CAMBRIDGE JOURNALS

Proof Delivery Form

Philosophy

Date of delivery: 12-05-2020

Journal and vol/article ref: phi PHI2000017

Number of pages (not including this page): 27

This proof is sent to you on behalf of Cambridge University Press.

Authors are strongly advised to read these proofs thoroughly because any errors missed may appear in the final published paper. This will be your ONLY chance to correct your proof. Once published, either online or in print, no further changes can be made.

****Please return your proofs within 3 working days ****

HOW TO RETURN YOUR PROOFS

Proofs **MUST** be sent to Assistant Editor, Matt Hewson

Email: **assistant@royalinstitutephilosophy.org**

You can mark up proofs either on screen using the enabled electronic editing tools (our preferred method), or by hand on a hardcopy print-out.

Marking up **electronically.**
All proofs are enabled to allow electronic annotation in the freely available Adobe Reader software. Using your cursor, select the text for correction and use the most appropriate single tool (i.e. 'Replace', 'Cross out', 'Insert' or 'Add note to text'). Please return the file as an attachment.

Marking up by hand. Alternatively, please print the PDF file, mark any amendments on the proofs, and list the corrections on a separate page, scan the marked-up proofs and email them to the address above.

If the amendments are minor, corrections may also be sent as a list, in the text of an email. If you specify the line number, and the correction to be made, the editorial team will enter it.

Changes should be limited to the correction of editing and typographical errors and the answering of any author queries. Any corrections that contradict journal style will not be transferred.

If you have **NO** corrections, please let assistant@royalinstitutephilosophy.org know.

Thank you for publishing in *Philosophy*. You will automatically receive a link to the PDF version of your article once it has been published online.

Philosophy

Please note:

• The proof is sent to you for correction of typographical errors only. Revision of the substance of the text is not permitted, unless discussed with the editor of the journal. Only **one** set of corrections are permitted.

• Please answer carefully any author queries.

• Corrections which do NOT follow journal style will not be accepted.

• A new copy of a figure must be provided if correction of anything other than a typographical error introduced by the typesetter is required.

• If you have problems with the file please contact **phiproduction@cambridge.org**

Please note that this pdf is for proof checking purposes only. It should not be distributed to third parties and may not represent the final published version.

Important: you must return any forms included with your proof**.**

Please do not reply to this email

NOTE - for further information about **Journals Production** please consult our **FAQs** at http://journals.cambridge.org/production_faqs

CAMBRIDGE UNIVERSITY PRESS

Journal – Copyright Transfer Form

Please complete both **Sections A and B**, sign, and return this page by post to Royal Institute of Philosophy, 14 Gordon Square, London WC1H 0AR, UK, or by email to journalscopyright@cambridge.org **as a scanned, signed (but not electronically signed) document,** as soon as possible. By completing, signing and returning this form you hereby agree to the Terms and Conditions enclosed (Doc.JCT.T&C.HSS.14.1).

Philosophy PHI

In consideration of the publication in **Philosophy**

of the contribution entitled: .....

.....

by (all authors' names): .....

.....

Section A – Assignment of Copyright (fill in either part 1 or 2 or 3)

1 To be filled in if copyright belongs to you

Transfer of copyright

I/we hereby assign to Royal Institute of Philosophy, full copyright in all forms and media in the said contribution, including in any supplementary materials that I/we may author in support of the online version.

I/we hereby assert my/our moral rights in accordance with the UK Copyright Designs and Patents Act (1988).

Signed (tick one) ☑ the sole author(s)

□ one author authorised to execute this transfer on behalf of all the authors of the above article

Name (block letters) Philip Pettit

Address UCHV, Marx Hall, Princeton University, Princetion, NJ 08544, USA

Institution/Company Princeton University

Signature *Philip Pettit* Date: 17 May 2020

(Additional authors' names, affiliations and nationalites should be provided on the second sheet of this form and all should be aware of, and accept, the terms of this form and accompanying form Doc.JCT.T&C.HSS.14.1.*)*

2 To be filled in if copyright does not belong to you

a Name and address of copyright holder .....

.....

.....

b The copyright holder hereby grants to Royal Institute of Philosophy, the exclusive right to publish the contribution in the Journal including any supplementary materials that support the online version and to deal with requests from third parties.

(Signature of copyright holder or authorised agent) .....

3 US Government exemption

I/we certify that the paper above was written in the course of employment by the United States Government so that no copyright exists.

Signature: ..... Name (Block letters): .....

Section B – Warranty and disclosure of conflict of interest

I/we warrant that I am/we are the sole owner or co-owners of the contribution and have full power to make this agreement, and that the contribution has not been previously published, contains nothing that is in any way an infringement of any existing copyright or licence, or duty of confidentiality, or duty to respect privacy, or any other right of any person or party whatsoever and contains nothing libellous or unlawful; and that all statements purporting to be facts are true and that any recipe, formula, instruction or equivalent published in the Journal will not, if followed accurately, cause any injury or damage to the user.

I/we further warrant that permission for all appropriate uses has been obtained from the copyright holder for any material not in my/our copyright including any audio and video material, that the appropriate acknowledgement has been made to the original source, and that in the case of audio or video material appropriate releases have been obtained from persons whose voices or likenesses are represented therein. I/we attach copies of all permission and release correspondence.

I indemnify and keep Royal Institute of Philosophy and Cambridge University Press, indemnified against any loss, injury or damage (including any legal costs and disbursements paid by them to compromise or settle any claim) occasioned to them in consequence of any breach of these warranties.

Name (block letters)……...

Signature ……*Philip Pettit*…… Date 17 May 2020…

(one author authorised to execute this warranty statement above and conflict of interest statement below on behalf of all the authors of the above article)

Please disclose any potential **conflict of interest** pertaining to your contribution or the Journal; or write 'NONE' to indicate you declare no such conflict of interest exists. A conflict of interest might exist if you have a competing interest (real or apparent) that could be considered or viewed as exerting an undue influence on you or your contribution. Examples could include financial, institutional or collaborative relationships. The Journal's editor(s) shall contact you if any disclosed conflict of interest may affect publication of your contribution in the Journal.

Potential conflict of interest ……

The information provided on this form will be held in perpetuity for record purposes. The name(s) and address(es) of the author(s) of the contribution may be reproduced in the journal and provided to print and online indexing and abstracting services and bibliographic databases.

Please provide additional authors' names, addresses, affiliations, and nationalities on this page. All authors should be aware of, and accept, the terms of this form and accompanying form 'Doc.JCT.T&C.HSS.14.1'

(Doc.JCT.T&C.HSS.14.1)

Terms and Conditions for authors to Humanities and Social Science journals published by Cambridge University Press.

By completing, signing and returning the Copyright Transfer form (Form JCT.14.1) you have agreed to abide by the following Terms and Conditions. Please retain this document for future reference.

Definitions[25]

Author's Original (AO)	Any version of the article that is considered by you to be of sufficient quality to be submitted for formal peer review by a second party. Content and layout as set out by you.
Submitted Manuscript Under Review (SMUR)	Any version of the article that is under formal review for inclusion in the journal. Content and layout follow the journal's stated submission requirements.
Accepted Manuscript (AM)	The version of the article that has been accepted for publication in the journal. This version may have been revised following peer review but may be subject to further editorial input by Cambridge University Press.
Version of Record (VoR)	The fixed version of the article that has been made available. This includes an "early release" article as long as it is citable via some permanent identifier(s). This does **not** include any "early release" article that has not yet been "fixed" by processes that are still to be applied, such as copy-editing or proof corrections. The VoR includes any corrected or enhanced VoR.
Personal Website	A non-commercial website maintained solely or in part by you.
Departmental / Institutional Repository	The online archive of intellectual output of the institution at which you were based when the article was written as well the repository of any future institution at which you are based.
Non-commercial Subject Repository	A repository relating to the subject area of your article which does not allow the content to be used commercially.
Commercial Repository	A repository which permits the content to be used for commercial gain.
Social Media Sites	Websites and computer programs that allow people to communicate and share information on the internet using a computer or mobile phone[26]

Re-use of your article.

All postings shall include a prominent bibliographical reference and statement of copyright ownership. Where possible all postings should include a link to the published article on Cambridge Journals Online. In relation to the posting of the AO or SMUR, a statement that such version has been accepted for publication and will appear in a revised form subsequent to peer review and / or editorial input by Cambridge University Press and / or the journal's proprietor ,as well as a link to the journal's site on Cambridge Journals Online, should be included.

[25] Adapted from NISO RP-8-2008, Copyright © 2008 by the National Information Standards Organization

[26] http://dictionary.cambridge.org/dictionary/british/social-media?q=social+media&utm_medium=widget_searchbox&utm_source=widget_searchbox_source&utm_campaign=widget_tracking

Cambridge does not permit full articles in AM or VOR form to be posted on Commercial Repositories or Social Media Sites including, but not limited to, Social Science Research Network ('SSRN'), ResearchGate, Academia.edu, Mendeley or LinkedIn.

Notwithstanding the assignment of copyright or grant of licence in your article, you retain the following non-transferable rights to deposit versions of your article, (subject to appropriate permission having been cleared for any third-party material):

	Personal Website	**Departmental / Institutional Repository**	**Non-commercial Subject Repository**	**Commercial Repository and Social Media Sites**
AO	At any time	At any time	At any time	At any time
SMUR	At any time	At any time	At any time	At any time
AM	On acceptance of publication.	On acceptance of publication.	On acceptance of publication.	Abstract only in PDF or HTML format no sooner than first publication of the full article.
VoR	Abstract only in PDF or HTML format no sooner than first publication of the full article.	Abstract only in PDF or HTML format no sooner than first publication of the full article.	Abstract only in PDF or HTML format no sooner than first publication of the full article.	Abstract only in PDF or HTML format no sooner than first publication of the full article.

Other

You may make hard copies of the article or an adapted version for your own purposes, including the right to make multiple copies for course use by your students, provided no sale is involved.

You may reproduce the article or an adapted version of it in any volume of which you are editor or author subject to normal acknowledgement.

If your reuse is not covered by the above please follow the 'Rights and Permissions' quick link on the Cambridge website for your region via www.cambridge.org.

Cambridge University Press co-operates in various licensing schemes that allow material to be photocopied within agreed restraints (e.g. the CCC in the USA and the CLA in the UK). Any proceeds received from such licences, together with any proceeds from sales of subsidiary rights in the Journal, directly support its continuing publication.

Cambridge University Press acts in accordance with the UK Bribery Act 2010 and the Data Protection Act 1998. Please refer to the Press's relevant policies, (http://www.cambridge.org/policy/privacy/), (http://www.cambridge.org/policy/abc_policy/) which may be revised from time to time.

Author Queries

Journal: PHI (Philosophy)

Manuscript: S0031819120000170jra

Q1 The distinction between surnames can be ambiguous, therefore to ensure accurate tagging for indexing purposes online (e.g. for PubMed entries), please check that the highlighted surnames have been correctly identified, that all names are in the correct order and spelt correctly.

My Three Selves*

PHILIP PETTIT Q1

Abstract
Having a self means being able think of myself under a certain profile that that is me: that is who I am, that is how I am. But if I raise the question as to who or how I am, there are three salient profiles in which I can cast myself, three selves with which we can identify. I can see myself just as an agent identified over time by the linkages between my experiences, my attitudes and my actions. I can see myself as the persona that I invite others to rely on and, if sincere, internalize. And I can see myself as the figure I cut in other people's eyes, whether or not I welcome that image. Such ambiguities help explain the complexity in philosophical discussion of the self as well as the conflict in everyday exhortations to be ourselves and know ourselves, yet also to forget ourselves and lose ourselves.

1. Introduction

Our traditions of thinking about the self are ambiguous, and the mantras they support conflicting. We are told to be ourselves, but also to let go of the self; to know ourselves but also to forget the self, even to see it as an illusion; to be true to ourselves but also to avoid the snares of self-concern. This lecture is an attempt to find a way through this thicket of confusion, outlining a map in which the self appears in three different roles and offers each of us three different points of orientation.

The paper is in two main parts. In the first, I set out some basic assumptions about the nature of agency, explain the connection between agency and selfhood and argue that the connection becomes much tighter when the agent is a person. And then in the second part, I rely on those assumptions to distinguish between three different ways in which the self must present to any one of us,

* This paper is based on the text of the Royal Institute of Philosophy Annual Lecture, London, Oct 2019. I am grateful for comments received on that occasion and also two others: at an earlier airing of the material in Oxford, as part of the Locke Lectures 2018–19, and in a later discussion at the Australian National University. I owe a particular debt to Victoria McGeer, and to the editors, Maria Alvarez and Bill Brewer, for their detailed comments on late drafts.

doi:10.1017/S0031819120000170

describing the first as the referenced self, the second as the personated self, and the third as the imputed self. The first part provides essential background, while the second presents the paper's headline claim. A short conclusion summarizes the overall view.

2. Agents, selves and persons

Agency

By all accounts an agent is a system that acts to bring about certain goals or purposes. These are states of affairs that it generates on a reliable basis, and not just by accident: on the basis, for example, of having been selected or designed to realize the goals or to follow certain procedures that identify goals to realize. But agency requires more than the pursuit of purposes. Otherwise the sunflower plant that tracks the sun – that reliably moves so as to maintain its orientation towards the sun – would count, implausibly, as an agent.

As we think of agency in our ordinary exchanges, the most obvious requirement over and beyond the pursuit of certain goals is that this pursuit should be maintained over variations in the circumstances of the system: variations that are more significant, intuitively, than those in the angle of the sun that the sunflower has to cope with. A system may count as an agent and be so simple that it has only a single goal. But still, it must be able to pursue that goal across different scenarios, adjusting its behavior ~~as~~ so as to realize the goal under the particularities of each situation.

Take the simple robot that is constructed to pursue the goal of raising certain objects on a flat surface to an upright position (List and Pettit 2011, Ch 1). Such a robot will have to be equipped with some apparatus for determining whether any glass or cup or bottle on the surface – say, a tabletop – is on its side or upright; think of this as an eye-like receptor that continually scans the objects on the surface. And then, presented with a~~n~~ bottle on its side, the robot will have to be organized across variations in the size and shape of the object, and its distance and direction, to adjust behaviorally so as, things going well, to put it upright: this, presumably, with the help of wheels for moving, levers for lifting, mechanisms for grasping, and so on.

In order for the robot to adjust its behavior appropriately in different scenarios –and count thereby, we may suppose, as an agent – it must be able in some sense to register the particularities of each scenario. It must change in response to receptor inputs, and its changed

state must then serve to shape the behavior – and to shape the behavior as it evolves in time (Hurley, 1998) – so as to achieve the goal. This is to say, in other terms, that the robot must form representations of how things are in each scenario where it acts, and let those representations direct its behavior there.

The general lesson, in familiar Humean terms, is that an agent is a system that reliably acts to fulfill its goal-seeking states according to representations that it reliably forms: in perhaps a deflationary sense of those terms, it acts to satisfy its desires according to its beliefs. The agent's reliability in forming beliefs constitutes epistemic rationality, its reliability in acting so as to satisfy desires constitutes practical rationality. The system may not be unfailingly reliable or rational but, if it is to count properly as an agent, then it must generally fail only when circumstances are abnormal by independent criteria. Thus, the robot may be misled under certain lighting about the position of an object; and it may not manage to put an object at the edge of the table upright: it may knock it to the floor. In the first case, it would fail epistemically, in the second practically.

This simple image of agency is not uncontroversial but it is at least familiar, being supported with variations by recent philosophers like David Lewis (1983), Robert Stalnaker (1984), and Daniel Dennett (1987). Others may place demands on agency over and beyond the requirements it encodes: they may demand capacities, for example, associated here with personhood. But still, it represents a model of agency that is an intelligible reconstruction of our commonsense assumptions, if not the only reconstruction admissible (Pettit 1993, Ch 1).

However simple, of course, our model of agency allows of many specifications, applying in different ways across the spectrum from robot to animal to human. Where we human beings pursue purposes of highly distinctive, contrasting kinds, for example, other agents may act only in pursuit of a few simple goals. Where we employ highly sophisticated representations of different types of situation, and represent how things are, were or will be, as well as how they may or must be, other agents may not range much beyond the here and now. And, of course, where we human beings use a common language to express and shape our purposes and representations, robots and animals do not do so in the same way.

Selfhood

Every agent, by virtue of what agency involves, treats itself as special. It acts on the basis of the memories and beliefs that its own

experiences support, not the experiences of any other agent. It forms desires on the basis of the sorts of things that attract it – its likes and dislikes – and not those of any other agent. And it acts on the basis of the intentions it itself forms, not the intentions that materialize in any other. Talk of experience, memory and attraction may be out of place with the robot, suggesting that it must be conscious. But even the robot treats itself as special in forming beliefs on the basis of the inputs it registers, in determining situation-specific goals on the basis of its general, hardwired goal, and in acting to pursue those situation-specific goals that it settles on.

It is noteworthy that every agent must treat itself as special, not just at a time, but over time. This entails that whether X at time t-n is the same agent as Y at the later time t shows up in the fact that X's experiences, tastes and intentions shape in a unique way the memories, beliefs, intentions and actions of Y. And whether Z at time t + n is the same agent as Y at the earlier time t shows up in the fact that similar relations to those between X and Y hold between Y and Z.

In order to treat itself as special in such ways, every agent, even one as simple as our robot, has to form representations or beliefs about itself. Thus, if the robot acts intentionally so as to put a glass on the table into an upright position, it will have to move towards that glass and at a certain point form a representation, whatever its exact content, to the effect that the glass is within reach; at that point, all going well, it will stop moving and reach out for the glass. The responses to this representation that the agent displays in ceasing to move and in reaching for the glass will be programmed into the robot, of course, but if we are to make sense of them from within the intentional stance – if we are to see them as the responses of an agent – then we must characterize the representation in a way that makes them rational. That means that it must be a representation or belief to the effect, from its standpoint, that the glass is within *my* reach, that it is graspable *by me*. The belief must register a relationship between the glass and the robot, not a property of the glass alone, and if it is to be expressed in words, as from a first-person viewpoint, the formula used has to deploy the first-person indexical 'I' or a variant (Perry, 1979).[1]

As even the simple robot will have to form this sort of belief about itself, so it will also have to form a range of other beliefs, if it is to act as

[1] The line of argument develops ideas used for other purposes in (List and Pettit, 2011, Ch 9) and (Pettit 2018b). It is very close to the picture developed by Alexandra Boyle (2018) in an insightful discussion of self-recognition in non-human animals.

it is designed to do. It will have to form beliefs to the effect that this or that object on the table is out of reach (of me), that it is to (my) right or left, that it is of the right size (for me) to lift, and so on. Does the fact that every agent has to form such beliefs about itself mean that it has a self in any sense? Surely not. Although it forms beliefs about itself, this simple sort of robot is little more than an impersonal mechanism, not something to which we could ever plausibly ascribe a self: it is a handy tool, not a colleague.

The beliefs that the robot must form about itself fail to argue for its having a self, plausibly, because they do not identify the self on which they bear. A belief will identify something for an agent when it picks it out from among other potential entities as a subject of which it predicates this or that property or relation. Thus, viewing it from within the intentional stance, we must credit the robot with identifying this or that glass as a subject of predication and then assigning a property or relation to it when it forms the belief that the glass is on its side.

The robot does not identify itself as a subject of predication in this manner, despite believing that it has certain relations to the objects on the table: that this or that glass is near to it or graspable by it, or whatever. The robot does not pick itself out from among other agents present or possible, and identifying itself among those candidates, form various beliefs about its relationships to things on the table. In positing those relationships, the robot's self-beliefs, as we may call them, lock onto itself; they do not first pick it out and then register that it, the subject identified, bears those relationships. The self-beliefs lock on to it as the only agent they could possibly apply to, and they apply to it of necessity.[2]

The robot's subject-predicate belief that a certain glass is on its side may err in either of two ways. Among the glasses on the table, it may misidentify the particular glass that constitutes the subject of the belief that it is on its side; a pulsating light may lead it to attribute that property to an upright glass nearby. Or, assuming that it has no problem in identifying the targeted subject, it may misattribute the property of being on its side; the glass may appear to be on its side due to its unusual balloon shape. Because its self-beliefs do not first identify itself and then attribute to it a certain relationship

[2] On David Lewis's (1983, Ch 10) account of *de se* belief, as he calls it, every belief involves the agent locating itself, whether in a possible world of a certain sort or at a particular place in such a world: say, near a glass on its side. The relationship between such a belief and the subject located will be of the locked-on kind described here. Thanks to Frank Jackson for drawing my attention to this.

with one or another object on the table, they cannot err on the first count. They may misattribute that relationship, taking an object to be nearer than it really is, for example. But they cannot misidentify the agent to whom the relationship is ascribed: in a case like this, as Gareth Evans (1982) puts it, there is no error through misidentification.

These observations invite us, by way of contrast, to consider an agent that is able to form identifying as well as locked-on beliefs about itself. An example would be an upgraded robot that is able, not just to form self-beliefs of the kind illustrated, but also to form beliefs that other entities are robots like itself, and in particular to form beliefs to the effect that it, the entity to which it has a locked-on, identification-independent relationship, is one of those robots. It can form beliefs that might warrant expression in first-person formulae like the following: I am this robot here, not that one there; I am the one currently moving towards the glass on the right, not the robot moving towards the glass on the left; I am the one who will reach its destination first; and so on.

We could reasonably say that this robot exists for itself in a way in which the earlier version did not. It connects in the same locked-on way with itself but identifies the agent with which it connects in that way as one among many possible candidates, and it treats the target of those two sorts of beliefs as one and the same. It exists for itself not just as the unmissable target of an indexical thought but also as this or that fallibly identified robot. It exists for itself, not just as a private reference point, in other words, but also as an entity located among other entities in a public world and characterized, like those entities, by various properties or relations.

It is plausible to suppose that an agent with a self, as distinct from an agent without a self, is at its most basic, an agent that exists for itself in this way: a *pour-soi*, in Sartre's (1958) terminology, not an *en-soi*. It is an agent that not only locates things in the public world, as presumably any agent must do, but that also identifies itself in that public world. Employing the private mode of self-reference available to any agent in the formation of beliefs and other attitudes, it identifies that agent – it identifies itself, so referred to – with a figure in the public world, characterized by public properties. It thinks thoughts of the form: that is me, that is how I am!

As we have described this upgraded robot, the beliefs it forms about itself are sophisticated in two ways. They are stimulus-independent beliefs insofar as they include beliefs beyond the here and now, as in the robot's believing that it will get to its destination sooner than others (Camp, 2009). And they are compositional

beliefs insofar it is true that any property the robot ascribes to itself, it can ascribe to others, and any property it ascribes to others it can apply to itself (Evans, 1982, 104). Might there be agents who can form a belief like 'That is me', but only in a non-compositional, stimulus-dependent way? The question bears on the self-recognitional capacities of those non-human animals who identify themselves in mirror reflections but we may put the issue aside in the present context (Boyle, 2018).

Personhood

Selfhood is associated with persons in particular and it's important that we introduce this category as well as that of agency. There is widespread agreement that not every agent is a person but little or no agreement on what it is that makes some agents into persons, others not. Some may hold by the traditional view, deriving from Boethius, that persons are agents of a reasoning character or nature – *rationabilis naturae* – or agents with similar, general capacities like the ability to communicate, to interpret others, to be aware of themselves in relation to others, and to enter reciprocal arrangements (Dennett, 1979).[3]

Many identify persons, however, not by their general capacities, but by specific, normatively relevant features. Thus, Locke (1975, s26) takes it that persons can assume and incur responsibilities, arguing that 'person' is 'a Forensick term appropriating Actions and their Merit'. John Rawls (2001, 23) argues that what distinguishes persons is that, of necessity, they can bear and assert rights: they are 'self-authenticating sources of moral claims'. And Harry Frankfurt (1971) identifies persons – as distinct from 'wantons' – as agents that engage the autonomy ideal of acting on desires that they desire to be moved by.

There is an approach to the characterization of persons that helps to explain the appeal of these different accounts and has a claim to our allegiance on that ground. In any case, it is the account ~~with which~~ we shall work with it here. Deriving from some brisk remarks by Thomas Hobbes (1994, Ch 16), it takes persons to be agents who have and exercise the capacity to 'personate' in relation to one another. Agents personate insofar as they assume a special authority

[3] Reasoning in this context may be taken to involve ratiocinative activity, not just the display of rationality in the formation of intentional states. For such a concept of reasoning, see (Pettit 1993, Ch 1) and (Broome, 2013).

in communicating their attitudes to others and, at least in general, live up to their words in practice: they act as the attitudes communicated would warrant. They speak for themselves, inviting others to trust their words: to rely on their displaying the profile or *persona* that they project in what they say (Pettit, 2008).[4]

The personation account entails the presence of the capacities that Boethius and Dennett and others ascribe to persons: agents can hardly speak for themselves in the manner envisaged without such abilities. But the account also explains the normative features that others take to be distinctive of persons. Connecting with Locke, it would make sense of why persons, inviting the reliance of others, must assume and incur responsibilities to prove reliable. Connecting with Rawls, it would explain why each person must have certain rights: the right to personate with others, for example, as well as the right to rely on others to live up to their personation. And, connecting with Frankfurt, it would make sense of why personhood implies an ideal, if not quite that which he envisages; here the associated ideal is that of displaying in actions the *persona* projected in words.[5]

What is it that makes people's communication of their attitudes into a form of personation? Why are persons said to assume authority for the communication, inviting others to rely on them? The key to the answer lies in the notion of commitment.[6]

We can define someone's being committed to holding by a certain attitude or to performing a certain action as their communicating that they have that attitude or that they will perform that action in a way that exposes them to a higher expense than normal – and so makes the communication more credible than normal – should they fail to act on the advertised disposition. Let the standard of normality be set by the expense they would incur if they merely reported on their disposition, as they might report on the disposition of a third person. Any communication of a current attitude or a projected action will be

[4] In support of his approach, Hobbes points out that the Latin *persona* refers to the mask through (per) which actors speak or sound (sonare) in presenting a figure in a play.

[5] It would also explain why, in Christine Korsgaard's (2009, 26) words: 'A good person is someone who is good at being a person'.

[6] I explicate the notion of commitment in what follows on lines set out at greater length in (Pettit, 2018a). But the notion, as I develop it, is indebted to a bunch of writers on the topic over the past couple of decades. On the general idea of commitment, I am enormously indebted to (McGeer, 1996; Moran, 1997; 2001; McGeer, 2008); and on the ideal of avowal, to (Bar-on 2004).

commissive insofar as it involves voluntarily incurring a higher risk of loss in the case of a miscommunication.

Suppose, unusually, that I were merely to report a belief or desire. Suppose that I said in a suitably tentative manner that I think I hold such and such a belief or desire or that I will perform such and such an action. Then, as with any report, I would be able to offer one of two salient face-saving explanations – one of two excuses – for failing to act accordingly. I might claim that I was misled by introspective evidence about my attitudes: my mind misled me. Or I might maintain that I changed my disposition since reporting it: I changed my mind.[7]

There are two prominent forms of commitment, avowals and pledges, that correspond to these excuses, and they figure prominently, as we shall see, in personation. The avowal of an attitude manifestly forecloses appeal to the misleading-mind excuse, thereby making the avowal more expensive and credible than a report. The pledge to perform an action – alternatively, the pledging of the intention to perform it – manifestly forecloses appeal to either the misleading-mind or the changed-mind excuse, making it more expensive and credible than even an avowal (Pettit, 2018a).

I will avow a belief that p, communicating the presence of that belief-state, if I just assert that p on the basis of having made up my mind, presumably in light of the evidence or data, that it is the case that p. The avowal will communicate, not just that p, but also that I believe that p, insofar as it is a matter of shared assumption that, absent insincerity or incompetence, anything I assert I also believe. But I do not learn that I have the avowed belief thereby communicated by introspecting the contents of my mind and relying on the evidence of what I find there; I do so by relying on the capacity manifested in making the assertion that p: the capacity to make up my mind. And so, the avowal manifestly precludes me from excusing a failure to act as if I had the belief by invoking the misleading-mind excuse. Not having been led by evidence about my mind, after all, I couldn't have been misled by it. Thus, the avowal of the belief will communicate the belief more expensively and more credibly than a detached report would have done.

As reporting on what I take to be the case will amount to avowing the belief that that is the case, so reporting on the presence of manifest

[7] In this presentation, I describe any explanation offered as a way of saving face, or more generally getting off the hook, as an excuse, ignoring the distinction between explanations that justify and explanations that excuse in a contrasting, narrower sense of the term.

desiderata in a certain option, reporting that it is the fairest or the most exciting alternative, for example, will generally constitute an avowal of that desire. I will form or reinforce the desire in fastening on the operative desiderata, letting them elicit or entrench the desire, as I will form or reinforce a belief in attending to the data that support it. By citing the desiderata, moreover, I will generally convey that I desire that option – why would I have cited its appealing feature otherwise? And I will convey this in a way that leaves me without access to the excuse, should I act as if I did not have the desire, that I must have been misled about its presence.[8] Not having been led by introspective evidence that it was present – knowing that it is present by virtue of knowing what I did in fastening on the desiderata – I cannot have been misled by such evidence. Thus, the communication will count as commissive in the same way as the avowal of a belief.[9]

I may avow an intention to do something, say to go to a concert with you, on the same desiderative basis as with a desire: 'that would be great fun', I may say, in response to a query about joining you at the concert. But in the case of an intention, and of the corresponding action, I may do something more. Recognizing that you won't go to the concert unless I join you, for example, and expecting to want to join you, I may make a pledge to be there. In doing so I will manifestly foreclose the changed-mind excuse as well as the misleading-mind excuse for having misled you, should I not turn up. You will naturally say 'But you promised!' if I try to excuse not turning up by saying that I changed my mind.[10]

Personation in broadly the Hobbesian sense involves commitments of the kind that avowal and pledging exemplify. When

[8] The background assumption here is that desires are linked with perceived desiderata such that it is intelligible that they should attract the agent, and unintelligible why an agent should be attracted to something in their absence. See (Anscombe, 1957).

[9] In communicating the desire in that way, of course, I will also communicate the belief that I hold the desire; on related matters see (Jackson and Pettit, 1998).

[10] I cannot pledge beliefs or desires, because I cannot guard against the change of data that might affect a belief, or the change of desiderata that might affect a desire: this, in the sense in which I maintain a desire for something only if I continue to like and desire it under the same desiderative aspect. While intentions are grounded, like desires, in the desiderata of what I come to intend, I will maintain that intention and act on it even if the desiderata change: even if the only desideratum remaining is that I said I would act in the corresponding way. See (Pettit, 2018a).

persons speak for their attitudes, they assume the authority that goes with their purported ability to communicate those attitudes while foreclosing the possibility of excusing a miscommunication by appeal to a misleading-mind mind or, where appropriate, to a changed-mind. To the extent that they put aside face-saving excuses for failure, ~~they~~ will stake their reputation on living up to their words; they will bet on themselves to prove reliable in that way.

Commitments of those kinds are common in social life, figuring prominently in conversational exchanges, whether about what is or might be the case or about what the participants should individually or jointly do. But commitments are even more common than this suggests, because commitments can assume a virtual as well as an active form. People will make commitments virtually when they fail to say 'Nay' to the manifest expectations of others; they will make them actively when they say 'Yea' in order to put novel expectations in place.

Suppose you and I live in a society, for example, where it is manifest to all that certain regularities hold and that everyone is expected in general to live up to them without question: this, for example, in avoiding violence, deception, fraudulence, infidelity and the like. I will know in the range of everyday interactions that others manifestly expect me, without any question, to accept those norms and to be willing to conform; indeed, I will also know that, should I fail, then appealing to excuses of the misleading- or changed-mind sort will not wash. This being a manifest matter between us, my not rejecting those expectations will communicate that I acquiesce. And by acquiescing those expectations, I will effectively avow acceptance of the norms and pledge conformity to them. Indeed, the same will be true across the spectrum of manifest, unrejected expectations that I and others form in dealing with one another.

Thus, by the account adopted here, persons are agents that personate in their relations with one another, and do so to the point where personation, active or virtual, is an inescapable aspect of their individual lives. As a byproduct of commissively communicating their dispositions – not, as an effort in narcissistic self-portrayal (Strawson, 2005) – they will each shape a *persona* or image of themselves that they project and generally seek to honor. They will hold out this *persona* to others, as if in proclaiming: this is who and how I am; this is who and how I back myself to be.

Personation is a social activity, by this account, and personhood can only materialize among a group of agents, not in the solitary

individual.[11] Hobbes (1994, 26.6) would seem to endorse this view, holding that it is impossible to make a commitment to yourself on the grounds that 'he that can bind, can release'. Even if we admit that the practice of commitment presupposes social life, however, induction in that practice will presumably enable persons to commit to themselves in an analogous, if not exactly similar way: to form resolutions, as we say. When I sincerely make commitments to others, indeed, there is a sense in which I will commit to myself at the same time; I will internalize the *persona* I project; I will think as well as say, 'this I who and how I am'. I may commit and personate insincerely in many social contexts, of course, and at a barely imaginable limit, I might even do this in all. I may make commitments that I do not internalize, in other words, impersonating a self that I do not have.[12]

By our earlier argument, agents may achieve self-identification and selfhood, and achieve it in a sophisticated fashion that is compositional and stimulus-independent, without being persons in the sense of personating agents. The self-identification of persons is bound to be similarly sophisticated: they will each be able to predicate various properties of themselves, whether they see themselves as 'I' or under a public identifier; they will be able to predicate those very same properties of others, when that is appropriate; and of course they will be able to do this in abstraction from current stimuli.

But apart from being sophisticated in that way, the self-identification of persons is going to be special in at least three respects. Self-identification in the personal case will characterize the self substantively, projecting a full attitudinal and practical profile. It will do this inescapably, as an essential part of what it is to personate. And it will do it as a matter of aspiration, not accomplished fact: it will hold out an idealized version of the figure with which the person identifies, albeit a version that they commit to realizing.

[11] It might be possible – in principle if not in practice – for a solitary individual to be able to self-impose a cost for not living up later to an attitude they enunciated, and thereby to give that earlier enunciation the character of a self-commitment.

[12] In this case I lie to others but let myself in on the lie. Is it possible to lie to myself, as in self-deception? Not perhaps in the strict sense of lying intentionally, but certainly in the sense of being negligently misleading. I may make some commitments sincerely, linking them with resolutions, in neglect of the fact that my record in keeping them is very poor and that they are likely to fail. That neglect may be epistemically culpable – I may know in my heart of hearts that I will fail – and it will certainly be morally culpable: I should attend to my record, plausibly, when inviting others to rely on me in certain ways.

3. My self in its three guises

Being a self, by the account sketched in the first half of this lecture, means being a subject that exists for itself under a certain profile: a subject that can identify itself – the unique agent with which it connects in a locked-on, identification-independent way – in other, public terms. Assuming it has the representational abilities of human beings, it will be able to practice this self-identification in a compositional, stimulus-independent fashion. And assuming that it has the personating capacity associated with persons, it will practice this self-identification in a substantive, inescapable and commissive mode.

If I am a self insofar as I exist for myself in a certain profile, then it is natural to say that if there are distinct, relatively independent profiles under which I may depict myself, then they correspond to distinct selves that I may bear. To treat such different profiles as distinct selves may threaten to reify them inappropriately but so long as we keep that danger in mind, it need not be a problem. The treatment is plausible, because there are indeed different profiles – different informational takes – under which I may identify myself, pointing to different answers I may give to the question of who and how I am. They represent images of me, one and the same person, but seen from distinct standpoints: roughly, ~~as we say,~~ those associated with the first, the second and the third person. We describe the first-person profile as the referenced self, the second-person as the personated self, and the third-person as the imputed self.

As there are three selves to discuss, so there are three questions we can raise about each of them. One is the metaphysical question as to what constitutes this self or character, another is the epistemological question as to how I know myself in this character, and a third is the practical question as to how far the character matters, or ought to matter, to me. We now look at the different selves, considering the three questions in relation to each. The treatment offered is inevitably sketchy but it may at least serve to identify the range of issues that need to be considered in a full investigation of the self.

The referenced self

Assuming that I am able to identify myself in public terms, and to do so in a sophisticated way – in a compositional, stimulus-independent manner – the most basic profile that I will be able to assign to myself under that aspect is as the agent referenced in my intentional states.

This is the agent whose experiences give me memories and beliefs, whose likes and dislikes lie at the origin of my desires, and whose intentions dispose me to perform corresponding actions. This agent is one among many subjects who populate the world and, if I ask myself who I am, then that agent provides one obvious answer to my question: this is who and how I am; this is me.

What is it about this referenced agent, to pose a metaphysical question, that makes it me? If I pose the question in the present about an agent in the past or indeed an agent in the future – if I ask about myself qua agent and not, for example, qua organism – then there is only one answer possible. This or that agent in the past or future is me myself now, just insofar as there are the familiar agential linkages between the three. The claim is not that I have those linkages with the past or present entity because they are me. Rather, they count as me – we count as the self-same individual over time – because those linkages obtain. What I am over time is constituted by the chained time-slices that connect in that way.

The chaining of those particular time-slices will be explained, no doubt, by all sorts of facts about my neural make-up; if my brain now were linked electronically with different brains in the past or the future, after all, then the path of the chain would be quite different. But there is an important sense, to use Derek Parfit's (1984) expression, in which there need be no deep fact about what gives me my self-identity – my identity as this referenced self – over time. Plausibly, the connections between the different links make it the case that I am present in each link and that, over time, I am constituted by the chain they form. There would be a deep fact about who I am, and about what binds my temporal stages together, only if things were the other way around: only if it was my presence at each linked point – my presence in some independent mysterious sense of *me* – that explained why the chain ran through just those locations.

If the metaphysics takes this form, and there is no deep fact about what makes me the referenced self I am, then various well-known science-fiction possibilities are open. I might survive as the same self in this sense, if I were tele-transported, for example, having my body reconstituted in duplicate form. And equally I, as I am now, might divide in the future, with distinct selves at the later time sharing a chain up to the moment of division, and with each referencing the links they have in common as they form relevant memories or act on relevant intentions. But I as a referenced self could hardly survive fusion with another, since that future agent would have conflicting connections into the past: it would have to serve the masters

provided by rival experiences, tastes, and intentions. And equally, in this particular identity, I could not survive the loss of agency that a vegetative state would imply.[13]

Moving to an epistemological question, how can I know myself as an enduring agent in this self or profile? As an enduring agent, I will be characterized, not by the states I happen to be in at any moment, but by the chain of connections with the states of past temporal stages and, presumptively, later stages in the future. Thus, there won't be much about this enduring self that will be revealed as I introspect at any moment. Things will be more or less as Hume (1978, I.6.3) famously describes them: 'when I enter most intimately into what I call *myself*, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch *myself* at any time without a perception, and never can observe anything but the perception'.[14]

But while I may not have much to learn about the nature of this referenced self in any form of direct introspection, there is still a sense in which it is bound to be accessible to me. This is the self to which I orientate, after all, when I try to remember a past experience, when I deliberate about what at some future point to do, and when I feel anxiety at the prospect of a visit to the dentist.[15] It contrasts with the agent to which I orientate when acting as a member of a group, for example, seeking to realize group goals according to group assumptions (Pettit, 2018b). It is a self I will see out of the corner of my eye, so to speak: in apperception, as we might call it, rather than perception. I know it, not by acquaintance and not by description, but as a self, referenced in all my agential adjustments. It is revealed to me in the way in which the viewpoint from which a landscape photograph was taken is revealed by a picture, despite that fact that it does not figure in the photograph itself.

Turning now to the practical question, how far does the referenced self matter to me, or how far ought it to matter? In the nature of the case, I am bound to care about getting the experiences of this self

[13] For a collection of pieces, classic and contemporary, on these and related issues in personal identity, as the topic is known, see (Perry, 2008). See also (Johnston, 2010).

[14] For a similar view, see Sartre's (1957) essay on The Transcendence of the Ego.

[15] Thus Georg Lichtenberg was mistaken to think that 'Cogito ergo sum' – 'I think, therefore I am' – conveys nothing more than 'Cogitatur, ergo id est': 'there is thought, so there is something'. The point is made by Bernard Williams (1978) in his study of Descartes.

right as I try to remember something, about keeping this self on an effective path in pursuing my goals, and about its not suffering too much in the visit to the dentist. But the observation needs to be qualified in a way that parallels a remark made in discussing the metaphysical question.

I do not care about my referenced self in this way because it is me in an independent sense. Rather, this self counts as me because it is the target of such care. If we can speak of self-love here, it is the innocuous form of self-love that Rousseau describes as *l'amour de soi,* as distinct from *l'amour propre*; it is a form of concern with self that he takes to be part and parcel of our nature (Dent, 1988).

Derek Parfit (1984) suggests that as I look further and further into the future in tracking the agent with whom I am linked in the fashion described, there will be fewer and fewer linkages between me now and that agent. He concludes that it may make little sense, then, for me to be moved prudentially by the interests of that agent: say, if I am young, by their interest in having a good superannuation to rely on. This does not follow with the sort of self-concern appropriate to the referenced self. Any connections of the relevant kind will mean that that future agent is me, even if those connections all go through intermediate stages: even if they do not include, as they surely may include, a direct connection like that established by an intention now that I enjoy a good superannuation then. And insofar as that future agent is me, however distant in time, it will attract my corresponding concern as a matter of necessity, not as a function of an optional – and, as it might seem, questionable – degree of prudential concern.

Imagine a future, then, in which I cease to exist properly as an agent. I enter a vegetative state and, while I retain my identity as an organism, I cease to be an agent and cease, *a fortiori*, to be the same agent as I am now. On the approach taken here, I cannot care in the agential way for that individual in the future. I can only care for that human being, as I might care for someone else, in a relatively altruistic and, as we might say in this case, a prudential manner.

The personated self

The personation that makes me into a person, by the account offered earlier, is an activity that I pursue as an agent, identified to myself in the first-person manner just sketched. But that activity itself consists in a form of self-identification and directs us to a distinct profile in which I may recognize myself: it yields a different self, as we may

say, that I, one and the same agent, may bear. This different self is that which I assume in my second-person relationships with others, and that which I self-ascribe when I am being sincere with my audience.

When I identify myself in personating or personal terms, I do not do so in the spirit of a self-reporter or autobiographer. I do not stand back from myself and describe for one or another audience – perhaps just for myself – the sort of figure I cut in the interpersonal world where I rely on others and invite them to rely on me. Such an autobiographical pen picture would carry no special authority. Indeed, given the bias or partiality that we each feel for ourselves, it might be especially suspect.

In writing as a self-reporter in this way, I could claim only the epistemic authority of someone determined to seek out all sources of evidence and to be responsive to the evidence mustered. In personation, however, I assume a distinct practical form of authority, claiming the capacity to commit only to attitudes that I can enact and to enact all those attitudes to which I commit. I stake my reputation for displaying this commissive-enactive capacity when I avow any beliefs, desires or intentions, putting aside the possibility of invoking a misleading-mind excuse for a miscommunication. And I do so in a yet more demanding way when I pledge an intention to act in one or another fashion, foreclosing appeal to either a misleading-mind or a changed-mind excuse for having been misleading.

In the case of either sort of commitment, of course, I may occasionally fail to enact the attitude to which I commit and there may even be un-foreclosed excuses that I can offer to save my reputation. I will be able to excuse a failure to act on an avowed belief or desire, or indeed an avowed intention, by a change of mind. And I will be able to excuse both a failure to act on an avowed attitude and a failure to execute a pledged intention by certain unforeseeable changes of circumstances: bereavement, accident, illness or whatever. When I have no excuse to offer for a failure, however, the only alternative will be an apology and, in token of sincerity, a renewal of the commitment. And if the failures become too frequent, I will jeopardize any claim to have the commissive-enactive capacity associated with personhood; I will begin to look like what Frankfurt calls a 'wanton' rather than a person.[16]

[16] I will stake my reputation for displaying a dual commissive-enactive capacity, only with attitudes that connect closely with action. I will risk little or nothing in avowing a highly specific degree of belief – say, a belief to degree 0.745 that p – since it is hard to imagine real-world actions that

The profile of myself that I present here constitutes a personated self that is distinct from the referenced self. Assuming sincerity and competence, this is a self that I will actually prove to bear. And assuming sincerity and competence, it will have to be a relatively unified self, not one committed in inconsistent ways. I may personate on different fronts with different audiences but I cannot personate sincerely – I cannot personate for myself, so to speak – in diverging, unreconcilable ways.

The personated self is metaphysically unproblematic. It is a character or *persona* that I create for myself in the exercise of commissive-enactive competence. It is grounded, then, in two aspects of my performance in relation to others – and to myself considered as if I were another. First, in the things I say, or the things I let go as not needing to be said, in speaking for myself. And second, in the things I do in giving life to the character to which my speech testifies. Where the referenced self is guaranteed to exist by the exigencies of agency, the personated self is something I construct to begin with and reconstruct ~~as~~ occasion demands.

To return to a point made earlier, however, this self is not a construct that I intend to create as such. Some theorists hold that human beings actively construct a narrative of their lives and of who they are, and that this is an inevitable aspect of personhood. That claim ties personhood, implausibly, to a highly intellectualized form of reflection and a pattern of self-scripting that sounds downright narcissistic, as critics have suggested (Strawson, 2005). But the idea here is not open to such objections, as the character that each of us is said to construct for others, and indeed for ourselves, is a byproduct –generally, we may presume, an unnoticed byproduct – of an independently intelligible practice of commitment.

That I construct and reconstruct my personated self in this manner does not preclude my deconstructing it, of course, as when I disown certain attitudes that I previously avowed or some intentions and actions that I previously pledged. Given changes in the data or desiderata I access, or in my appreciation of them, it is entirely intelligible why I should change beliefs or desires I previously avowed. And

---

would show that I did not have precisely that degree of confidence. And I will risk little or nothing in pledging a highly conditionalized intention – an intention, should such and such conditions materialize, to do so and so – if the conditions are unlikely ever to be realized. But in interaction with others, I will generally speak for myself more plainly and will engage my reputation more directly. See (Pettit, 2018a).

given unforeseen circumstances, or transformative experiences (Paul, 2014), it is equally intelligible why I should change pledged intentions or policies as well.

Turning from the metaphysical to the epistemological issue, how do I know the personated self I assume and display in the manner described? The answer, in brief, is: by means of a maker's, not an observer's, knowledge. In exercising my commissive-enactive authority, I make it the case that I commit only to attitudes I can enact and that I enact all the attitudes to which I commit. Insofar as I know what I am doing in such an exercise I will be in a position to know the personated self I project, although I may not be conscious of it as such.

The exercise of my commissive-enactive competence may take the form of active avowals or pledges or of acquiescence in the manifest expectations of others. But in either case the knowledge of what I am doing will give me a base for knowing who in the personal sense I am; it will be capable of revealing my personated character to me. The problem of how I know what I am doing in any activity is a recognized problem, of course, but assuming it is soluble in the general case, it will yield a solution to the specific epistemological issue raised by my personated self.

But while my personated self will be knowable to me for these reasons, ~~it is important that~~ it may take effort to achieve a full knowledge of who and how in this sense I am. That will require a unified sense of the different fronts on which I am committed and of the package that those commitments constitute. Thus, it may take time and trouble for me to develop such a sense of where I am committed. There is going to be a point, therefore, to the exhortation in this case to know yourself; I may have to practice a degree of discipline and meditation to achieve self-knowledge in this sense.

Finally, to the practical issue. Ought I to care about myself in this personal guise? And am I likely in any case to do so? The answer in each case is affirmative.

I cannot help but care about living up to the commitments I make sincerely to others, and the self I thereby project. My reputation as someone others can rely on, after all – indeed my reputation as someone I can myself rely on – depends on my displaying such fidelity. For me to make commitments sincerely without any care for whether I proved reliable would be impossible; it would undermine the very notion of commitment.

But not only would it be unlikely that I should not have any care for keeping my commitments, and remaining faithful to my personated self. Such a lack of care would also be undesirable both for me and for others. It would make it impossible for others to be able to rely

on me, and impossible for me to be able to elicit their reliability in return. It would make for a loss on all sides.

The case for being faithful or true to the personated self over time is memorably put by Shakespeare, of course, in words he ascribes to Polonius, when bidding farewell to Laertes, his son. Polonius may be presented in an otherwise unflattering light by Shakespeare but he is surely credited with wisdom, when he says to Laertes:

> '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'.

This is wisdom, at any rate, if Polonius has the personated self in view. To be true to the personated self is just to live up to active and virtual commitments, letting your words and actions testify to the same *persona*. And, as a result of that convergence, it is to speak reliably to others, avoiding falsehood or duplicity.

One final comment, however, on the personated self. The fact that I unify this self, know it properly and care for it appropriately does not mean that it is morally admirable. For all that is required by the account offered, I might assume and enact the identity of a Nietzschean *Übermensch*, who treats certain others with disdain and makes serious commitments only to an elite of perceived equals. The ideal of the personated self is a structural ideal, not an ideal of a substantive kind. It teaches a lesson about how I should be – unified, self-knowing and stable – regardless of the character that I actually have.

The imputed self

In personation I am invested in establishing a persona in the minds of others that, when I personate sincerely, I can identify with: I can see as me. But for each of us there is a character that is not sculpted in personation, or that is sculpted only partially in that manner. This is the image of me that exists in the minds of others as they view me at a distance, so to speak, beyond the reach or control of my second-person avowals and pledges. It is a character created by third-person gossip about my attitudes and dispositions, by the available record of my achievements and failures, by the labels and stereotypes under which I am seen in our society and, if I am exposed to publicity in any domain or degree, by what becomes accepted as a matter of common or public assumption about me.

This character, alien and uncontrolled, may not seem like a candidate for being seen as me. But this is the self, known under my public name, that those with whom I do not regularly interact will take to be me. And since the way those others treat me will reflect the character that they ascribe, I cannot detach myself wholly from this identity; I cannot pretend it is someone else. Whatever the self imputed to me, of course, I may take different attitudes towards it: I may relish it, I may rail at it, or I may regard it with relative indifference. But even when I am indifferent, ~~however,~~ I cannot ignore it completely; it follows me as a shadow in social space, whether I like it or not.

The inescapability of this imputed self is nicely captured by Jorge Luis Borges (1962, 246–47) in a short essay, entitled 'Borges and I'. Seeing himself cast as a result of his publicity in a *persona* that is created by others as much as by himself, he comments: 'The other one, the one called Borges, is the one things happen to'. While this Borges is beyond his control, it is not a figure he resents. 'It would be an exaggeration', he reports, 'to say that ours is a hostile relationship; I live, let myself go on living, so that Borges may contrive his literature, and this literature justifies me'.

Where Borges displays an amused indifference to his imputed self, the protagonist in Jean Paul Sartre's (1948) essay 'Portrait of the Anti-semite' represents someone who relishes it. This young man, unsure of his identity, is taken to be hostile to Jews, perhaps quite mistakenly, and finds that the image impresses acquaintances and prompts them to make allowances for his attitudes. Enjoying that recognition, then, he comes to cherish the image in which he finds himself cast. He will do this, Sartre suggests, not necessarily because the image appeals independently, but because it makes him into a somebody: it gives him bearings by which to navigate and rescues him from a sea of existentialist indecision. The lesson is a general one. Each of us is an unconstrained center of decision-making, not someone predestined or predetermined to have certain attitudes or to act in certain ways. And as a result, we are each subject to the temptation to espouse in bad faith any character we find imputed to us.

But the self imputed to us, whether we like it or not, may also be one we resent and reject. If I belong to a disadvantaged group, for example, and am cast in a stereotype of a religious, ethnic or gendered kind, then I am extremely likely to resist that aspect of the character foisted upon me. And equally, if I am defamed – if, at the limit, a negative characterization is imposed on me, as a matter of common awareness in my community –then I must shrink from the character imputed.

In either case, but especially, in the stereotyping one, the image pinned on me may not only run counter to how I personate sincerely and reliably; it may silence any attempts to correct the image imposed and to change the expectations raised (Langton, 2009). It may deprive me of the capacity to speak for myself in a way that commands the credence of others, discounting or recasting the sorts of things I might say. It may compound the injustice of the misrepresentation, in other words, with the epistemic injustice of disabling me from putting it right (Fricker, 2007).

The fact that others invariably give us a character means that there is a third sort of self in which I or anyone else may be invested. This imputed self does not have the same interest as the referenced or personated self but it is still worth putting on the map. While the metaphysical and epistemological questions it raises are not very significant, there are real issues about why it matters, and about how far it ought to matter, to us.

Metaphysically, the imputed self is constituted by the opinion in which others hold me, independently of how I personate, and will appear in as many fractured and unfriendly forms as such opinion assumes. Insofar as I sincerely personate, and want my imputed self to conform to its personated counterpart, this will give me an interest in unifying the imputed self and bringing it into line with how I personate. But there may be strict limits ~~on~~ how far I can hope to achieve that result, creating a single image on this front of who and how I am.

Epistemologically, the imputed self is bound to be elusive, being constituted by the independent opinions that people hold of me. I may be painfully aware of a prejudicial category in which I am cast, or of some personally defaming gossip. Or if I am a public figure, I may take pleasure in the positive things that are said about me. But I will generally find it hard to gauge the opinion in which I am held by others. While people may gossip among themselves about me, after all, they are unlikely to gossip to me about myself. If someone is a friend, they may tell me what others are saying about me, whether to congratulate me about the good that is said or to warn me about the bad. But gossip proper belongs essentially to the third-person perspective on my performance rather than to the second-person relationship in which I can personate with my interactants.

Metaphysically fractured and epistemologically elusive though it is, however, my imputed self is bound to matter to me. And with good reason. Like others, I am likely to care about how people think of me, this being part of our social nature. But even if I don't

worry about their opinion of me as such, I will certainly care about how that opinion leads them to treat me. Thus, I have good instrumental reason for caring about the self imputed, when there is any danger of being individually defamed or of being cast with others under a prejudicial label. And equally, I have good instrumental reason to care, if I occupy a public position – say, a political office – and success in my sphere of activity, electoral or otherwise, depends on how I am viewed beyond the bounds where I may hope to personate.

Assuming that we can put such special cases aside, however unrealistic the assumption may be, is there good reason in general to care about my imputed self? The question is not whether there is reason, ~~when I have evidence,~~ to welcome the imputation of a positive character or to bemoan the imputation of a negative; clearly, the answer there is, yes. The question rather is whether there is reason, first, to desire to shape the imputed, un-personated self and, second, to try to shape it by filtering or massaging the information available about me – by running a personal public-relations exercise. By long and solid tradition, the answer to both questions is, no. The standard wisdom is that~~, on the contrary,~~ there are reasons against trying to shape this self, beyond the bounds where personation is possible, and reasons against even wanting to make such an attempt.

The reasons against trying to shape the imputed self in a personal public-relations exercise derive, at base, from the fact that this will put me in zero-sum competition with others. Whatever the domain of performance, I will want to establish how well I do in comparison with others, not on some absolute scale; after all, it is the performance of people generally that will set the standard in any domain (Brennan and Pettit, 2004). This means that trying to shape ~~up~~ my imputed image may put me in fruitless competition with others. We may each make costly efforts to improve our relative standing when those efforts may cancel out one another, leaving us individually no better off than we were before we started.

The reasons against even desiring or wanting to make such efforts – the reasons against caring, in that sense, about the imputed self – are many. One is that indulging that concern may distract me from focusing on something intuitively much more important: the self that I forge for others and for myself in personation. A second is that it may be impossible ever to tell if that concern is satisfied: the epistemic elusiveness of the imputed self means that trying to see if it is in an appealing shape may be as fruitless as trying to paint a picture of someone in the dark. And a third consideration is that if the concern is to be satisfied, I had better keep it hidden from others;

no one is going to achieve a high standing in the opinion of their peers if, beyond the context where defamation or prejudice threatens, they are seen to be concerned to achieve such standing. 'The general axiom in this domain', as Jon Elster (1983, 66) says, 'is that nothing is so unimpressive as behaviour designed to impress'.[17]

A final reason against even wanting to try to shape the imputed self is that it is an inherently insatiable concern: it can drive me to seek, not just to do well by local standards, but to outscore all others, and by as much as possible. As Rousseau (2020, 232) argued, following Hobbes, 'the first feeling excited by this comparison is the desire to be first'. It is this love of being first, this love of pre-eminence, that Rousseau castigated as *l'amour propre*, as distinct from the innocent *l'amour de soi*. Seeing such self-love as a quest for supremacy, Kant (2006, 167) also condemned it, finding it present in 'the manias for honor, dominance, and possession': in these manias, he said, 'the human being becomes the dupe'.

Conclusion

We began by noting the conflict and inconsistency in the various mantras associated with the self: to be yourself and to let go of the self; to know yourself and to forget about the self; to be true to yourself and to avoid self-concern. With the distinctions generated in the course of discussion, it is possible to find a way through this maze of confusing advice.

The self you should be, the self you should know and the self you should be true to is surely the personated self, as we have characterized it. This self becomes prominent on the view of agency and personhood sketched in the first part. If I am to be an agent that personates, exercising a capacity to make and enact commitments to others, then the self I project should be a self I realize, a self I try to know, and a self I take as an ideal: a self, in Polonius's words, to which I am true.

That self contrasts naturally with the imputed, un-personated self that puts me in zero-sum, potentially profitless, competition with

[17] See (Brennan and Pettit, 2004) for an argument that this observation does not rule out the possibility of an economy of esteem. The core argument is that people may be rewarded by esteem and penalized by disesteem, and may be reinforced or inhibited as a result, without ever seeking esteem or shunning disesteem in a strategic manner. They may be deeply subject to the influence of esteem without being guided by the desire for esteem.

others. This is a self that may distract from a focus on the personated self, that lies beyond what I can effectively control or reliably know, and that I cannot openly pursue without undermining that very enterprise. The concern for this self may also prove insatiable, in the manner of a mania, as Kant puts it. Plausibly, it is this self that I should let go, this self that I should forget, and this self that I should have little or no concern with – no concern, at any rate, beyond that required to counter defamation or to join with others in combatting prejudice.

What, finally, of the referenced self? This is the self in perhaps its most intriguing guise. It exists, to be sure, but in the manner of a networked sequence of links in a chain, not an underlying substance. It is there at every juncture to command the focus of the agent but eludes any attempt to grasp its nature introspectively. And while it commands the care and concern of the agent, even as it is imagined into the far future, it commands this as an exigency of agency itself, not as an attitude that might be voluntarily assumed. Where the personated self identifies a commanding ideal, and the imputed self represents the voice of a siren, the referenced self lies deeper than normative concern, in the very constitution of our nature.

References

G. E. M. Anscombe, *Intention* (Oxford: Blackwell, 1957).

D. Bar-on, *Speaking My Mind: Expression and Self-knowledge,* (Oxford: Oxford University Press, 2004).

J. L. Borges, *Labyrinths: Selected Stories and Other Writings*, (New York: New Directions, 1962).

A. Boyle, 'Mirror Self-Recognition and Self-Identification', *Philosophy and Phenomenological Research* 97 (2018), 284–303.

G. Brennan and P. Pettit, *The Economy of Esteem: An Essay on Civil and Political Society,* (Oxford: Oxford University Press, 2004).

J. Broome, *Rationality through Reasoning,* (Oxford: Wiley Blackwell, 2013).

E. Camp, 'Putting Thoughts to Work: Concepts, Systematicity, and Stimulus-Independence', *Philosophy and Phenomenological Research* 78 (2009) 275–311.

D. Dennett, *Brainstorms,* (Brighton: Harvester Press, 1979).

D. Dennett, *The Intentional Stance*, (Cambridge, Mass.: MIT Press, 1987).

N. J. H. Dent, *Rousseau*, (Oxford: Blackwell, 1988).

J. Elster, *Sour Grapes,* (Cambridge: Cambridge University Press, 1983).

G. Evans, *The Varieties of Reference,* (Oxford: Oxford University Press, 1982).

H. Frankfurt, 'Freedom of the Will and the Concept of a Person', *Journal of Philosophy* 68 (1971), 5–20.

M. Fricker, *Epistemic Injustice: Power and the Ethics of Knowing,* (Oxford: Oxford University Press, 2007).
T. Hobbes, *Leviathan*, ed E. Curley, (Indianapolis: Hackett, 1994).
D. Hume, *A Treatise of Human Nature,* (Oxford: Oxford University Press, 1978).
S. Hurley, *Consciousness in Action,* (Cambridge, Mass.: Harvard University Press, 1998).
F. Jackson, and P. Pettit, 'A Problem for Expressivism', *Analysis* 58 (1998), 239–51; reprinted in F. Jackson, P. Pettit and M. Smith, *Mind, Morality and Explanation*, (Oxford: Oxford University Press, 2004).
M. Johnston, *Surviving Death,* (Princeton: Princeton University Press, 2010).
I. Kant, *Anthropology from a Pragmatic Point of View,* (Cambridge: Cambridge University Press, 2006).
C. Korsgaard, *Self-Constitution: Agency, Identity, and Integrity,* (Oxford: Oxford University Press, 2009).
R. Langton, *Sexual Solipsism: Philosophical Essays in Pornography and Objectification,* (Oxford: Oxford University Press, 2009).
D. Lewis, *Philosophical Papers Vol 1,* (Oxford: Oxford University Press, 1983).
C. List and P. Pettit, *Group Agency: The Possibility, Design and Status of Corporate Agents,* (Oxford: Oxford University Press, 2011).
J. Locke, *An Essay Concerning Human Understanding,* (Oxford: Oxford University Press, 1975).
V. McGeer, 'Is "Self-knowledge" an Empirical Problem? Renegotiating the Space of Philosophical Explanation', *Journal of Philosophy* 93 (1996), 483–515.
V. McGeer, 'The Moral Development of First-Person Authority', *European Journal of Philosophy* 16 (2008), 81–108.
R. Moran, 'Self-Knowledge: Discovery, Resolution, and Undoing', *European Journal of Philosophy* 5 (1997), 141–61.
R. Moran, *Authority and Estrangement: An Essay on Self-knowledge,* (Princeton: Princeton University Press, 2001).
D. Parfit, *Reasons and Persons,* (Oxford: Oxford University Press, 1984).
L. A. Paul, *Transformative Experience,* (Oxford: Oxford University Press, 2014).
J. Perry, 'The Essential Indexical' *Nous* 13 (1979), 3–21.
J. Perry, Ed. *Personal Identity*, (Berkeley: University of California Press, 2008).
P. Pettit, *The Common Mind: An Essay on Psychology, Society and Politics,* (New York: Oxford University Press, 1993).
P. Pettit, *Made with Words: Hobbes on Language, Mind and Politics,* (Princeton: Princeton University Press, 2008).
P. Pettit, *The Birth of Ethics: Reconstructing the Role and Nature of Morality,* (Oxford: Oxford University Press, 2018a).

P. Pettit, 'Consciousness Incorporated', *Journal of Social Philosophy*, 49 (2018b) 12–37.
J. Rawls, *Justice as Fairness: A Restatement,* (Cambridge, Mass: Harvard University Press, 2001).
J. J Rousseau, *Emile* (New York: Public Domain Books, Kindle ed, 2020).
J. P. Sartre, *Portrait of the Anti-Semite*, (London: Secker and Warburg, 1948).
J. P. Sartre, *The Transcendence of the Ego: An Existentialist Theory of Consciousness*, (New York: Farrar, Straus and Giroux, 1957).
J. P. Sartre, *Being and Nothingness,* (London: Methuen, 1958).
R. C. Stalnaker, *Inquiry,* (Cambridge, Mass.: MIT Press, 1984).
G. Strawson, *Against Narrativity*, (Oxford: Blackwell, 2005).
B. Williams, *Descartes* (Harmondsworth: Penguin, 1978).

PHILIP PETTIT *is Professor of Human Values in Princeton and Professor of Philosophy in the Australian National University. His recent books include* Just Freedom *(Norton 2014),* The Robust Demands of the Good *(OUP, 2015) and* The Birth of Ethics *(OUP 2018).*