

Meeting 7: Chapter 4, 1st Half (Nozick on Free Will and Tracking Value)

I. Overview of Chapter 4

Sometimes the topic of free will is formulated as a problem about punishment and responsibility:

How can we punish someone (or hold a person responsible) for an action that was causally determined by factors originating before the person's birth?

And how can we punish someone (or hold a person responsible) for an action that was random, like the time at which an alpha particle is emitted by a sample of uranium 238?

For Nozick, though, the central problem is that either of these options (causal determination or randomness) seems to undercut our value and importance: rather than being the valuable originators of our actions, instead we would seem simply to be a place where things happen, whether via earlier causes or spontaneously.

In Part I, Nozick develops (and rejects) an indeterminist account of free will that would fully preserve our value as agents.

In Part II, Nozick develops a determinist account of free will on which our value as agents is diminished, but something of value still remains.

In Part III, Nozick formulates a rationale underlying punishment in retribution for a wrong.

II. An Indeterminist View of Free Will

Suppose you are deciding between acts A, B, . . . , Z. There are various reasons for and against each of these. In considering these reasons, one arrives at a view of which reasons have more weight than the others, and then acts on the basis of the weightiest reasons.

Nozick's proposal:

The competing reasons do not come with previously given precisely specified weights.

At least sometimes, during deliberation, one makes it the case that the reasons have a certain weight: "the decision process is not one of discovering . . . weights but of assigning them" (p. 294).

In these cases, one freely acts in choosing to (say) do act A on the basis of reasons R_A .

Further comments on his proposal:

Although the specific weight of a reason is "up to us," whether a given consideration is a reason at all is not "up to us."

The weights bestowed by one's decision are not exact quantities, but they suffice to make some options have more reason in their favor than others.

Moreover, these bestowed weights set a precedent for future decisions.

If person S decides to do act A for reasons R_A rather than alternative act B for reasons R_B , then Nozick claims the following will all be the case:

S "was caused to do act A by (accepting) R_A " (p. 295).

But if S had decided to do B instead, "it then would have been R_B that caused him to do B" (ibid.).

So although R_A in fact caused S to do A, S's doing A was not causally determined.

Thus S exhibits "contra-deterministic freedom" in doing A.

Are these free decisions random?

No, says Nozick, for “there are not fixed factual probabilities for each action” (p. 302).

Are these free decisions arbitrary?

Maybe not, says Nozick, if the decision is *self-subsuming*: “the weights it bestows may fix general principles that mandate not only the relevant act but also the bestowing of those (or similar) weights” (p. 300).

A self-subsuming decision will also be *reflexive*: “the weights are bestowed in virtue of weights that come into effect in the very act of bestowal” (p. 304).

But aren't there different and conflicting self-subsuming decisions that could be made?

And what about choices not made on the basis of self-subsuming decisions?

In the end, Nozick rejects this proposal because he has “worries about terming this bestowal nonarbitrary and nonrandom because it is self-subsuming and reflexive” (p. 305).

I have worries, though, about how to even understand the proposal. One key issue: what, exactly, causes person S to do act A? Is it R_A itself, or is it S's acceptance of R_A ?

- *first option*: R_A itself is the cause.

One potential worry: R_A will be considerations like the fact that my parents will be happy if I do A, or the fact that I promised her I would do A, or the fact that my friendship with her requires that I do A. But it's not clear how these sorts of facts can cause me to do A.

Another potential worry: now choices made on the basis of false beliefs can never be free. Suppose I think R_A is the case, but in fact it isn't; then R_A can't cause me to do A.

- *second option*: S's accepting R_A is the cause.

Does S's accepting R_A have causes? If not, then we have an uncaused cause somewhere in the network, so this proposal is more mysterious than it at first seemed.

But if S's accepting R_A does have causes, then it's not clear we can tell the same story over again about how S's accepting R_A has causes but is not causally determined, on pain of entering into a regress of reasons for our weighting of reasons for our weighting of . . . reasons. (Or maybe there is such a regress, and only self-subsumption can stop it?)

III. Types of Value

Nozick distinguishes between four types of value (pp. 311-13):

- *intrinsic value*: the value something has in itself, independently of what it leads to or its further consequences.
- *instrumental value*: the value something has due to its actual or expected causal consequences.
- *originative value*: the value something has due to the value (intrinsic, instrumental, or otherwise) it newly introduces into the world.
- *contributory value*: the value something has due to there being more value in the actual situation in which that thing is present than a situation in which it isn't but other factors were reorganized to minimize its absence.

Nozick insists that if causal determinism is true, then our actions lack originative value.

But I don't see why that's so. If our actions are causally determined, then maybe they can't bring any new instrumental value into the world. But what blocks them from being able to bring new originaive or contributory value into the world?

Nozick goes on, in Part II, to sketch a view on which although our causally determined actions lack originaive value, they can possess contributory value, when we act in a way that tracks the rightness of what we do.

IV. Tracking Rightness

Let us assume, for the moment, that some of our actions are better than others (an evaluative assessment), and that some of our actions are permitted whereas others are forbidden (a deontic assessment).

Nozick uses 'rightness' or 'bestness' as a placeholder for the normative property (whatever it is) we aim to track when intentionally acting.

He then proposes that just as knowledge (= tracking the truth) is a special, exalted form of belief, so too is there a special, exalted form of action that involves acting in a way that tracks the rightness of what we do.

Here is the most straightforward way of adapting Nozick's tracking theory of knowledge to the case of action:

Nozick's tracking theory for action (first pass): Person S's doing of act A tracks rightness if and only if:

1. A is right;
2. S does A;
- III. if A weren't right, S wouldn't do A;
- IV. if A were right, S would do A.

But, Nozick thinks, there is a problem (p. 707, n. 31):

If we interpret 'right' to mean 'permitted' throughout, then IV is implausible: if donating to charity X and donating to charity Y are both permitted, I might track rightness in donating to X even though there is a nearby possible world in which donating to X is still permitted but I donate to Y instead.

If we interpret 'right' to mean 'mandatory' throughout, then III is implausible: if donating to charity X is mandatory but the nearest possible world in which it isn't is one in which donating to charity X and donating to charity Y are both permitted, it doesn't seem to be a mark against me qua tracker of rightness in the actual world that I would still donate to X in that nearby world.

Nozick's fix is to interpret 'right' differently in III and IV, like so:

Nozick's tracking theory for action (second pass): Person S's doing of act A tracks rightness if and only if:

1. A is right [in what sense?];
2. S does A;
3. if A weren't permitted, S wouldn't do A;
4. if A were mandatory, S would do A.

The basic idea: when a person's action tracks rightness, it isn't an accident that the person does the right thing (just as when a person knows that p , it isn't an accident that the person truly believes that p).

The crucial thing: S's doing of A can track rightness even if it was causally determined that S would do A.

Note that when evaluating the subjunctive conditionals in 3 and 4, we don't postulate "some grand moral change in the universe" (p. 322): we don't, for example, consider possible worlds in which utilitarianism is true instead of Kantianism. Rather, we consider possible worlds in which the purely descriptive facts in virtue of which the normative facts hold are different.

Just as we can complicate the tracking theory for belief by taking into account the *method* by which a belief was formed (= the basis of one's belief), so too can we complicate the tracking theory for action by taking into account one's *motive* for acting (= the basis of one's action).

Nozick never explicitly states the modified theory once motives are taken into account. Here is my best attempt at formulating it (when we set to one side the possibility that our motives for acting are overdetermined):

Nozick's tracking theory for action (third pass): Person S's doing of act A tracks rightness if and only if there is a motive M for acting such that:

1. A is right;
2. S does A for M;
- 3'. if A weren't permitted and S were to consider, on the basis of M, whether (or not) to do A, S wouldn't do A;
- 4'. if A were mandatory and S were to consider, on the basis of M, whether (or not) to do A, S would do A.

This allows us to avoid the following counterexample to the second-pass proposal: “[S]omeone might perform a mandatory action after carefully considering its rightness, yet if it weren't right, the issue of its moral quality would never have arisen for him [and] still he would have been led to decide to do it” (p. 320).

Nozick writes, “It seem plausible to think that the only way an action can track an evaluative fact is via . . . the person's knowledge of the fact” (p. 321).

The suggestion here is that S's doing A only tracks rightness when (a) S's doing A tracks S's belief *that A is right*, (b) S's belief *that A is right* tracks the truth, and (c) this is not one of those cases where tracking fails to be transitive.

Later he adds that, in addition, it must be that (d) S has a desire to do what is right. (This leads him to consider whether to add a requirement that one desire to believe the truth to his analysis of knowledge.)

But I don't see why any of this is required.

Tracking rightness by having a belief about rightness that tracks the truth (plus a desire to do what is right) could be one way of tracking rightness, but we shouldn't rule out other ways.

So I think Nozick would be better off taking that belief (plus desire) to be one motive M that we can plug into his third-pass account of tracking rightness, without restricting the account to that motive.

Nozick points out that his account of tracking bestness allows for closure failures:

For example, *my going to the library* might track rightness, and this action might entail that I go outside sometime this decade, but *my going outside sometime this decade* might not track rightness (because I would still go outside sometime this decade if it weren't permitted).

But, Nozick laments, “though there is this sort of nonclosure there to be noticed, it has not played any deep role in the free will issue” (p. 330).

He concludes that “the notion of an action's tracking value [is] simply a nice description of something else, something we might be happy to have, but not free will” (p. 327; see also p. 332).

I am inclined to agree. The idea that tracking rightness is a way of freely acting is a bizarre suggestion. (Why would only right acts be free? Moreover, there is no temptation to think that his tracking theory of knowledge is an account of how beliefs can be free.)

What Nozick appears to have here is a very interesting, very ahead-of-its-time account of what it is to *act from moral worth* (a hot topic among philosophers in recent years).