena. However, it does allow us to state that both are conceptualisations that Nietzsche thinks are problematic and need changing (reconceptualisation) precisely because thinking about certain mental states in this way can have profound, long-term effects. We need to learn how to think differently (*umzulernen*), Nietzsche states in D 103, in order to „finally, perhaps very late, achieve even more: to feel differently" (D, KSA 3, 92). Our conscious interpretation can affect our affects, change what they mean – punisher or reward[15] – and how they guide the way we act.[16]

In the following section, I will try to show that Nietzsche attempts to eliminate our deeply embedded *conceptualisation* of our first-person awareness as ‚free will' in order to replace it with a ‚better' conceptualisation of what our awareness as causally efficacious agents actually refers to. The goal is not to eliminate the self or agency once and for all (because it is not what we thought it was) and replace it with a bundle of drives (even though we can learn a great deal from such a reduction). The goal is to understand what Nietzsche thinks our ‚feeling of doing' refers to, and only then can we ask questions about whether selves might be able to do otherwise, and if conscious self-control might be compatible with such a view.

---

14 They need changing because these conceptions have profound effects: they make man who is already „*the* sick animal" (GM, KSA 5, 367) „even sicker" (ibd., 388).

15 In contemporary affective models of consciousness, affects are seen as effective because they function as ‚punishers' and ‚rewards'. Edmund T. Rolls, for example, suggests that an „affective state is produced by an external stimulus, with the whole process of stimulus representation, evaluation in terms of reward or punishment, and the resulting mood or affect being referred to as emotion" (*Emotion, Higher-Order Syntactic Thought, and Consciousness*, in: *Frontiers of Consciousness: Chichele Lectures*, Lawrence Davies, Martin Davies (eds.), Oxford 2008, 135). Emotions are operationally defined as „states elicited by rewards and punishers that have particular functions […] A reward is anything for which an animal (which includes humans) will work. A punisher is anything that an animal will escape from or avoid" (Edmund T. Rolls, *The Affective Neuroscience of Consciousness: Higher-Order Syntactic Thoughts, Dual Routes to Emotion and Action, and Consciousness*, in: *The Cambridge Handbook of Consciousness*, Philip D. Zelazo, Morris Moscovitch, Evan Thompson (eds.), Cambridge 2007, 831). Within this model, the „adaptive value of higher-order thoughts and thus consciousness" is considered to lie in their usefulness for correcting, non-conscious, affect-based, lower-order thought (Edmund T. Rolls, *Emotion*, 146). That Nietzsche frequently conceives of pleasure and pain in exactly this reinforcing, functional way is clear already from his early discussion in 1875 when he states, against Arthur Schopenhauer and Eugen Dühring, that „pain is only the *excess* of satisfying this lack and need. Therefore both, pleasure and pain, positive, in so far as they alleviate a lack, but pain at the same time creating a new need, demanding a decrease of the stimulus. Pleasure *demands* an increase of the stimulus, pain a decrease: therein both are negative" (NL 9[1], KSA 8, 156).

16 For evidence that Nietzsche holds that conscious beliefs, in tandem with drives and affects, can have an important effect on affects and actions, see the story he recounts of helping a man who breaks down in front of him (D, KSA 3, 114).

Journal: Nietzscheforschung (ISSN: 1869-5604) vol. 20, 2013: 235–247
(Postprint, may differ from published version, for research only)

3. The Phenomenology of Agent Causation

> What do people actually want when they want ‚knowledge'? Nothing more than this: something unfamiliar is to be traced back to something *familiar* […] For ‚what is familiar is known': on this they agree. Even the most cautious among them assume that the familiar can at least be *more easily known* than the strange; that for example sound method demands that we start from the ‚inner world', from the ‚facts of consciousness', because this world is *more familiar to us*. Error of errors! The familiar is what we are used to, and what we are used to is the most difficult to ‚know' – that is, to view as a problem, to see as strange, as distant, as ‚outside us' (…) (GS, KSA 3, 539 ff.).

In the well-known passage on knowing in GS 355, Nietzsche guards his readers against one of the most common errors: to take the reduction of something unknown to something already known as a means to knowledge, for insights, and to conclude that what is thereby known counts as knowledge. In particular, those matters which we seem to know so intimately, our ‚inner world', the ‚facts of consciousness', are the hardest to even conceive as problematic and worthy of proper study. Nietzsche does not believe that we know our inner world well. Psychology, which studies the ‚elements of consciousness', is especially hard because the latter are so familiar. For Nietzsche, a distancing from our all-too-familiar understanding of ourselves is in order.

Why do we, despite evidence to the contrary coming, for example, from the neurosciences, continue to believe in ourselves as causally efficacious? Why is it that we so often *feel* efficacious when we (allegedly) are not? Nietzsche puts forward an explanation for our recalcitrance in giving up this self-conception.[17] This passage is found in an 1885 notebook in which Nietzsche provides a description of the phenomenology of agent causation and offers an explanation for the obstinacy of our belief in ‚free will'. The passage is interesting and significant because Nietzsche went back to revise it. In its earliest version it read as follows: „When we encounter a resistance and have to give in, we feel *unfree*, when we do not give in, *free*. It is this *feeling of our increase of force*, which we name ‚freedom of the will': our force, which *compels*, against a force that is compelled" (KGW IX:1, N VII 1, 1st version of 34[250]). In this version, Nietzsche is silent on whether ‚feeling unfree or free' and the ‚feeling of an increase in force' are conscious first-person mental states. However, when he later revises the passage he explicitly addresses this issue and adds that our concept of ‚free will' is rooted in our *conscious* awareness of effort or, better put, *conscious awareness of self-efficacy within a resistance relationship*.[18] With his revisions, the passage reads: „When we encounter a resistance

[17] One of his earliest notes on the subject hints at the direction his later philosophy will take. He writes: „The pleasure in power is to be explained from the displeasure experienced a hundred times from dependency" (NL 23[63], KSA 8, 425).

[18] This is also the key to Nietzsche's later conception and use of the term ‚freedom'. In ‚*My concept of freedom*', the ‚highest types' are said to be maximally free because they seek and endure maximal resistance (TI, KSA 6, 139 f.). On free will and Nietzsche's later use of the concept of freedom, see my *Freedom, Resistance, Agency* in: *Nietzsche on Mind and Nature*, Peter Kail, Manuel Dries (eds.), Oxford forthcoming.

and have to give in, we feel *unfree*, when we do not give in *but compel it to give in to us*, *free*. *I. e.,* it is this *feeling of our increase of force*, which we name ‚freedom of the will': *the conscious awareness that* our force *compels*, *in relation to* a force that is compelled" (ibd., 2nd version of 34[250]).[19]

Nietzsche's analysis of the feeling of doing, the phenomenology of agency, suggests that our deeply embedded belief in ourselves as efficacious agents originates in resistance relationships of which a self-system is consciously aware – from the immediate awareness of self-efficacy simultaneously inferred with a resistance relationship. While it might have nothing to do with any substantive faculty, this awareness, I take Nietzsche to claim here, nevertheless provides first-person access to, and feedback on, the relational status of a self-system.

Nietzsche forms a kind of explanatory hypothesis about self-systems, which is that our sense of ownership and agency expresses (we might even say with Antonio Damasio ‚is tracking') a self-system's overall resistance relationships, i. e. its levels of efficacy as an embedded agent, its relative strength and weakness in any given situation. And while this sense of agency might lead to the false belief in some fictitious Cartesian pilot, a soul substance, as evidence for a ‚doer behind our deeds', Nietzsche nevertheless sees this sense of agency not simply as epiphenomenal but also as intrinsically motivating, functioning like a standing desire. In other words, self-systems need to feel themselves as efficacious, which is why they are motivated to seek resistances, and do so deliberately.[20]

4. Restoring Self-Efficacy Through Resistance

In GM, Nietzsche uses this hypothesis to explain the emergence of what he calls pejoratively ‚slave morality'. Early group or society formation, Nietzsche argues, prevents members from discharging their drives in the way they used to (and need to in order to maintain their core sense of self as efficacious agents). As a consequence, these agents suffer from a lack of external resistance: „Lacking external enemies and obstacles, and forced into oppressive narrowness and conformity of custom" motivates and leads to an

---

19 I use the English ‚compel' in my translation of the German ‚zwingen', as the German describes a phenomenology that refers to psychological force, authority, and control (e. g. zwingende Gründe/compelling reasons) rather than any act of forcing in a physical sense (although that connotation is there also).

20 As Reginster has recently shown, Nietzsche's concept of ‚will to power' depends on resistance and the „deliberate seeking" of resistance of many forms, but not on domination or overpowering (Bernard Reginster, *The Will to Power and the Ethics of Creativity*, in: *Nietzsche and Morality*, Brian Leiter, Neil Sinhababu (eds.), Oxford 2007, 36). Galen Strawson denies that self-experience must necessarily involve experience of agency, but he agrees, quoting William James, that it involves experience of activity, that „the ‚central active self'; ‚the very core and nucleus of our self […] is the sense of activity which certain inner states possess'" (Galen Strawson, *Selves*, Oxford 2009, 93). For Strawson, „human minds are powerfully governed by deep, natural, non-agentive principles of operation", and sense of agency is „not a ‚transcendental' or constitutive condition of self-experience". Nevertheless, „choice or decision is […] no less one's own for occurring outside consciousness (it is certainly no one else's). It flows from oneself, from one's character and outlook, from what one is, mentally" (ibid., 196); and, like Gemes, Strawson does not want to rule out that „the occurrence of our thoughts and choices can be partly caused by genuinely intentional mental actions on our part" (ibid., 197).

internalisation of man: „All instincts which are not discharged outwardly *turn inwards*" (GM, KSA 5, 322). The absence of resistance scenarios that, according to Nietzsche's hypothesis, sustain the sense of self-efficacious agency, motivate the „desire to give form to *oneself* as a piece of difficult, *resisting*, suffering matter" (ibd., 326; my emphasis). By turning themselves into resistance scenarios then, human beings restored their sense of self-efficacy. The feeling of low self-efficacy (unfreedom), of being constrained by society without being able to do anything about it, motivates actions to restore the latter. Ultimately, this ‚turn' – this ‚internalisation' – leads to a new interpretive-affective framework, which includes a conscious conceptual framework (that posits an Ego that can command, be blamed, and be held responsible) that ‚interpreted' which affects (punishers and rewards) were to be associated with which drives: „the drives are transformed into demons, against which there is strife, etc."[21] Within this conscious, conceptual framework self-efficacy beliefs greatly increased. After all, the sense of self now referred to substantive souls that were so efficacious that they resisted and outlasted even their physical deaths. Furthermore, it also gave them ‚their will': they could now rely on this interpretive-affective framework to achieve self-control, for example, by associating, in a conscious mental simulation, the desire for „a murder for revenge" with „everlasting punishment in hell" (D, KSA 3, 97). Crucially, or rather cruelly, the very same interpretive-affective framework lent itself to all kinds of justificatory reasoning.

5. The Axiom of Resistance

Nietzsche assumes a kind of experiential feedback mechanism that fixes the meaning of our sense of agency. Those who suffered most from heteronomy not only turned themselves into resistance scenarios (restoring their self-efficacy and sense of agency through ascetic self-mastery) but later misappropriated it even further. They devised a language game in which they endowed themselves (in opposition to their oppressors) with ‚selves', ‚free will', ‚absolute freedom', and ‚absolute responsibility'. Based on his phenomenological analysis, Nietzsche posits what I call his ‚resistance axiom' of agency: (RA) The degree to which a self-system is aware of itself as an efficacious agent and author of its actions is a function of a self-system's efficacious effort in its resistance relationships. Awareness of inefficacy motivates self-systems to restore their sense of agency.

The self-system feels itself to be an efficacious agent when it is engaged in resistance relationships with which it can cope. The self-system feels inefficacious, is aware of other-agency, when it is engaged in resistance relationships with which it cannot cope.

I cannot here go into contemporary discussions of these issues. All I wish to indicate is that Nietzsche's hypothesis of a relationship between effort and the feeling of agent causation accords with results recently presented by Tim Bayne and Neil Levy: „Although the experience of authorship is […] a central component of the phenomenology of agency

[21] „The internalisation occurs when powerful drives, that are denied external discharge with the installation of peace and society, seek internal compensation, with the assistance of the imagination. The need for enmity, cruelty, revenge, violence turns backward, ‚withdraws'; in the pursuit of knowledge is avarice and conquest; in the artist the withdrawn power of the imagination and lying appears; the drives are transformed into demons, against which there is strife, etc." (NL 8[4], KSA 12, 335).

generally, it appears to be particularly vivid in experiences of effort. […] If the experience of authorship ever takes the experience of agent causation, we suggest that it is in contexts in which the experience of effort is particularly vivid."[22]

Nietzsche's reconceptualisation of our feeling of agent causation is not really a reduction or elimination. What has to go is the strong sense of agent causation and it has to go not because it is a mere fiction or epiphenomenal but because it is a conceptualisation that is, though initially restorative of self-efficacy, ultimately harmful.[23] He hypothesises that: (1) the sense of self as author of action is a function of a self-system's ability to cope with resistance; (2) the more effort and resistance a self-system can master, the higher its degree of self-efficacy and sense of agent causation; and (3) if a self-system becomes aware of other-efficacy it is motivated to action that restores self-efficacy.

6. The Embodied Sense of Agency

Nietzsche speculates that conscious awareness of efficacy, our sense of self, is very old, evolutionarily speaking, and similarly deeply embedded within the organism. His naturalistic understanding of self-systems as sense-making (‚interpreting') biological organisms applies the axiom of resistance to the organic level.[24] „In the entire organism", Nietzsche writes, „there is constantly the overcoming of innumerable resistances/inhibitions". This is where a sense of agency first emerges: „because we live in a state of innumerable individual pleasurable incitations, this expresses itself in the feeling of well-being of the entire person" (NL 9[1], KSA 8, 156). When Nietzsche hypothesises that resistance relationships and awareness thereof exist already on the organic level, he describes what cognitive science today refers to as proprioception, interoception, and exteroception, our unconscious and non-conceptual conscious awareness of the states of our internal organs and muscles, sensitivity to stimuli originating inside and outside the body, which he believes result in a proto-conscious, pre-reflective sense of agency: „the sense of well-being as the feeling of *power* [better: *efficacy*] triggered by little obstacles: because in the entire organism there is constantly the overcoming of innumerable resistances/inhibitions, – this *victorious* feeling becomes conscious as the *overall feeling* [*Gesammtgefühl*] of gaiety, ‚freedom'" (NL 5[50], KSA 12, 204).[25]

---

22 Timothy J. Bayne, Neil Levy, *The Feeling of Doing: Deconstructing the Phenomenology of Agency*, in: *Disorders of Volition*, Natalie Sebanz, Wolfgang Prinz (eds.), Cambridge, MA 2006, 63.

23 Nietzsche argues that this interpretive-affective framework is ultimately nihilistic, i. e. it undermines itself and leads to disorientation, despair, and a total breakdown of the capacity to ‚will', to set oneself goals, and to maintain an affirmative attitude towards life.

24 Shortly after his remark on freedom he explains both experience and memory as successful resisting and dealing with complexity: „Enabling of *experience*, by simplifying real events both on the side of the impacting forces as well as on the side of our creative forces" (N VII 1, 3). See also: „Experience is only possible with the help of memory: memory is only possible by means of an abbreviation of a mental event into a *sign*" (ibd., 6). On signs, see Günter Abel, *Zeichen der Wirklichkeit*, Frankfurt/M. 2004.

25 See, for example, neurobiologist Antonio Damasio on empirical evidence for the hypothesis of a ‚protoself' consisting of primordial feelings that „reflect the current state of the body along varied

It comes as no surprise that sentient self-systems cannot free themselves from the assumption of agent causation, our „Freiheitsgefühl" (ibd.) that is confirmed by our overall sense of well-being, and which tracks the proper functioning and self-regulation of an embedded organism's resistance relationships.[26]

It is not only on the level of the functioning organism that the sentient self-system engages in resistance activities. Nietzsche uses the resistance axiom to explain also the more complex, higher-order cognitive functions of self-systems. He sees the self-system's intentional sense-making ability – i. e. to see something *as* something – as confirmation that it is always dealing with and seeking resistances. Even the higher cognitive activities, such as naming, thinking, and imagining, he explains, again applying RA, are resistance activities motivated to sustain the self-system's feeling of self-efficacy. As long as a self-system is able to do so successfully, it feels itself as causally efficacious.[27] While Nietzsche rules out soul substances and homunculi, he nevertheless assumes something like a core self-mechanism: resistance and resisting activity and awareness leads to our *Grundglaube*, „our basic belief", or „fundamentum": „[t]hat we are *efficacious* beings, forces" (NL 34[250], KSA, 11, 505).

7. Nietzsche's Phenomenology of Agency between Bandura and Damasio

I want to close with two brief remarks, the first on self-efficacy and the second on homeostasis.[28] First, a large body of empirical research today backs up Nietzsche's acute

dimensions, for example, along the scale that ranges from pleasure to pain" (*Self Comes to Mind. Constructing the Conscious Brain*, New York 2010, 22, 272 ff.).

[26] See also: „before ‚thinking' began, there must have been already a ‚composing', the shaping faculty is more original than that of ‚thinking'" (NL 40[17], KSA 11, 636).

[27] Nietzsche does not claim that self-systems engage in higher-order cognitive functions only as a means to sustaining a sense of self-efficacy. He indeed recognizes many other motives that incline toward these activities. (I would like to thank Paul Katsafanas for this point.) Nevertheless, he seems to hold that RA plays an important role also in higher-order cognitive functions, thus contributing to the sense of self-efficacy.

[28] There is an important and, to my knowledge, underexplored connection that links Nietzsche's sense of agency to a tradition within philosophy of mind that starts with Fichte. Fichte conceived of the self no longer as any ‚substance' but rather as an activity issuing from nothing other than resistance. „The I posits itself as determined by the not-I" is Fichte's formula for, as Wayne Martin has recently shown, his „fundamental law of consciousness", which is based precisely on a „principle of resistance". Experience of self and of an independent world depends on the encounter of resistance within experience: „The conscious experience of oneself as having aims and of the world as resistant are inextricably interlinked" (Wayne M. Martin, *Fichte's Transcendental Phenomenology of Agency*, in: *Fichte. System der Sittenlehre*, Jean-Christophe Merle, Andreas Schmidt (eds.), Berlin forthcoming). Fichte's transcendental account is commonly misunderstood as reducing everything to the self and Nietzsche's naturalistic account is misunderstood as eliminating the self. There is a long-standing tradition in the philosophy of mind that recognises the special status of self-awareness. On the contemporary relevance of Fichte's ‚fundamental insight' that self-consciousness cannot be explained through a reflective relation whereby a subject ‚grasps' itself *as* an object but that instead there must be a pre-reflective sense of self, see Manfred Frank, *Non-objectal Subjectivity*,

sense of the importance of unconscious and conscious awareness of self-efficacy, as well as unconscious and conscious self-efficacy beliefs. As Albert Bandura and others have shown, self-efficacy depends on „inferences from somatic and emotional states of personal strength and vulnerabilities", on „mastery experiences" and changes through social persuasion. Self-efficacy levels are extremely important and have been proven, in numerous experiments, to affect „level of motivation, quality of functioning, resilience to adversity and vulnerability to stress and depression".[29] Nietzsche assumes all of the above and sees this as a prime motivator for action.[30]

Second, Nietzsche's model of the self as it emerges from his examination of the phenomenology of agency is continuous with results and models used in contemporary neurobiology and neuroscience. For Antonio Damasio, for example, the ‚self comes to mind' precisely because self-systems are constantly engaged in homeostatic self-regulation at different levels of which they are aware. Selves understood as such homeostatic systems are, he speculates, composed of ‚primordial feelings' and a ‚protoself' that is mapping internal and external engagements. In order, for example, to create a map of an object „the brain must adjust the body in a suitable way" and it is these changes in the protoself that „inaugurate the momentary creation of the core self"[31], a kind of ‚protagonist': „the self comes to mind in the form of images, relentlessly telling a story of such engagements."[32] Like Nietzsche, he also speculates about the natural basis of certain values: „What we have come to designate as valuable, in terms of goods or actions, is directly or indirectly related to the possibility of maintaining a homeostatic range in the interior of the organism […] The diagnosis requires no special expertise but merely the fundamental process of consciousness: optimal ranges express themselves in the conscious mind as pleasurable feelings; dangerous ranges, as not-so-pleasant or even painful feelings."[33] There is, however, an important difference. While he would agree with Damasio that moral rules „responded to the detection of imbalances caused by social behaviour that endangered individuals and the group",[34] Nietzsche is not just describing but criticising the specific, ascetic nature of the ‚homeostatic equilibrium' thereby achieved.

---

in: *Journal of Consciousness Studies* 14, 2007, 152–173. The phenomenological tradition assumes, as for example Dan Zahavi argues, that „pre-reflective self-awareness and a minimal sense of self are integral parts of our experiential life" (*Subjectivity and Selfhood. Investigating the First-Person Perspective*, Cambridge, MA 2005, 146). Nietzsche would be opposed to much this tradition has helped itself to through this pre-reflective sense of self, but his naturalistic, embodied conception nevertheless belongs to this tradition.

29 Albert Bandura, *Self-efficacy*, in: Vilayanur S. Ramachaudran (ed.), *Encyclopedia of Human Behavior* (vol. 4), New York 1994, 81.

30 On self-efficacy and motivation, see Albert Bandura, *Self-Regulation and Motivation Through Anticipatory and Self-Reactive Mechanisms*, in: Richard A. Dienstbier (ed.), *Nebraska Symposium on Motivation: Perspectives on Motivation*, vol. 38 (pp. 69–164). Lincoln 1991.

31 Antonio Damasio, *Self Comes to Mind*, 215.

32 Ibid., 216.

33 Ibid., 59.

34 Ibid., 310.

Conclusion

I began with the question of whether agent causation is illusory for Nietzsche, and in what sense, if any, conscious self-control and ‚willing' might be possible for the kinds of selves Nietzsche thinks we are. I briefly explored the sense in which ‚slave morality' relied, unreflectively, on what Nietzsche terms the ‚mechanism of willing' (GS 127) and how, in mental simulations, these ‚slaves' were able to use their interpretive-affective framework to practice self-control.[35] I then tried to show how Nietzsche understands our ‚feeling of doing', i. e. what selves are (and are not). The embodied self works for Nietzsche as a homeostatic feedback control system.[36] When the self-system is in (feels) an overall state (*Gesamtgefühl)* of low efficacy, for example, due to lack of essential nutrients (or due to conditions of heteronomy), a drive expresses this negative system state either unconsciously or consciously. For example, the painful feeling of hunger (or resentment) expresses low self-efficacy and puts the self-system in a negative state or a state of tension that motivates (through ‚reinforcing' affects) sense-making and agency. Once a drive-induced affective orientation[37] leads to the satisfying of the drive, a kind of equilibrium or homeostasis is reached until the next need arises and puts the system back into disequilibrium. In Nietzsche's model, awareness of low efficacy motivates the bundle of drives (including the drive to invent ideals in the case of the ‚slave') to coordinate in such a way as to increase the self-system's sense of agency. It is Nietzsche's analysis of the phenomenology of agency that underpins his hypothesis that human beings function by means of an evolved „*instinct of freedom*" (GM, KSA 5, 326).[38]

Brian Leiter and Ken Gemes are right: the kind of agent causation and conscious ‚I' that have to do with any mysterious property that enables autonomous agents to consciously initiate causal chains *ex nihilo* do not exist. Yet, as conceptually articulated conscious mental states they are effective – and this is why the conceptualisations of our first-person experience need changing. I hope to have shown here why Nietzsche thinks that belief in agent causation is so tenacious, what function he attributes to the first-person awareness of self-efficacy, and in what sense, and by which ‚mechanism', the ascetic exercises their will. The question I end with, then, is: might there be a model of the will that utilises the ‚mechanism of willing' both reflectively and non-ascetically – effectively?

---

35 I explore this further in my forthcoming *Nietzsche's Simulationist Model of the Will*.

36 Put crudely, it can be pictured like a thermostat.

37 On drives and affective orientations, see Paul Katsafanas, *Value, Affect, Drive*, in: Peter Kail, Manuel Dries, *Nietzsche on Mind and Nature*, Oxford forthcoming.

38 Nietzsche's choice of terminology is perhaps unfortunate – chosen for affective effect – but ‚will to power' means not much more than the unconscious and conscious awareness of the embedded, sentient self-system's self-efficacy status that motivates orientation and action through pro- and con-affects (within an interpretative framework): „eben jener *Instinkt der Freiheit* (in meiner Sprache geredet: der Wille zur Macht)" (GM, KSA 5, 326).

Journal: Nietzscheforschung (ISSN: 1869-5604) vol. 20, 2013: 235–247
(Postprint, may differ from published version, for research only)

Journal: Nietzscheforschung (ISSN: 1869-5604) vol. 20, 2013: 235–247
(Postprint, may differ from published version, for research only)