

1. Personhood

There are many different accounts of what it is that distinguishes persons from other agents.

Content-based. Boethius to Dennett: *rationabilis*, interpreting, reciprocating, self-conscious..

Duty-based. Locke: 'a Forensic Term appropriating Actions and their Merit'

Right-based. Rawls: 'self-authenticating sources of moral claims'; i.e. bearers of rights.

Ideal-based. Kant to Frankfurt: their desires answerable to reason or second-order desires....

It would be nice to have an account of persons that made sense of these associations,

explaining why persons reason+; incur duties; enjoy rights; and engage an elusive ideal.

It would also be nice to be able to explain why there are legal persons—and might be robotic.

It turns out that a quasi-Hobbesian, language-first account can deliver on these fronts.

Hobbes introduces the notion of personating: playing a part, as in a character in a play;

from '*persona*': literally, a role-related mask through (per) which an actor sounds (sonare).

He then characterizes persons as agents with a(n exercise-dependent) capacity to personate:

to speak for themselves in words they 'author' and 'own', inviting others to rely on them.

I personate if I commit to you, manifestly speaking for myself with a presumptive authority:

in avowal, I rule out the possibility of misreading my mind, in pledging, changing my mind.

I may commit actively by what I say but I will more often commit in a virtual way: this,

when it is manifest that others ascribe an attitude, or expect an action, and I do not demur.

Committing under this practice, I communicate the expectations to which others may hold me.

As a natural person I self-personate but, as Hobbes insists, I may personate others as well;

I may speak for those others: e.g., when I am authorized as spokesperson for a group agent.

He endorses the medieval account of the legal person, represented by an authorized voice,

and reworks it to give us a general account of all persons, including natural persons.

The personation theory explains a range of issues, *relative to the practices of commitment*:

why persons have to reason, interpret, reciprocate, be self-aware...: needed for the practice;

why they incur duties: to renounce an excuse for not X-ing is to commit, *pro tanto*, to X-ing;

why all persons must enjoy some rights: these will identify areas where they may commit;

why being a person identifies an ideal: that of being able to make and honor commitments;

how there can be corporate persons (and robotic?), with suitable abilities, duties, rights.

Note. Suppose I cannot commit to myself (except by mimicking the act of committing to others);

suppose, like Hobbes, that this is ruled out by my self-control : 'he that can bind, can release'.

In that case, we persons can exist only in one another's company, under commissive practices.

This illustrates an outside-in view of our capacities, similar to that upheld in earlier lectures.

2. Selfhood

Every agent's attitudes will implicate itself *qua* the agent that its beliefs and desires guide,

that is given memories by its experiences, that preserves its formed attitudes...

It must also have beliefs about itself (*de se*) under an aspect such that reference cannot fail;

even our robot has indexical beliefs that things are 'near' (to itself) that cannot mis-refer.

To count as having a self, plausibly, an agent must not only have such self-engaging beliefs,

but also beliefs (attitudes) that refer to itself *de dicto*; it must exist for itself. (Locke)

As a person I count as having a self, for I commit myself as 'I', and am committed as P.P. But that's only the start of the story, for I can identify my-self under various aspects: as a 1st-person, *referenced self*, a 2nd-person *personated self*, or a 3rd-person *alienated self*.

i). The first-person, *referenced self* is the self I engage (*de se*), in the manner of any agent.

Metaphysically, this self is *that* agent, the one engaged by my attitudes and my actions. I am and will survive as this self insofar as I continue as the referenced agent; no 'deep fact'. Thus, I might survive re-embodiment (teletransportation), or fission into two later agents, but I would hardly survive fusion with another, or becoming a merely vegetative being.

Epistemically, I cannot misidentify this self (Evans) but may know little about it (Hume). But not nothing: after all, *Cogito, ergo sum* says more than *Cogitatur, ergo id est* (Williams). Conscious of my attitude contents, I will *apperceive* the self that they guide and reference; it will be available in the way the angle of a photo-shot is available in what it depicts. I know this self *by presupposition*, we might say, not by acquaintance and not by description.

Practically, this is a self I am invested in/care for as an agent, as in Rousseau's *amour de soi*. But this care is constitutive, not selfish: PP in the future is me because I care, not vice versa. There is no reason why I should not care for myself in this way over my life. 'Connecting fact'.

ii). The second-person, *personated self* is the self I invite others (*de dicto*) to take me to be.

Metaphysically, this is the authorized *persona*/image of self I project in my commitments; these will multiply as I commit myself *virtually*, not rejecting others' manifest expectations. While it is something I construct and reconstruct, it is not the product of a narrative (Strawson) but a byproduct of my commitments: 'this is who I am'; 'this is who I back myself to be'.

Epistemically, I will have access to this self on the basis of a maker's knowledge of what I do in making this or that commitment, or acquiescing in one or another manifest expectation.

Practically, I am bound to care about this personated self, trying to live up to it (*amour juste*); it is central to my standing as a reliable person. 'This, above all: to thine own self be true, And it must follow, as the night the day, Thou canst not then be false to any man'. Polonius is right! To lack this standing would deprive me of bearings ('integrity'), not just hurt me with others.

iii). The third-person, *alienated self* is a self identified *de dicto* by others (by their dictum).

Metaphysically, this self is the changing figure(s) that I cut among (one or more sets of) others: the 'character' that others give me, whether manifestly or not, that I may not control.

Epistemically, this self is only as accessible to me as the opinions of others; third parties may be better able to identify it, for no one is likely to gossip to me about me!

Practically, it is probably self-destructive to care about this self (Rousseau's *amour propre*). Acting on the concern may undermine *amour juste*; it may be fruitless, even self-defeating. And it may prompt a generally damaging search for positional advantage over others (Hobbes). Thus, it may unleash *Ehrsucht*, *Herrschaft* and *Habsucht*, in Kant's terms, and introduce a zero-sum game: the pursuit of Hobbesian 'eminence' in prestige, power and possession