

D\]cgcd\ m

Appears in: Philip Pettit, 'My Three Selves', *Philosophy*, Vol 95, 2020, 363-389 (Royal Annual Philosophy Lecture, London, 2019)

8UH'cZXY`j Yfm %&!\$)!&\$&\$

>ci fbU`UbX`j c`#UfhjVW`fYZ

d\]

D< =&\$\$\$\$%#

Bi a VYf`cZdU[Yg`fbch]bVM X]b[`h`]g`dU[YL` &+

page 1 of 2

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

5 i h c f g `U f Y g l f c b [` m U X j] g Y X ` l c ` f Y U X ` h Y g Y d f c c Z g ` h c f c i [\ ` m V Y W U g Y U b m Y f f c f g ` a] g g Y X a U m U d d Y U f ` j b ` h Y Z] b U ` d i V ` g l Y X ` d U d Y f ` h] g ` k] ` V Y ` n e i f ` C B @ M W U b W Y ` l c ` W e f f Y W h i n c i f d f c c Z ` C b W ` d i V ` g l Y X Z Y] h Y f ` c b `] b Y ` c f `] b ` d f] b l z b c ` Z f h Y f ` W U b [Y g ` W U b ` V Y ` a U X Y "

†D`YUgY`fYhi fb`nci f`dfccZg`k]h]b" `k cf_]b[`XUng`†

<CK`HC`F9HI`FB`MCI`F`DFCC:G

Proofs **AI GH** be sent to Assistant Editor, Matt Hewson

Email: **Ugg]ghUbh# fcnU]bgh]hi hYd\]cgcd\ m'cf[**

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 Y`YWfcb]WU`m'

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 **BC** 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.

D\]cgcd\ m

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 **cbY** 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 **d\]dfcXi W]cb4 W]a Vf]X[Y'cf[**

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.

⇒ **a dcfHUbh** you must return any forms included with your proof"

D`YUgY`Xc`bch`fYd`mh`c`h`]g`Ya`U]`

NOTE - for further information about **>ci fbUg`DfcXi W]cb** please consult our : **5 Eg** at http://journals.cambridge.org/production_faqs

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, Princeton, NJ 08544, USA

Institution/Company Princeton University

Signature Philip Pettit

Date 17 May 2020

(Additional authors' names, affiliations and nationalities 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.

Definitions²⁵

| | |
|---|---|
| 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
Form JCT 14.1 and Doc.JCT.T&C.HSS.14.1

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 to 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.

Philip Pettit

44 describing the first as the referenced self, the second as the personated
45 self, and the third as the imputed self. The first part provides essential
46 background, while the second presents the paper's headline claim. A
47 short conclusion summarizes the overall view.
48

50 2. Agents, selves and persons

51 *Agency*

52
53
54 By all accounts an agent is a system that acts to bring about certain
55 goals or purposes. These are states of affairs that it generates on a re-
56 liable basis, and not just by accident: on the basis, for example, of
57 having been selected or designed to realize the goals or to follow
58 certain procedures that identify goals to realize. But agency requires
59 more than the pursuit of purposes. Otherwise the sunflower plant
60 that tracks the sun – that reliably moves so as to maintain its orienta-
61 tion towards the sun – would count, implausibly, as an agent.

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

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

83 In order for the robot to adjust its behavior appropriately in differ-
84 ent scenarios –and count thereby, we may suppose, as an agent – it
85 must be able in some sense to register the particularities of each scen-
86 ario. It must change in response to receptor inputs, and its changed

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

92 The general lesson, in familiar Humean terms, is that an agent is a
93 system that reliably acts to fulfill its goal-seeking states according to
94 representations that it reliably forms: in perhaps a deflationary
95 sense of those terms, it acts to satisfy its desires according to its
96 beliefs. The agent's reliability in forming beliefs constitutes epi-
97 stemic rationality, its reliability in acting so as to satisfy desires con-
98 stitutes practical rationality. The system may not be unfailingly
99 reliable or rational but, if it is to count properly as an agent, then it
100 must generally fail only when circumstances are abnormal by inde-
101 pendent criteria. Thus, the robot may be misled under certain light-
102 ing about the position of an object; and it may not manage to put an
103 object at the edge of the table upright: it may knock it to the floor. In
104 the first case, it would fail epistemically, in the second practically.

105 This simple image of agency is not uncontroversial but it is at least
106 familiar, being supported with variations by recent philosophers like
107 David Lewis (1983), Robert Stalnaker (1984), and Daniel Dennett
108 (1987). Others may place demands on agency over and beyond the re-
109 quirements it encodes: they may demand capacities, for example, asso-
110 ciated here with personhood. But still, it represents a model of agency
111 that is an intelligible reconstruction of our commonsense assumptions,
112 if not the only reconstruction admissible (Pettit 1993, Ch 1).

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

124 125 126 *Selfhood*

127
128 Every agent, by virtue of what agency involves, treats itself as special.
129 It acts on the basis of the memories and beliefs that its own

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

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

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

166 As even the simple robot will have to form this sort of belief about
167 itself, so it will also have to form a range of other beliefs, if it is to act as
168

169 ¹ The line of argument develops ideas used for other purposes in (List
170 and Pettit, 2011, Ch 9) and (Pettit 2018b). It is very close to the picture de-
171 veloped by Alexandra Boyle (2018) in an insightful discussion of self-recog-
172 nition in non-human animals.

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

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

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

200 The robot's subject-predicate belief that a certain glass is on its side
201 may err in either of two ways. Among the glasses on the table, it may
202 misidentify the particular glass that constitutes the subject of the
203 belief that it is on its side; a pulsating light may lead it to attribute
204 that property to an upright glass nearby. Or, assuming that it has
205 no problem in identifying the targeted subject, it may misattribute
206 the property of being on its side; the glass may appear to be on its
207 side due to its unusual balloon shape. Because its self-beliefs do not
208 first identify itself and then attribute to it a certain relationship
209

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

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

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

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

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

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

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

268 *Personhood*

270
271 Selfhood is associated with persons in particular and it’s important
272 that we introduce this category as well as that of agency. There is
273 widespread agreement that not every agent is a person but little or
274 no agreement on what it is that makes some agents into persons,
275 others not. Some may hold by the traditional view, deriving from
276 Boethius, that persons are agents of a reasoning character or nature –
277 *rationabilis naturae* – or agents with similar, general capacities like
278 the ability to communicate, to interpret others, to be aware of them-
279 selves in relation to others, and to enter reciprocal arrangements
280 (Dennett, 1979).³

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

291 There is an approach to the characterization of persons that helps to
292 explain the appeal of these different accounts and has a claim to our
293 allegiance on that ground. In any case, it is the account ~~with which~~
294 we shall work with it here. Deriving from some brisk remarks by
295 Thomas Hobbes (1994, Ch 16), it takes persons to be agents who
296 have and exercise the capacity to ‘personate’ in relation to one
297 another. Agents personate insofar as they assume a special authority
298

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

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

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

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

324 We can define someone's being committed to holding by a certain
325 attitude or to performing a certain action as their communicating that
326 they have that attitude or that they will perform that action in a way
327 that exposes them to a higher expense than normal – and so makes the
328 communication more credible than normal – should they fail to act on
329 the advertised disposition. Let the standard of normality be set by the
330 expense they would incur if they merely reported on their disposition,
331 as they might report on the disposition of a third person. Any com-
332 munication of a current attitude or a projected action will be
333

334
335 ⁴ In support of his approach, Hobbes points out that the Latin *persona*
336 refers to the mask through (per) which actors speak or sound (sonare) in pre-
337 senting a figure in a play.

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

340 ⁶ I explicate the notion of commitment in what follows on lines set out
341 at greater length in (Pettit, 2018a). But the notion, as I develop it, is indebted
342 to a bunch of writers on the topic over the past couple of decades. On the
343 general idea of commitment, I am enormously indebted to (McGeer,
344 1996; Moran, 1997; 2001; McGeer, 2008); and on the ideal of avowal, to
(Bar-on 2004).

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

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

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

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

381 As reporting on what I take to be the case will amount to avowing
382 the belief that that is the case, so reporting on the presence of manifest
383

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

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

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

414 Personation in broadly the Hobbesian sense involves commit-
415 ments of the kind that avowal and pledging exemplify. When
416

417 ⁸ The background assumption here is that desires are linked with per-
418 ceived desiderata such that it is intelligible that they should attract the
419 agent, and unintelligible why an agent should be attracted to something in
420 their absence. See (Anscombe, 1957).

421 ⁹ In communicating the desire in that way, of course, I will also commu-
422 nicate the belief that I hold the desire; on related matters see (Jackson and
423 Pettit, 1998).

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

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

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

448 Suppose you and I live in a society, for example, where it is
449 manifest to all that certain regularities hold and that everyone is ex-
450 pected in general to live up to them without question: this, for
451 example, in avoiding violence, deception, fraudulence, infidelity
452 and the like. I will know in the range of everyday interactions
453 that others manifestly expect me, without any question, to accept
454 those norms and to be willing to conform; indeed, I will also
455 know that, should I fail, then appealing to excuses of the mislead-
456 ing- or changed-mind sort will not wash. This being a manifest
457 matter between us, my not rejecting those expectations will
458 communicate that I acquiesce. And by acquiescing those expecta-
459 tions, I will effectively avow acceptance of the norms and pledge
460 conformity to them. Indeed, the same will be true across the spec-
461 trum of manifest, unrejected expectations that I and others form in
462 dealing with one another.

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

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

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

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

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

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

509 ¹² In this case I lie to others but let myself in on the lie. Is it possible to
510 lie to myself, as in self-deception? Not perhaps in the strict sense of lying in-
511 tentiously, but certainly in the sense of being negligently misleading. I may
512 make some commitments sincerely, linking them with resolutions, in
513 neglect of the fact that my record in keeping them is very poor and that
514 they are likely to fail. That neglect may be epistemically culpable – I may
515 know in my heart of hearts that I will fail – and it will certainly be
516 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.

Philip Pettit

560 This is the agent whose experiences give me memories and beliefs,
561 whose likes and dislikes lie at the origin of my desires, and whose in-
562 tentions dispose me to perform corresponding actions. This agent is
563 one among many subjects who populate the world and, if I ask myself
564 who I am, then that agent provides one obvious answer to my ques-
565 tion: this is who and how I am; this is me.

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

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

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

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

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

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

633 Turning now to the practical question, how far does the referenced
634 self matter to me, or how far ought it to matter? In the nature of the
635 case, I am bound to care about getting the experiences of this self
636

637
638 ¹³ For a collection of pieces, classic and contemporary, on these and
639 related issues in personal identity, as the topic is known, see (Perry, 2008).
640 See also (Johnston, 2010).

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

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

646 right as I try to remember something, about keeping this self on an
647 effective path in pursuing my goals, and about its not suffering too
648 much in the visit to the dentist. But the observation needs to be qual-
649 ified in a way that parallels a remark made in discussing the metaphys-
650 ical question.

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

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

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

681
682 *The personated self*
683

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

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

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

701 In writing as a self-reporter in this way, I could claim only the epi-
702 stemic authority of someone determined to seek out all sources of evi-
703 dence and to be responsive to the evidence mustered. In personation,
704 however, I assume a distinct practical form of authority, claiming the
705 capacity to commit only to attitudes that I can enact and to enact all
706 those attitudes to which I commit. I stake my reputation for display-
707 ing this commissive-enactive capacity when I avow any beliefs,
708 desires or intentions, putting aside the possibility of invoking a mis-
709 leading-mind excuse for a miscommunication. And I do so in a yet
710 more demanding way when I pledge an intention to act in one or
711 another fashion, foreclosing appeal to either a misleading-mind or a
712 changed-mind excuse for having been misleading.

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

727
728 ¹⁶ I will stake my reputation for displaying a dual commissive-enactive
729 capacity, only with attitudes that connect closely with action. I will risk
730 little or nothing in avowing a highly specific degree of belief – say, a belief
731 to degree 0.745 that p – since it is hard to imagine real-world actions that

Philip Pettit

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

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

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

761 That I construct and reconstruct my personated self in this manner
762 does not preclude my deconstructing it, of course, as when I disown
763 certain attitudes that I previously avowed or some intentions and
764 actions that I previously pledged. Given changes in the data or desid-
765 erata I access, or in my appreciation of them, it is entirely intelligible
766 why I should change beliefs or desires I previously avowed. And
767

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

775 given unforeseen circumstances, or transformative experiences (Paul,
776 2014), it is equally intelligible why I should change pledged inten-
777 tions or policies as well.

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

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

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

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

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

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

Philip Pettit

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

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

825 ‘This, above all: to thine own self be true,
826 And it must follow, as the night the day,
827 Thou canst not then be false to any man’.

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

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

844 845 846 *The imputed self*

847
848 In personation I am invested in establishing a persona in the minds
849 of others that, when I personate sincerely, I can identify with: I can
850 see as me. But for each of us there is a character that is not sculpted in
851 personation, or that is sculpted only partially in that manner. This is
852 the image of me that exists in the minds of others as they view me at a
853 distance, so to speak, beyond the reach or control of my second-
854 person avowals and pledges. It is a character created by third-
855 person gossip about my attitudes and dispositions, by the available
856 record of my achievements and failures, by the labels and stereo-
857 types under which I am seen in our society and, if I am exposed
858 to publicity in any domain or degree, by what becomes accepted
859 as a matter of common or public assumption about me.

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

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

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

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

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

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

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

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

943 Metaphysically fractured and epistemologically elusive though it
944 is, however, my imputed self is bound to matter to me. And with
945 good reason. Like others, I am likely to care about how people
946 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;

990 no one is going to achieve a high standing in the opinion of their peers
991 if, beyond the context where defamation or prejudice threatens, they
992 are seen to be concerned to achieve such standing. ‘The general axiom
993 in this domain’, as Jon Elster (1983, 66) says, ‘is that nothing is so un-
994 impressive as behaviour designed to impress’.¹⁷

995 A final reason against even wanting to try to shape the imputed self
996 is that it is an inherently insatiable concern: it can drive me to seek,
997 not just to do well by local standards, but to outscore all others,
998 and by as much as possible. As Rousseau (2020, 232) argued, follow-
999 ing Hobbes, ‘the first feeling excited by this comparison is the desire
1000 to be first’. It is this love of being first, this love of pre-eminence, that
1001 Rousseau castigated as *l’amour propre*, as distinct from the innocent
1002 *l’amour de soi*. Seeing such self-love as a quest for supremacy, Kant
1003 (2006, 167) also condemned it, finding it present in ‘the manias for
1004 honor, dominance, and possession’: in these manias, he said, ‘the
1005 human being becomes the dupe’.

1006 1007 1008 **Conclusion**

1009
1010 We began by noting the conflict and inconsistency in the various
1011 mantras associated with the self: to be yourself and to let go of the
1012 self; to know yourself and to forget about the self; to be true to your-
1013 self and to avoid self-concern. With the distinctions generated in the
1014 course of discussion, it is possible to find a way through this maze of
1015 confusing advice.

1016 The self you should be, the self you should know and the self you
1017 should be true to is surely the personated self, as we have character-
1018 ized it. This self becomes prominent on the view of agency and
1019 personhood sketched in the first part. If I am to be an agent that per-
1020 sonates, exercising a capacity to make and enact commitments to
1021 others, then the self I project should be a self I realize, a self I try
1022 to know, and a self I take as an ideal: a self, in Polonius’s words, to
1023 which I am true.

1024 That self contrasts naturally with the imputed, un-personated self
1025 that puts me in zero-sum, potentially profitless, competition with
1026

1027 ¹⁷ See (Brennan and Pettit, 2004) for an argument that this observation
1028 does not rule out the possibility of an economy of esteem. The core argu-
1029 ment is that people may be rewarded by esteem and penalized by disesteem,
1030 and may be reinforced or inhibited as a result, without ever seeking esteem or
1031 shunning disesteem in a strategic manner. They may be deeply subject to the
1032 influence of esteem, without being guided by the desire for esteem.

1033 others. This is a self that may distract from a focus on the personated
1034 self, that lies beyond what I can effectively control or reliably know,
1035 and that I cannot openly pursue without undermining that very en-
1036 terprise. The concern for this self may also prove insatiable, in the
1037 manner of a mania, as Kant puts it. Plausibly, it is this self that
1038 I should let go, this self that I should forget, and this self that I
1039 should have little or no concern with – no concern, at any rate,
1040 beyond that required to counter defamation or to join with others
1041 in combatting prejudice.

1042 What, finally, of the referenced self? This is the self in perhaps its
1043 most intriguing guise. It exists, to be sure, but in the manner of a net-
1044 worked sequence of links in a chain, not an underlying substance. It is
1045 there at every juncture to command the focus of the agent but eludes
1046 any attempt to grasp its nature introspectively. And while it
1047 commands the care and concern of the agent, even as it is imagined
1048 into the far future, it commands this as an exigency of agency itself,
1049 not as an attitude that might be voluntarily assumed. Where the
1050 personated self identifies a commanding ideal, and the imputed self
1051 represents the voice of a siren, the referenced self lies deeper than
1052 normative concern, in the very constitution of our nature.

References

- 1056 G. E. M. Anscombe, *Intention* (Oxford: Blackwell, 1957).
1057 D. Bar-on, *Speaking My Mind: Expression and Self-knowledge*, (Oxford:
1058 Oxford University Press, 2004).
1059 J. L. Borges, *Labyrinths: Selected Stories and Other Writings*, (New York:
1060 New Directions, 1962).
1061 A. Boyle, 'Mirror Self-Recognition and Self-Identification', *Philosophy and*
1062 *Phenomenological Research* 97 (2018), 284–303.
1063 G. Brennan and P. Pettit, *The Economy of Esteem: An Essay on Civil and*
1064 *Political Society*, (Oxford: Oxford University Press, 2004).
1065 J. Broome, *Rationality through Reasoning*, (Oxford: Wiley Blackwell, 2013).
1066 E. Camp, 'Putting Thoughts to Work: Concepts, Systematicity, and
1067 Stimulus-Independence', *Philosophy and Phenomenological Research* 78
1068 (2009) 275–311.
1069 D. Dennett, *Brainstorms*, (Brighton: Harvester Press, 1979).
1070 D. Dennett, *The Intentional Stance*, (Cambridge, Mass.: MIT Press, 1987).
1071 N. J. H. Dent, *Rousseau*, (Oxford: Blackwell, 1988).
1072 J. Elster, *Sour Grapes*, (Cambridge: Cambridge University Press, 1983).
1073 G. Evans, *The Varieties of Reference*, (Oxford: Oxford University Press,
1074 1982).
1075 H. Frankfurt, 'Freedom of the Will and the Concept of a Person', *Journal of*
Philosophy 68 (1971), 5–20.

Philip Pettit

- 1076 M. Fricker, *Epistemic Injustice: Power and the Ethics of Knowing*, (Oxford:
1077 Oxford University Press, 2007).
- 1078 T. Hobbes, *Leviathan*, ed E. Curley, (Indianapolis: Hackett, 1994).
- 1079 D. Hume, *A Treatise of Human Nature*, (Oxford: Oxford University Press,
1080 1978).
- 1081 S. Hurley, *Consciousness in Action*, (Cambridge, Mass.: Harvard University
1082 Press, 1998).
- 1083 F. Jackson, and P. Pettit, 'A Problem for Expressivism', *Analysis* 58 (1998),
1084 239–51; reprinted in F. Jackson, P. Pettit and M. Smith, *Mind, Morality
1085 and Explanation*, (Oxford: Oxford University Press, 2004).
- 1086 M. Johnston, *Surviving Death*, (Princeton: Princeton University Press,
1087 2010).
- 1088 I. Kant, *Anthropology from a Pragmatic Point of View*, (Cambridge:
1089 Cambridge University Press, 2006).
- 1090 C. Korsgaard, *Self-Constitution: Agency, Identity, and Integrity*, (Oxford:
1091 Oxford University Press, 2009).
- 1092 R. Langton, *Sexual Solipsism: Philosophical Essays in Pornography and
1093 Objectification*, (Oxford: Oxford University Press, 2009).
- 1094 D. Lewis, *Philosophical Papers Vol 1*, (Oxford: Oxford University Press,
1095 1983).
- 1096 C. List and P. Pettit, *Group Agency: The Possibility, Design and Status of
1097 Corporate Agents*, (Oxford: Oxford University Press, 2011).
- 1098 J. Locke, *An Essay Concerning Human Understanding*, (Oxford: Oxford
1099 University Press, 1975).
- 1100 V. McGeer, 'Is "Self-knowledge" an Empirical Problem? Renegotiating the
1101 Space of Philosophical Explanation', *Journal of Philosophy* 93 (1996),
1102 483–515.
- 1103 V. McGeer, 'The Moral Development of First-Person Authority', *European
1104 Journal of Philosophy* 16 (2008), 81–108.
- 1105 R. Moran, 'Self-Knowledge: Discovery, Resolution, and Undoing',
1106 *European Journal of Philosophy* 5 (1997), 141–61.
- 1107 R. Moran, *Authority and Estrangement: An Essay on Self-knowledge*,
1108 (Princeton: Princeton University Press, 2001).
- 1109 D. Parfit, *Reasons and Persons*, (Oxford: Oxford University Press, 1984).
- 1110 L. A. Paul, *Transformative Experience*, (Oxford: Oxford University Press,
1111 2014).
- 1112 J. Perry, 'The Essential Indexical' *Nous* 13 (1979), 3–21.
- 1113 J. Perry, Ed. *Personal Identity*, (Berkeley: University of California Press,
1114 2008).
- 1115 P. Pettit, *The Common Mind: An Essay on Psychology, Society and Politics*,
1116 (New York: Oxford University Press, 1993).
- 1117 P. Pettit, *Made with Words: Hobbes on Language, Mind and Politics*,
1118 (Princeton: Princeton University Press, 2008).
- P. Pettit, *The Birth of Ethics: Reconstructing the Role and Nature of Morality*,
(Oxford: Oxford University Press, 2018a).

My Three Selves

1119 P. Pettit, 'Consciousness Incorporated', *Journal of Social Philosophy*, 49
1120 (2018b) 12–37.

1121 J. Rawls, *Justice as Fairness: A Restatement*, (Cambridge, Mass: Harvard
1122 University Press, 2001).

1123 J. J Rousseau, *Emile* (New York: Public Domain Books, Kindle ed, 2020).

1124 J. P. Sartre, *Portrait of the Anti-Semite*, (London: Secker and Warburg,
1125 1948).

1126 J. P. Sartre, *The Transcendence of the Ego: An Existentialist Theory of
1127 Consciousness*, (New York: Farrar, Straus and Giroux, 1957).

1128 J. P. Sartre, *Being and Nothingness*, (London: Methuen, 1958).

1129 R. C. Stalnaker, *Inquiry*, (Cambridge, Mass.: MIT Press, 1984).

1130 G. Strawson, *Against Narrativity*, (Oxford: Blackwell, 2005).

1131 B. Williams, *Descartes* (Harmondsworth: Penguin, 1978).

1132 PHILIP PETTIT is Professor of Human Values in Princeton and Professor of
1133 Philosophy in the Australian National University. His recent books include *Just
1134 Freedom* (Norton 2014), *The Robust Demands of the Good* (OUP, 2015) and
1135 *The Birth of Ethics* (OUP 2018).

1136

1137

1138

1139

1140

1141

1142

1143

1144

1145

1146

1147

1148

1149

1150

1151

1152

1153

1154

1155

1156

1157

1158

1159

1160

1161