

Rationality and Selfhood

Definition VII. That thing is called free, which exists solely by the necessity of its own nature, and of which the action is determined by itself alone. On the other hand, that thing is necessary, or rather constrained, which is determined by something external to itself to a fixed and definite method of existence or action. (Spinoza, *Ethics*)

Introduction

My thesis is that selfhood, rationality, freedom and responsibility form a conceptual system in that each concept must be understood in terms of the others. Buying, selling and money, for example, form a conceptual system: if you grasp one of these concepts, you understand all three. You might say they are three sides of the one coin.

I will proceed by examining in turn each of the six relationships between the four concepts. Along the way I hope to indicate the specific meanings I attribute to the four terms.

1. Free actions are actions of a self

Only a self can be free. To be free is to act autonomously, from one's own nature, not to be pushed around by forces extraneous to the self.

But what is a self? I understand a self, in an Aristotelean fashion¹, as a mode of

¹Or a Dennettian fashion. See Dennett, Daniel C. (1989) *The Origins of*

organization, a way that a person reacts to the world, a function (or set of functions). In a simple, mechanical or billiard-ball view of the world, a cause impinges on an object and unleashes an effect in accordance with a simple, deterministic law. An organism with a self-structure interrupts this linear sequence: it absorbs stimuli and processes them in a complex manner by categorizing them linguistically, by relating them to the accumulated life history of the organism, by evaluating them in the light of its values, and by judging where they fit into future projects – in a word, by deliberating about them. Only then does it deliberately originate a response which is appropriate to the situation. Note that I am not saying that there is an object called a self which owns such a structure; I am not attributing the structure as a property to some substance or reality underlying the organization. The self is not a *thing*, physical or metaphysical, but is simply the mode of functioning by which some organisms integrate current information with past formation and future aims.

The self is the tip of an iceberg: Most of the things which happen in our brains are not part of the self-structure. The cognitive processes involved in perception, for example, are not integrated into the self. The identification of an object as a table, the recognition of the surface before me as a face, the attractiveness and familiarity of the face, all happen independently of the self which has no insight into how they occur and has no control over them. Only the end-products of these processes enter the stream of experience of the self. In a parallel way, when the self decides upon an action such as walking down the street, the innervation of the muscles, the coordination of the

Selves. *Cogito* 3 163-173.

movements and the maintaining of balance are left to brain structures outside the self. In a time when the brain was even less understood than today, it is not surprising that Occasionalists attributed such opaque processes to God: an individual is physically in front of my body and miraculously I recognize their face; I will to walk and, look, there go my feet! Complex insights, novel ways of saying things, or even significant decisions about one's life may also sometimes occur outside of the self with only the product being experienced. Since in these cases (unlike the case of perception) our expectations are that the self will be in control, the event erupts into our lives as a surprise and so may be interpreted as a Call from God, the Inspiration of a Muse, or the Creative Act of a Super-self. Insofar as such decisions or actions escape integration into the self-structure they are alien to the self and they fail to be integrated into the history, values, or projects of the self. Such actions are therefore not free.

My action is called free, then, as Spinoza points out in my masthead quotation, in so far as it is in accord with my nature. To be free is to act in accord with my own values, with my own goals, in function of who I am and who I have become. I fail to be free, that is, I am governed by necessity in so far as my response is determined by something external to myself, external to this structure. Stimuli such as reflexes which directly release responses, without first being integrated into the self, are not free, are not *my* actions, actions of my *self*. Insofar as a causal chain of events does not pass through the structure we call selfhood, to that extent an action is not under my control; it is not free.

2. A self is responsible for its actions

Selfhood is, as Locke claims (Locke Ch 27 Para 26), a forensic concept, that is, it is the basis for responsibility, for praise and blame, for reward and punishment. Where there is no self, in the case of animals for instance, we don't hold individuals responsible for their actions. In multiple personality disorder where (maybe) there are many alters in the one body, we do not hold one alter responsible for actions committed when another alter is "out." If, as in one of Locke's or Parfit's body-switching scenarios, someone (Jane) commits a crime on Monday but wakes up on Tuesday morning with Joe's memories and no recollection of Jane's actions, that is, with Joe's self in Jane's body, we would not hold Joe's self responsible and punish him just because he is now trapped in the body of the perpetrator. Would it not be grotesque to punish a war criminal who, due to Alzheimer's, had no recollection of his crimes? While I don't wish to commit myself to the theory that selfhood is solely a matter of memory, these scenarios bring out the conceptual relationship between selfhood and the attribution of responsibility.

3. Responsible acts are free

To hold someone responsible for an action presupposes that the action is free. If the action is coerced we do not consider the individual responsible. As Campbell claims, "any threat to freedom ... is a threat to moral responsibility." (Campbell 36) Campbell's main concern in "In Defence of Free Will" is the belief that in a deterministic universe no one could be held responsible for their actions. While I agree with his claim that all responsible acts are free ones, Campbell goes on to say that

From our practical self-consciousness we gain a notion of genuinely creative act – which might be defined as an act which nothing determines save the agent’s doing it. ... But the critic says ‘No!’ this sort of thing cannot be. A person cannot without affront to reason be conceived to be the author of an act which bears, *ex hypothesi*, no intelligible relation to his character. (Campbell 50-51)

Clearly I side with his critic. If an action does not flow from my historically developed self, from my values and projects, then I cannot be held responsible for it. If his creative act was, for example, a crime, would Campbell plead to the judge that it wasn’t *him* – Campbell the philosophy professor of good reputation – who did the act? “It was out of character,” he might say, or “Something just came over me,” “It wasn’t really me.” Responsibility only makes sense for an ongoing self who got himself into the situation and who is going to be around to accept the consequences. (You’ll be relieved to know that Campbell believes such out-of-character actions to be “comparative rarities.” (46) What I mean by free actions, on the other hand, are very common; they are the everyday decisions for which I as a self accept responsibility.)

4. A responsible act is one that can be justified by reasons

What is a to be responsible? I’m not responsible for my reflexes. I am responsible only for my deliberate actions, that is, actions about which I have deliberated, or could have deliberated. Such deliberation may be buried in the past: months or years ago I may have examined

the issue in the light of my goals and my values and arrived at a default decision which carries forward by habit to the action, which, without current deliberation, I perform today. If challenged about today’s action, I would be able to resurrect my former reasons and offer them as current justification. In the case of behaviour for which I am not responsible, I can offer no such justification. And I am also responsible for actions I should have deliberated about but didn’t. Actions for which I am responsible, then, are actions to which I have an alternative, a choice, and for which I had a reason to choose one rather than the other. If I am unable to give a reason for my action, then it is not an action for which I am responsible. Socially, we want people to be responsible because we want them to act not on impulse but after considering the consequences and social implications of the proposed action and to be prepared to offer the reason for the action as an explanation of it, or a justification for it.

One might object that we hold people responsible for acting “irresponsibly.” When we accuse people, in popular language, of acting “irresponsibly,” I take this to mean that such people are indeed responsible for their actions, otherwise there would be no point in accusing them; we simply condemn the reasons for which they acted. We hold Gordon Campbell responsible for his “irresponsible” driving because his reason for driving while drunk – to get back to his hotel – was a bad reason under the circumstance. The famous Toronto case of a murder by someone who was sleep-walking illustrates that when someone is incapable of having a reason for their actions, in this case because of sleep, we hold him to be irresponsible in the strong sense that his behaviours are ones for which he is not responsible because he is not rational. But what is rationality?

By rationality I mean the guiding and justifying of actions by reasons. An animal's reactions are based on genetic factors or individual learning. Insofar as humans live by culture we can reflect on and deliberate about our actions and offer reasons to justify these actions to ourselves or to others. This process is intrinsically linked to language. An animal's behaviour depends on many factors such as genes, pack conformity, or learning, but since it lacks language, we don't expect it to deliberate or be able to give reasons. Humans have an added factor: language. While some of our behaviour is governed by the same factors as animals, those behaviours, namely actions, which are subject to linguistic deliberation are ones we call free. Insofar as our actions are determined by purely physical or genetic factors, they are not guided by reasons and so are neither free nor responsible.

As Scanlon, discussing what he calls "reasons in the standard normative sense" (21) puts it:

A rational creature is, first of all, a reasoning creature – one that has the capacity to recognize, assess and be moved by reasons ... These reflective capacities set us apart from creatures who, although they can act purposely [animals] cannot raise or answer the question whether a given purpose provides adequate reason for action. We have this capacity, and consequently every action that we take with a minimum of deliberation about what to do reflects a judgment that a certain reason is worth acting on.
(Scanlon 23)

It is difficult to say in the abstract what a good reason for action is. This is not a defect in the analysis but is due to a certain flexibility in the notion of reason itself. What qualifies as a (good) reason depends on the context. If you ask me why I performed the action of moving a small piece of marble, I might justify it by saying it leads to checkmate. But this qualifies as a reason only within the context of chess. To engage in chess is to accept a regime within which a set of reasons are possible. The global action of playing chess authorizes a set of possible reasons for subordinate actions. Note that this is not a teleological point. I may move my queen in order to take a pawn, but the move is related to the game in a different way. I do not move my queen as a means to accomplishing my goal of playing chess; rather my moving the queen *is* my playing of chess, or at least part of it. Chess is the framework which establishes the possibility of certain kinds of reasons. If the Dean offers high grades as a reason for awarding a scholarship, such justification only makes sense within an educational institution and in the context of the values that it espouses. "To value something is to take oneself to have reasons for holding certain positive attitudes toward it and for acting in certain ways in regard to it."
(Scanlon 95)

Note that since not all values are ethical, not all reasons are moral ones, so when I speak of justification, I don't necessarily mean moral justification: I simply mean offering an account which indicates how the action was rational. One is responsible for one's moves in chess in so far as one can give reasons for them – accidentally upsetting the board is thereby excluded – yet these are not moral reasons.

5. A free act is one for which a reason can be given

Are rational actions, actions done for a reason, free? What is it for an action to be free? There are two requirements. First, as I have already claimed, the causation of the action must be processed by the self-structure. The second requirement is that a reason can be given for the action. If asked to justify a free action, I can offer a reason. Consider non-free actions. If I perform an action because I've been physically coerced, hypnotized, or because of a neuro-probe in my brain, I can offer no reason why I did it. I may, or may not, be in a position to offer a causal account of why the action occurred, but would be unable to offer a reason which could *justify* the action. Ultimately of course such an occurrence is not an action at all; in so far as it does not originate in my self, it is not mine.

6. A self acts for (long-term) reasons

My most central thesis is the claim that selfhood and rationality coincide, that the actions of a self are not only free and responsible actions, but can be justified by reasons. Why is this so?

I have said above that reasons require contexts. We might think of the self as the overall context within which any reason can occur. Schechtman claims, "we expect a person's beliefs, desires, values, emotions, actions and experience to hang together in a way that makes what she says, does and feels psychologically intelligible." (Schechtman 97) The structure of the self – which Schechtman maintains is a narrative structure – "means that the narrator should be able to explain why he does what he does,

believes what he believes, and feels what he feels. ... we take it for granted that people are able to answer questions about themselves. ... I expect an explanation in terms of [their] long-term goals, personality traits, talents, likes and dislikes."

(Schechtman 114) Insofar as I understand my life as having the structure of a narrative, I can justify specific actions by explaining their role within the story of my life. Why did you go on that peace march? Because I have been in the peace movement in the past, I am the kind of person who values peace, and future world peace is one of my projects.

Without committing myself to her narrative theory of the self, I think Schechtman is right in claiming that selfhood is the structure within which it makes sense to ask for reasons for actions. This is because we can think of the self as among other things, a set of values, a context within which particular reasons makes sense. A psychopath, for whom respect for others is not a value, cannot justify a action he has performed by citing kindness as a reason. If I have developed into the kind of self for whom peace is not a value, I cannot legitimately offer opposition to war as a justification for marching down the street. An action which does not originate from my values, which is independent of my formed character (one of Campbell's "creative" actions for instance) would be an action without a context within which a justifying reason could be offered. Such a creative action would appear to the self as extraneous, as escaping the history, values and projects that make up the self-structure. In so far as no reason could be given for such an action, such an action could not, as I see it, be one we could hold the self responsible for.

Already I can hear the objections. A purely rational characterization of a self is a

caricature. Selves are spontaneous and creative, they have feelings and emotions, they are given to passion, and so on. But there is rationality and rationality!

A. Rationality

Rationality can be understood in either a short-term or a long-term sense. If I decide every moment what it is most rational for me to do at that moment, then long-term projects become impossible. To commit myself to long-term project requires that I blind myself to the attractions or distractions of stuff which it might be rational for me to do at each moment along the way. It may for instance be in my short-term interest to lie on a particular occasion, but it may be in my long-term interest to become the kind of person others know they can trust.

While not all reasons are teleological, many are, and it is to teleological reasons that I wish to bring the short-term/long-term distinction. I have interests, such as survival, reproduction, or wealth, and I use my intelligence to deliberate about the best means to obtain these goals. The means I choose may be short-term, as when I am hungry I grab a piece of bread, or long-term, as when I educate myself in order to get a job in order to be able to buy bread after retirement. Long-term rationality usually involves cooperation with other individuals in a reciprocal, ongoing, social relationship which requires mutual trust. If Mary does a job of work for Joe today, Mary must trust Joe to pay her tomorrow. But there is a problem: if Joe is into short-term rationality then tomorrow, when Mary's work for him is completed, it is in his interest to renege on the deal and keep his money. On the wider social scale this is the difficulty that Brian Penrose referred to last fall as the free-rider problem. From the viewpoint of long-term rationality, Mary's problem is to determine

whether Joe is the kind of person she should cooperate with. But of course the situation is a problem for Joe too. From his point of view, if he wants to convert from a short-term to a long-term perspective, he needs to figure out how to convince Mary and others that he is trustworthy. He could simply tell her that he will keep his promises, but talk is cheap and Mary is not likely to trust him just on that basis: Will he not be free tomorrow to do what is in his immediate self-interest? How can Joe convince Mary that he is not a free-rider, that he will not renege on her tomorrow? "So here is the main problem: not just how can you make yourself into an agent that can be trusted in commitment problem cases, but how can you credibly advertise the fact that you are to be so trusted?" (Dennett 205)

This problem is faced by all cooperating groups. For most animals, genetics has solved the problem since groups in which individuals were not cooperative have become extinct. For humans, who live by reason, culture, and language, however, the problem must be solved differently. Robert Frank proposes that the solution is to be found in passion, in human emotion. What Joe must do is become the kind of person who will be so passionate that Mary will know that he will be unable to break his commitments in the future no matter how rational (in the short-term sense) it may be for him to do so. He must train himself in virtue in such a way that he will not be free to guide his actions by short-term reasons. If he promises that he will work with me for the next year on the peace campaign, I have good reason to trust him because I can see his passionate commitment to the cause, a passion which will override any short-term interests which may appear during the year. When, on a short-term basis, he wants to sleep, his passion will keep him awake and working to accomplish his, and my, project.

The passion I experience, the passion I display, both serve important roles in rationality. The people may elect me as a politician because they see my passion for peace and therefore trust me to pursue the cause. If I simply state the reasons which justify my commitment to peace, my constituents may think that perhaps I'll find new reasons tomorrow and betrayed the project for which they are electing me. The same considerations apply to all cooperative relationships, including reproductive and economic one. One evolutionary function of human passions is to signal to others what I am in fact likely to do, so they can predict my actions and act accordingly.

Paradoxically this means that a person must make herself short-term irrational in order to be long-term rational. "Amartya Sen has called the caricature of the shortsighted self-interested person a 'rational fool'. If the rational fool turned out to be taking shortsighted decisions then he is not being rational just shortsighted. He is indeed a fool who fails to consider the effect of his actions on others." (Ridley 137)

When I claim, then, that free and responsible actions are rational ones, I mean true rationality, not the foolish kind. Rationality in this long-term sense – the only true sense – includes those actions based on passions which carry us beyond short-term interests and keep us focussed on our long-term rational goal of cooperation with others. As Dennett puts it, we must *be* good in order to *appear* good. (Dennett 202)

B. Selfhood

So what kind of being is capable of carrying out this trick? I argue that it is a "self" in contradistinction to a Cartesian "ego."

Descartes' interest in the ego is to find an absolute epistemological foundation. Clear and distinct ideas are certain, but only for the instant, since memory of the past and expectations of the future are untrustworthy. Indeed, the past and the future are so disconnected from the present instant that only the miraculous intervention of God can establish a relationship between them.

For a lifespan can be divided into countless parts, each completely independent of the others, so that it does not follow from the fact that I existed a little while ago that I must exist now, unless there is some cause which as it were creates me afresh at this moment -- that is, which preserves me. For it is quite clear to anyone who attentively considers the nature of time that the same power and action are needed to preserve anything at each individual moment of its duration as would be required to create that thing anew if it were not yet in existence. (Descartes, Meditation #3)

The notion that I exist as a sequence of independent points – let us call it "punctualism" – may have some value in epistemology but is a disaster when we come to consider freedom, rationality, and selfhood. Punctualism restricts rationality to consideration of the best thing to do at this current instant.

Opposing Philebus' notion that the good is immediate enjoyment, Socrates points out:

If you had not memory, it obviously follows that you would never so much as remember that you had known enjoyment before, and that the present moment's visitation of

pleasure would leave no trace of itself behind. ... without the power of calculation you could not even calculate that you will get enjoyment in the future; your life would not be that of a man, but of one of those marine creatures whose bodies are confined by a shell. (Plato, *Philebus* 21C)

Unlike a punctual Cartesian ego, a self is a structure which integrates experience over time. A self is, and understands itself to be, identical with who it was in its past and because of this acknowledgement it accepts responsibility for its previous actions, promises and decisions. A self has projects to which it has committed itself in the future.

There is no reason for me to deliberate about or undertake a project unless I know that my current decision will bind me in the future. The kind of rationality which a self has is historical. A punctual self which acted on the instant in the light of its immediate goals at each moment would erase its history and start life anew every instant. A punctual consciousness couldn't have any long-term goals. Indeed it couldn't have any goals at all, because a goal requires that I care about the future, which is possible only if there is a continuing self who I will still be next year. The notion of short-term rationality becomes meaningless when we push it to its limit in the instant. As Merleau-Ponty puts it, – arguing I suspect against Sartre – freedom cannot be absolute if by that we mean that at every moment we are condemned to reconsider our choices.

The very notion of freedom demands that our decision should plunge into the future, that something should have been *done* by it, that the subsequent instant should benefit

from its predecessor and, though not necessitated, should be at least required by it. If freedom is doing, it is necessary that what it does should not be immediately undone by a new freedom. Each instant, therefore, must not be a closed world; one instant must be able to commit its successors and, a decision once taken and action once begun, I must have something acquired at my disposal, I must benefit from my impetus, I must be inclined to carry on, and there must be a bent or propensity of the mind. (Merleau-Ponty 437)

Real freedom, Merleau-Ponty claims, requires a subject that can engage itself for the future with the assurance that its future self will remain faithful to the commitment the current self is embarking upon.

Dennett (p. 205) puts the issue very graphically:

Sometimes one problem can be solved by another problem. This is especially true when the problem is confronted by Mother Nature, that master opportunist. We have a problem of self-control that is truly hard – costly – for us to solve. According to Frank, the fact that it is costly to solve is a blessing, not a curse. It is the problem exemplified by Ulysses and the Sirens, where the trick is to devise some way of tying yourself to the mast and blocking your sailors' ears with wax so that you can't act on your strongest inclination of the moment. (The trick is to arrange it so that "at time *t*" your will is *ineffective*.) Ulysses knows perfectly well the long-term benefits of adopting the policy of avoiding the Sirens when they sing their seductive song, but he also knows he is disposed in many circumstances to over-value immediate payoffs, so he needs to

protect himself from a somewhat misshapen preference structure that he expects will impose itself on him when time t rolls around. He knows himself, and he knows what evolution has provided for him: a slightly second-rate faculty of reason that will cause him to take the immediate payoff ("I couldn't do otherwise," he'll say, as he jumps into the Sirens' arms) unless he takes steps now to distribute his decision-making over more favorable times and attitudes. His seduction by the Sirens is not *inevitable*, provided he has enough lead time to prepare his avoiding move. As Frank observes,

It is important to stress that the experimental literature does not say that immediate payoffs get *too much* weight in every situation. It says only that they always get *very heavy* weight. On balance, that was likely a good thing in the environments in which we evolved. When selection pressures are intense, current payoffs are often the *only* ones that matter. The present, after all, is the gateway to the future. (Frank 1988, p. 89)

In the limit, a punctual consciousness with absolute freedom every instant would not be a self at all. A similar theme can be found in Ricoeur who claims that selfhood requires "permanence in time" and one of his models for such permanence is "the constancy of the self in promising." (Ricoeur 124) "[The model] is that of keeping one's word in faithfulness to the word that has been given. I see in this keeping the emblematic figure of ... [personal] identity." (Ricoeur 123) A self is not a *substance* which endures through time in a mechanical manner, a piece of reality-in-itself; it sustains its own identity by *fidelity*: it is faithful to its past commitments and trusts itself to carry out in the future the decisions it is currently

making. The difference between a self and punctual consciousness is that a self is rational in this long-term sense.

C. Conclusion

We can now see how selfhood and rationality fit together. A self, as a structure that sustains itself over time, acts freely and responsibly by being rational. But this rationality is not immediate, short-term rationality, a concept which in the limit is meaningless. Long-term rationality involves commitment which shows itself as passion. It is precisely this passion, intrinsic to selfhood, which signals to others that I am the kind of person with whom they should cooperate. Selfhood and long-term rationality fit like hand in glove. A social order is possible because other cooperators can trust me precisely in so far as I have developed a self capable of self-control, that is, the ability to avoid the distractions of the moment, of short-term gain, in so far, that is, as I have developed the virtue of remaining faithful to the promises and commitments I have made.

Rational decisions must be converted into character, that is a long-term set of habits on which both I and others can depend. A self is constructed by its history. The kind of being which allows its history to determine its current character to is a self. One's character is what one has made of oneself has is rational to make oneself into character one can trust. And incidentally, that others can trust.

Bibliography

Campbell, C.A. *In Defence of Free Will*. New York: Humanities Press, 1967.

Dennett, Daniel C. "The Origin of Selves." *Cogito* 3 (1989) 163-173. Also at <http://ase.tufts.edu/cogstud/papers/originss.htm>

Dennett, Daniel C. *Freedom Evolves*. Toronto: Viking/Penguin, 2003.

Frank, Robert H. *Passions within Reason: The Strategic Role of the Emotions*. New Norton, 1988.

Locke, John. *An Essay Concerning Human Understanding*. New York: Dover, 1959.

Merleau-Ponty, Maurice. *The Phenomenology of Perception*. London: Routledge & Kegan Paul, 1962.

Ricoeur, Paul. *Oneself as Another*. Chicago: Chicago University Press, 1992.

Ridley, Matt. *The Origins of Virtue: Human Instincts and the Evolution of Cooperation*. Toronto: Penguin, 1996.

Scanlon, T. M. *What We Owe to Each Other*. Cambridge: Harvard University Press, 1998.

Schechtman, Marya. *The Constitution of Selves*. Ithaca and London: Cornell University Press, 1996.