Troubles with integrationist theories

The general idea behind integrationist theories is one of harmony between motivational states of different levels: »It is in securing the conformity of his will to his second order volitions, that a person exercises freedom of the will. And it is in the discrepancy between his will and his second order volitions, or in his awareness that their coincidence is not his own doing but only a happy chance, that a person who does not have this freedom feels its lack, »(H. G. Frankfurt: Freedom of the Will and the Concept of a Person, reprinted in G. Watson: Free Will, OUP, 1982. p. 90.).

One can discern three components which enter this kind of theory:

1. *locus of control component:* The real source of freedom, and the focus of identification of the agent are his higher order motivational states.

2. *conformity component:* Freedom consists in the counterfactually secured conformity between higher and lower order motivational states.

3. *structure component:* It is the structural properties of the motivational system of a person which alone count in determining whether the person is free or not.

Each of these component is highly problematic.

1. It has often been stressed that there is no special reason to assign locus of control to higher order motivational states. I would only add that some psychologists think it is much easier to manipulate people's higher order desires (linked with selfesteem, with features of selfconcept and the like), than the down-to-earth first order ones. (I owe this point to Viktor Božičević). If the integrationist postulates indefinitely high hierarchies of motivational states — the way K. Lehrer does in his recent writings — his theory looses psyhological plausibility.

2. It is doubtful that on the folk-psychological level integration and conformity have much to do with freedom. Take two characters — Jack Sober and Jim Wilde. Jack Sober is punctual, somewhat autoritarian and conformistic. He has a very strong ability of self-control, and he never feels remorse or even regret. If he had a different ideal of his self, he would be strong-willed enough to change his first order preferences. Being as he is, he is a somewhat boring and unattractive fellow. Jim Wilde is spontaneous and nonconformistic. He often does things on impulse which he later often regrets. In fact, he would like to be like Jack Sober, but he could never bring himself to give up his spontaneous behaviour. Now, the integrationist theory would make us pronounce Jack Sober a (reasonably) free man, and Jim Wilde an unfree creature. The intuitons of many people would decide to the contrary, and some people would say that the evidence is not sufficient. I do not

know anybody who would take Jack Sober as a paradigm of a free man, as opposed to Jim Wilde. Freedom just does not seem to consist in conformity between higher and lower order motivational states.

3. It seems that some ways of »motivating« people are incompatible with freedom, no matter how well they fit in the structure.

Take the following fictional (counter-) example:

Orestes is a hesitant, Hamlet-like character, who is basically good natured and meek. He has some wild fancies and second order desires: he would like to be cruel and barbarous, especially as he notices that girls in his native Argos admire local hoodloms. Of course, he cannot bring himself to form adequate first order desires, so his actions are in fact kind and gentle. Enters Pilades. He is Orestes friend, and he has telepathic powers over him. So he decides to help him. As soon as Orestes forms a second order desire, say, to be able to beat an innocent bystander, Pilades instills in him the corresponding first order desire, upon which Orestes then acts. Had Orestes had a different second order desire, Pilades would provide him with the corresponding firstorder volition. So, on the given occasion, Orestes beats up an innocent person. Is he responsible for his deed? Not exactly. He was not really free.

If the example is succesful — and I hope it is — it shows that some ways of motivating people as a rule preclude freedom.

These considerations might perhaps lead us to the right track. It seems that a central deficiency of integrationist hypothesis lies in its structuralist bias. An integrationist sees the property of being free as a properties at a certain time. So his a structuralist current — time slice principle governed hypothesis (Current time-slice principle, a term borrowed from Nozick says that freedom-related properties of a motivational state are determined by how the state is inserted in some wider motivational structure). The hypnosis example shows that certain states are un-free no matter how well they fit into one's motivational structure. They carry their stigma not in virtue of their structural relationships at a given time, but in virtue of their *etiology.*

So structure is not enough, and history counts.

To start with the question of necessary conditions for a state being free, it seems that in order to qualify at all, the state must not have a stigma-conferring history. A motivational state induced by brain washing, hypnosis or a phobia simply does not have an appropriate history.

(There is an analogy here with historical or causal theories in other domains. A structuralist theory of X (denotation, knowledge, justice) says that an item I has the property X in virtue of its structural builtup, regardless of the aetiology. A historical-causal theory claims at least that having a right kind of history is a necessary condition for

I to have X: in order for a name to denote Moses it has to have been linked to the man Moses in the right fashion, in order for a distribution to be just it is necessary that it originates in a right way from a certain state, etc.).

It seems that the fact that other-induced hypnosis automatically disqualifies the produced state clearly shows that having a right kind of history (or, not having a wrong kind of history) is a necessary precondition of a state being free.

This is in itself quite important, because it entails that *no purely structuralist theory* shall account for freedom. Some history is certainly needed.

Can this account be extended and strengthened? Can one frame sufficient conditions for a motivational state being free in terms of history or aetiology of the state?

Given the wide range of unconstrained motivational states (wanting to drink tea just because one likes it, wishing to go to the concert because this is what everyone in the neighbourhood does, and so on in indefinitum) it seems that one cannot. There are so many causal histories, so many spurious kinds, and so little solid knowledge about them!

However, it might be possible to make a crude and very optimistic skech if we assume a more thoroughly naturalistic stance.

Suppose that being free is not an outlandish or heavily idealistic possibility for motivational states. Suppose that human behavior in this sense is unconstrained in natural or at least in quite favourable circumstances. Some people might agree to this supposition (Aristotle certainly would, for one). Then we could turn the tables, and somehow identify freedom with naturalness, as in the following thesis:

(F) a motivational states is free if it has been generated in a natural way.

Thesis (F), vague and ambiguous as it is, squares well with two sets of facts: first, facts concerning the inappropriateness of hypnosis, phobia and the rest to generate free motivational states, second, facts concerning usual or normal motivation (from hunger to love) which, in spite of their strength we do not deem usuitable for producing free motivational states.

However, there is a problem with the epithet »natural«: malfunctioning is as »natural« as functioning, disease as »natural« as health, at least by one reading of »natural«. It is therefore better to revert to overtly teleological notions of normal or proper functioning (in line with authors like A. Goldmann, R. Milikan and D. Dennett).

With this in mind we propose the following thesis:

(F^+) A motivational state is free if it is produced by normal motivational processes.

It will be useful to stretch the analogy with the cognitive realm, and to distinguish the normal sources of motivation from the normal transmissions of motivations. An analogous distinction in the cognitive realm has been proposed by A. Goldmann.

One set of plausible candidates for motivation sources would certainly be needs. All needs or only some of them?

There has been a lot of discussion about which needs are somehow »legitimate« and it might turn out that not all are. However, at least some subset of actual needs of people could certainly serve as a paradigm example of what is a normal motivational source.

How about motivational state transitions?

The most important candidate for the normal transition process is practical reasoning. Suppose that Peter wants to see a movie, and that he also knows that the only way to see the movie is to buy a ticket. If he then forms the intention to buy the ticket, by usual reasoning, then his motivation has been transfered from the goal to the means in the appropriate way.

One virtue of this cursory account is that it makes sense of the idea that most of our everyday activities which are by commonsense standards unconstrained are as a matter of fact free.

For instance, when I take my coffee in the morning, I do this unconstrained, although I would be quite unhappy were I deprived of the coffee. My taking is highly predictable, not only for a laplacean scientist, but for my family and friends as well. I am not sure about my higher — order preferences. If I prefered to prefer to have a tea instead then... Well, I do not know what would have happened then. Maybe I would not have prefered tea. This does not make my drinking coffee unfree. My preference for the coffee has arisen out of normal needs, has become a matter of habit in an unconstrained process of habituation, and it is not immune to revision in light of other considerations: if coffee suddenly became ten times as expensive as it now is, I would give it up.

The causal account also makes sense of the appearing idea that one can be ufree without realising it, and this without invoking complicated counterfactuals about high-order preferences. This is how it should be. Suppose Peter's psychiatrist tells him that his greed for sweets is a compulsive drive, stemming from certain events from his childhood. No distant counterfactuals enter the scene — the nature of Peter's motivational state is established by appeal to its aetiology, and that is the end of it.

It might seem, however, that this account is unable to do justice to the more dramatic aspects of moral life to a person's struggle to achieve what she thinks is a worthy and valuable life, to live up to her own expectations and to similar morally relevant phenomena.

In order to make the picture more realistic it is useful to introduce some additional considerations about the way motivational states are generated.

It is of interest for any organism, and so also for human beings, to have a reasonably coherent set of motivations and preferences, smply because an incoherent set is impossible to satisfy. On the other hand new experience provide new motivations. Coherence is thus frustration-preventing. Now, one interesting way of mainting a balance between coherence and dynamic changes, is to have one's first order states revised in the light of overall second order preferences. Being able to »step back« and consider one's first order motives at a given time is essential for maintaining coherence and overall balance. There is an analogy here with cognitive processes. It is very important for an information processing system to be able to resist first order pressures (The pressure to conclude from: This stick looks bent to This stick is bent), and to be able to revise its opinions in the light of second order epistemic or methodological principles.

It is therefore to be expected that it will be part and parcel of the proper functioning of the human motivational system that it will allow or even press for second-order motivation. The ability to resist first order »temptation«, and to judge first order motivational states in the light of one's global self-concept is certainly essential for the proper functioning of the human mind.

This is then the kernel of truth in the integrationist's hypothesis. The integrationist sees the control over one's first order motivational states as being fundamental for freedom. He is aware of the fact that higher-order preferences can be manipulated as easily, if not more easily, than the first-order ones, and he consequently tries to push the locus of control as high up as possible. This ascent of control is quite dramatic — it seems that it is always at the next higher level that we shall encounter the true self, the very heart of our being. The hope is illusory — higher order preferences do not guarantee freedom. But the integrationist has glimpsed something important, namely the fact that a capacity for taking a distanced view of oneself, and of ones present motives does play an essential role in the way human motivational states are generated.

The sketch of a causal-teleological account of freedom that is here proposed is very crude indeed. The account is in some ways parallel to the classical naturalistic accounts of motivation and freedom, most closely to those of Aristotle and Spinoza. It needs a lot of further work, fleshing out the details, as well as providing answers to some central questions like What makes a motivation — creating process normal or proper? But this we leave for some further occasion.

The integrationist strikes back

Confronted with the causal ploy, the integrationist is, of course, far from being defenceless. He might offer some persuasive counter-arguments. I shall adress two of them.

a) The agent and her twin

The first argument has been proposed by Keith Lehrer, in discussion, but it goes back to his paper in »Action Theory«.

Suppose that a twin universe, an exact copy of our universe, is created at this very moment. Each of us has an exact replica in that universe: there is a Twin Me, and a Twin You. Now, there is a strong intuition that a Twin You is free only in case X you are free, and is unfree, only in case you are unfree. However, your twin has not the same history you have, because he has none, and so his motivational states do not have the same aetiology as yours do.

Let us call being free and being unfree »freedom values« of motivational states. Let ustake the expression »states A covaries with the state B« to mean that they are equivalent with respect to freedom values: A covaries with B = Def A is free iff B is free.

The upshot of the argument, then, seems to be this: it cannot be the case that being free is dependent on causal history, because here we have a case in which radically different causal histories do not count, and the only thing that does count is the similarity between you and Twin You at a given moment.

Notice that the story is not exact counterexample to the causal theory. Nowhere in the theory is it being claimed that two divergent causal histories cannot result in both final states being free (or both unfree). It does not even show that we typically disregard aetiology. It might be that in the case of normal history we go by causal eatiology, but in the case of sudden creation we help ourselves to some subsidiary criterion, eg. similarity to aetiologically clear cases. So the story is not fatal to the causal theory.

But there is more to it.

The story is as much dangerous to the integrationist as it is to the causalist. Suppose that you have just freely chosen coffee over tea and your twin is created at this very moment. By the top-down integrationist story at least the following is true of you: if you have preferred to choose tea instead of coffee (if you had a different second order preferences, he would have had different first order preferences (because he would then simply not have existed). So, by integrationists standards, he is not free. So, you are free, and your twin is not, which contradicts the intuitive assumption. So, if we stay by the intuition, the integrationist's account is false.

This is of course an *ad hominem* argument. If it is valid, it shows that the integrationist cannot use the Twin Universe example againtst

the causalist. But someone else could (say, Peter van Inwagen). So, the story is not yet over.

There is another way out for the causalist, namely to deny the relevance of this kind of thought experiments on the ground that they violate the natural assumptions about causality which in fact govern our imaginative power. This way has been suggested to me by P. Bieri, and it has been argued for in different context by K. Wilkes. I would accept it only if I had no choice, because it deprives the philosopher of the most exciting ways to argue and to test his intuitions.

Instead of this, I would question the intuition that states and twin states covary in respect to freedomvalue. Suppose that Bloody Mary has been planning yet another murder. She hates the victim, she has set the trap, and she only has to strike. The minute before the stroke she reflects for a second: should I kill or not? But the hatred prevails. She kills, we would say, freely.

Her Twin is created at the very moment of her reflection. So, the Twin Mary also takes a pause, asks herself the question, and answers in the positive: Yes, I shall kill.

Now is Twin Mary in the same way responsible for her act in which Mary is? Is she not rather a helpless creature, created to the likeness of a criminal, but not fully responsible for her choice? I think there is a great difference in degree of responsibility.

It is natural, then to suppose that there is a difference as to the freedom. Twin Mary's states are, so to say, imposed upon her, she did not do anything to arrive at them. Her final decision has been in a way forced upon her, and in the way Mary's decision was not.

I admit that intuition is here not as firm as it should perhaps be. This is natural with such kind of science-fiction story. At least it is not perfectly obvious that freedom values do covary, and it might be the case that they do not. But, this is all the causalist needs. Different aetiologies yield different freedom-values.

Let me resume:

The Twin Universe argument starts from the assumption that freedom-values of motivational states covary across universes regardles of aetiologiess of those states.

First, the argument, even if correct, would render the causal theory only implausible, it would not show it to be false.

Second, the argument is not available to the integrationist, because it yields negative results for integratioism — if the argument is valid, integrationism is false (at least the top — down version).

Third, and most important, the argument is flawed. It rests on the wrong impression, that our inuition would always confirm the hypothesis of covariation, that it would always deliver the verdict that X is free iff Twin X is free. This is incorect, and so the argument fails.

NENAD MIŠČEVIĆ: NATURALIZIRANJE SLOBODNE VOLJE — PRVI KORACI

Sažetak

U članku se razmatraju kompatibilističke teorije slobodne volje. Podrobno se analizira hijerarhijska teorija i iznose se kritičke primjedbe. U drugom dijelu članka iznosi se skica prijedloga za alternativnu kompatibilističku teoriju koja tvrdi da je radnja slobodna ako je motivirana normalnim motivacijskim sustavom.