PROOF COVER SHEET

Author(s):	Kristin Andrews
Article Title:	Pluralistic folk psychology and varieties of self-knowledge: an exploration
Article No:	RPEX1032116
Enclosures:	1) Query sheet
	2) Article proofs

Dear Author,

1. Please check these proofs carefully. It is the responsibility of the corresponding author to check these and approve or amend them. A second proof is not normally provided. Taylor & Francis cannot be held responsible for uncorrected errors, even if introduced during the production process. Once your corrections have been added to the article, it will be considered ready for publication.

Please limit changes at this stage to the correction of errors. You should not make trivial changes, improve prose style, add new material, or delete existing material at this stage. You may be charged if your corrections are excessive (we would not expect corrections to exceed 30 changes).

For detailed guidance on how to check your proofs, please paste this address into a new browser window: http://journalauthors.tandf.co.uk/production/checkingproofs.asp

Your PDF proof file has been enabled so that you can comment on the proof directly using Adobe Acrobat. If you wish to do this, please save the file to your hard disk first. For further information on marking corrections using Acrobat, please paste this address into a new browser window: http://journalauthors.tandf.co.uk/production/acrobat.asp

2. Please review the table of contributors below and confirm that the first and last names are structured correctly and that the authors are listed in the correct order of contribution. This check is to ensure that your name will appear correctly online and when the article is indexed.

Sequence	Prefix	Given name(s)	Surname	Suffix
1		Kristin	Andrews	

Queries are marked in the margins of the proofs, and you can also click the hyperlinks below. Content changes made during copy-editing are shown as tracked changes. Inserted text is in red font and revisions have a red indicator ⊥. Changes can also be viewed using the list comments function. To correct the proofs, you should insert or delete text following the instructions below, but **do not add comments to the existing tracked changes**.

AUTHOR QUERIES

General points:

1. **Permissions**: You have warranted that you have secured the necessary written permission from the appropriate copyright owner for the reproduction of any text, illustration, or other material in your article. Please see http://journalauthors.tandf.co.uk/permissions/usingThirdPartyMaterial.asp.
2. **Third-party content**: If there is third-party content in your article, please check that the rightsholder details for re-use are shown correctly.
3. **Affiliation**: The corresponding author is responsible for ensuring that address and email details are correct for all the co-authors. Affiliations given in the article should be the affiliation at the time the research was conducted. Please see http://journalauthors.tandf.co.uk/preparation/writing.asp.
4. **Funding**: Was your research for this article funded by a funding agency? If so, please insert 'This work was supported by <insert the name of the funding agency in full>', followed by the grant number in square brackets '[grant number xxxx]'.
5. **Supplemental data and underlying research materials**: Do you wish to include the location of the underlying research materials (e.g. data, samples or models) for your article? If so, please insert this sentence before the reference section: 'The underlying research materials for this article can be accessed at <full link>/ description of location [author to complete]'. If your article includes supplemental data, the link will also be provided in this paragraph. See <http://journalauthors.tandf.co.uk/preparation/multimedia.asp> for further explanation of supplemental data and underlying research materials.
6. The **CrossRef database** (www.**crossref**.org/) has been used to validate the references. Changes resulting from mismatches are tracked in red font.

QUERY NO.	QUERY DETAILS
AQ1	Please provide a short biography of the author 'Kristin Andrews'.
AQ2	Please provide complete postal address for the author 'Kristin Andrews'.
AQ3	Please provide a short abstract of no more than journal limit words and journal range keywords.
AQ4	Figure 1 was not cited in the text so a citation has been inserted. Please provide a correction if this is inaccurate.
AQ5	The year for 'Darwin (1872)' has been changed to 'Darwin (1886)' to match the entry in the references list. Please provide revisions if this is incorrect.
AQ6	The CrossRef database (www.crossref.org/) has been used to validate the references. Mismatches between the original manuscript and CrossRef are tracked in red font. Please provide a revision if the change is incorrect. Do not comment on correct changes.

QUERY NO.	QUERY DETAILS
AQ7	Please provide the missing page range for reference 'Bar-On (2011)'.
AQ8	Please provide the missing page range for reference 'Bar-On and Long (2003)'.
AQ9	Please provide complete page range for reference 'Hill (1993)'.
AQ10	Please provide the missing page range for reference 'Neta (2011)'.
AQ11	Please provide complete page range for reference 'Nisbett and Wilson (1977)'.
AQ12	Please provide the missing page range for reference 'Olson et al. (1996)'.
AQ13	Please provide complete page range for reference 'Vazire (2010)'.
AQ14	The reference 'Martin (1998)' is cited in the text but is not listed in the references list. Please either delete in-text citation or provide full reference details following journal style.

How to make corrections to your proofs using Adobe Acrobat/Reader

Taylor & Francis offers you a choice of options to help you make corrections to your proofs. Your PDF proof file has been enabled so that you can edit the proof directly using Adobe Acrobat/Reader. This is the simplest and best way for you to ensure that your corrections will be incorporated. If you wish to do this, please follow these instructions:

1. Save the file to your hard disk.

2. Check which version of Adobe Acrobat/Reader you have on your computer. You can do this by clicking on the "Help" tab, and then "About".

If Adobe Reader is not installed, you can get the latest version free from http://get.adobe.com/reader/.

3. If you have Adobe Acrobat/Reader 10 or a later version, click on the "Comment" link at the right-hand side to view the Comments pane.

4. You can then select any text and mark it up for deletion or replacement, or insert new text as needed. Please note that these will clearly be displayed in the Comments pane and secondary annotation is not needed to draw attention to your corrections. If you need to include new sections of text, it is also possible to add a comment to the proofs. To do this, use the Sticky Note tool in the task bar. Please also see our FAQs here: http://journalauthors.tandf.co.uk/production/index.asp.

5. Make sure that you save the file when you close the document before uploading it to CATS using the "Upload File" button on the online correction form. If you have more than one file, please zip them together and then upload the zip file.

If you prefer, you can make your corrections using the CATS online correction form.

Troubleshooting

Acrobat help: http://helpx.adobe.com/acrobat.html
Reader help: http://helpx.adobe.com/reader.html

Please note that full user guides for earlier versions of these programs are available from the Adobe Help pages by clicking on the link "Previous versions" under the "Help and tutorials" heading from the relevant link above. Commenting functionality is available from Adobe Reader 8.0 onwards and from Adobe Acrobat 7.0 onwards.

Firefox users: Firefox's inbuilt PDF Viewer is set to the default; please see the following for instructions on how to use this and download the PDF to your hard drive: http://support.mozilla.org/en-US/kb/view-pdf-files-firefox-without-downloading-them#w_using-a-pdf-reader-plugin

Philosophical Explorations, 2015
Vol. 18, No. 2, 280–294, http://dx.doi.org/10.1080/13869795.2015.1032116

This paper is part of the special issue 'Self-knowledge in perspective' guest edited by Fleur Jongepier and Derek Strijbos

Pluralistic folk psychology and varieties of self-knowledge: an exploration

Kristin Andrews*

York University

(*Received 10 March 2015; final version received 10 March 2015*)

Self-knowledge is one of the oldest topics of Western philosophy, with the ancient Greeks advising us to *know thyself*. But what does this advice amount to? What sort of knowledge are we being advised to gain, exactly?

In discussions among contemporary analytic philosophers, we find talk of two rough categories of knowledge types: sensations and thoughts. So we might interpret the *know thyself* advice as directing us to attend to our thoughts and sensations. However, advice is usually given in order to direct us to do things that we might not *already* be doing, so we should interpret the ancient Greek advice as directing us to do something *new*. Furthermore, advice is usually given to direct us to do things that we *can* do.

When we turn to look at our facility with thoughts and sensations it is pretty clear that we have a close relationship with at least some of them. Indeed, the untamed mind can distract us with the occurrent thoughts that flitter by: ~~*What Should I Make for Dinner? Cute Pants. The Bread is Almost Done. I'm Tired. The Autumn Leaves Smell Good. Why Am I Grinding My Teeth?*~~ And humans seem to be pretty familiar with lots of sensations too; we can sense our hunger, our bladder, our state of alertness or tiredness, pain and pleasure. Since the Greeks would not have advised us to breathe, or to eat, or to have sex, as we are already pretty good at those things, perhaps they had something else in mind in addition to attending to our surface-level occurrent thoughts and sensations.

One way of exploring the scope of self-knowledge is to use the same method we use when we try to understand other people. We have a commonsense understanding of other minds, called *folk psychology*, which is a lens though which we understand other people. If we interpret the advice as telling us to turn our other-considering lens on ourselves, then we have a self-knowledge practice that we can engage in, which may give us additional information about the self, and that is, at least in some cases, achievable.

Standard theories of folk psychology suggest that we know others by knowing their beliefs and desires. But this description of our understanding of others is quite austere, and does not reflect the multitude of ways we engage with others – not just with their beliefs and desires, but also with their emotions, goals, moods and character traits. We interact with people based on the social roles they play, and the histories they have experienced. My account of Pluralistic Folk Psychology aims to provide a fuller descriptive account of our understanding of others (Andrews 2012). The strategy in this paper is that adopting the

*Email: andrewsk@yorku.ca

methods of Pluralistic Folk Psychology for our own minds will lead us toward understanding the types of self-knowledge we can have. Once we know what self-knowledge consists in, we will be in a better position to begin analyzing whether any claims to knowledge of the self are accurate and to examine the extent to which accuracy is achievable. The end result of this investigation will offer us insights into what is of value in self-knowledge, and into the nature of the self.

Why self-knowledge?

The first step in the Pluralistic Folk Psychology approach to examining how we understand other minds is to examine the functions of this knowledge. Philosophers have stressed that folk psychology allows us to predict others' behavior, and to explain it. The functions of self-knowledge may be different from the functions of other knowledge. Rather than just *predicting* our future behavior, when wondering about what we are going to do next we are often *deciding* what to do (or, in some cases, deciding to let spontaneity be the guide because the decision does not matter very much). We also explain our own behavior, but more often we explain our behavior to others, and the explanation also serves as justification for behavior that others did not expect. In other cases we may need to explain ourselves to ourselves. I might recognize my desire to earn the next belt in Aikido, and believe that I can earn the next belt by going to the dojo three times a week. That should lead me to decide to go to the dojo three times a week. But if I do not feel like working hard right now, that emotion will interact with the belief and the desire, and I will either have to overcome my lethargy or I will not earn the next belt. And if I decide not to go this week, I can keep my self-model consistent by explaining that decision in terms of my other mental states.

Gaining self-knowledge will have effects on ourselves as well as on those around us. Knowing your own emotions, commitments, needs, and preferences and acting consistently with them makes you not only more transparent to yourself, but also makes you more transparent to others. Acting against your nature, ignoring your needs, or pushing aside your emotions also have consequences that affect self and others. It may cause illness, and may lead to an inability to fulfill responsibilities. It may cause stress and tension, and general irritability toward self and others. When you do not know what you want or need, you are going to have a hard time predicting what will happen when you do not get what you want or need. This makes you unpredictable to others as well. With increasing amounts of self-knowledge, however, we can make commitments to the self, and to others, that we strive to live up to. This makes our behavior more predictable, and makes it easier for us to coordinate with others.

The process of gaining self-knowledge changes ourselves, and by changing ourselves we also change our interactions with the people around us. If I come to realize that I need a clean environment to feel peaceful and productive, I may ask the people around me to help achieve that goal. If I come to realize that I am stressed at work, it may cause me to take on fewer tasks, and this will impact others who have to pick up the slack. Self-knowledge has other effects, because of the looping effects that go along with seeing yourself in a new way. Realizing that you need something and proclaiming that you need it create a standard that you hope others will live up to, but one that you yourself try to live up to as well. It regulates behavior of self and other.

In small groups of similar individuals, the path toward self-knowledge is also a path toward group knowledge. When everyone does the same thing in the same circumstance, predictions are easy to make, and self-knowledge is not difficult to achieve, because

knowing about others is the same as knowing about the self. Imagine a self-knowledge quest among schooling fish, each of whom knows what to expect of others given their position in the environment. Nothing would change in terms of the group behavior.

However, when groups are large and heterogeneous, increased self-knowledge can help others to also come to know that self, and thereby gain increased ability to predict and coordinate increasingly complex sorts of behavior. When the ancients travelled to other places, and accepted visitors from other cultures, great differences in human behavior would have been evident. Without the ability to think about his own nature, the immigrant's ability to interact and coordinate with members of a very different culture would be difficult. The differences in behavior lead to questions – requests for explanations for dress, for manners, for ceremonies, for words and concepts. And such explanations are more forthcoming with some amount of self-knowledge.

If we think of self-knowledge as a way of facilitating increasingly complex forms of cooperation in large heterogeneous groups, then we can take the functions of self-knowledge to include increased social coordination. Using folk psychology applied to the self, we can examine the types of self-understanding, as well as ask whether these types of understanding count as self-knowledge, or an accurate understanding of the self. Before we engage in this analysis, we need to detour so I can sketch the theory of Pluralistic Folk Psychology.

Pluralistic folk psychology

Before we engage in the analysis of our self-conception using the same framework in which we understand others, we need to understand what that framework is. Pluralistic Folk Psychology is a theory of folk psychology according to which our understanding of others involves building models of other individuals and types of people. These models are not limited to information about people's beliefs and desires, but include information about other mental states and events such as moods and emotions, and social information about personality traits, dispositions, and historical facts (For a detailed discussion of the points in this section, see Andrews 2012). It is this framework that I think will be helpful in answering questions about the nature of self-knowledge.

Any theory of folk psychology aims to answer the question *how do we understand other people*? A standard answer is that we understand what others believe and desire, and by plugging those beliefs and desires into a theory or by running a simulation with them we can generate predictions about how others will act next. In the first case, the understanding of others is like the kind of understanding scientists gain when they can make predictions about how things will go from here. In the second case, we use our own cognitive resources to determine how we would act in this situation. Either way, understanding permits predictive facility.

But we can also predict things that we do not feel that we understand after thinking about the causes. Turning to a concrete case, I can predict that the toilet will remove the waste after I flush, but I do not understand exactly how the toilet works, why the force of water is sufficient to remove the waste, why the water stops running when the tank is full, etc. And we can predict social behavior without understanding it. Sometimes I have a student who shows up for class, but never pays attention to anything that happens. I know that she will come to class – I predict her behavior – but it is a real puzzle why she is there. I do not understand her at all.

These cases suggest that our sense of understanding that facilitates prediction does not need to go alongside an understanding of the causes of the behavior. In addition,

in some cases knowing people's commitments that lead to actions can cause problems predicting what they will do. Maybe most people would not run into a burning building to save a cat, but I know that you are brave, not fearful of fire, and a cat-lover, so I worry that you *might* risk your life to rescue the cat stranger. Whereas I confidently predict of the random person that he will not enter the building, it is *you* that I am uncertain of.

Understanding, in terms of prediction and explanation, but also in terms of other practices such as coordination, feeling-together, and regulation, requires more than considering others' beliefs and desires. I have proposed that when we get to know people well enough to understand them, we have built models of them. These models are not skeletal lists of mental states, but are richly drawn out like characters in a good novel. We understand people when we know where they are coming from, what they typically do, what they like to do, and we know to what extent their verbal commitments and self-descriptors match their behavior. This understanding can not only help us to predict what an individual will do next, but it also helps us to explain and justify the person's actions, and to coordinate joint social activities.

The core of Pluralistic Folk Psychology is its pluralism – a commitment to the existence of a number of different cognitive mechanisms that support the ability to understand other people. Different mechanisms are needed because attributing propositional attitudes alone is not sufficient. For one, attributing attitudes in the face of behavior is difficult because the propositional attitude set is underdetermined by the observable behavior and theory. This leads to a computational intractability when trying to decide which attitudes a person holds (Apperly 2010; Zawidzki 2013). Second, attributing full-fledged propositional attitudes requires a number of cognitive abilities, including the ability to think about beliefs and to understand the evidential and logical relationships between them. However, though they are folk psychologists, children do not seem to realize that people can have false or true beliefs until about 4 years old (Wellman, Cross, and Watson 2001), they do not seem to realize that attitudes are opaque and must be treated *de dicto* until middle childhood (Apperly and Robinson 1998, 2003), and they have difficulties with evidential relations as young teens (Moshman 2004). Children have problems understanding the hidden intention in irony and double bluff scenarios (Capelli, Nakagawa, and Madden 1990). Conceptual changes in the understanding of belief continue into later childhood and adolescence (Chandler, Boyes, and Ball 1990; Wellman 1990).

Finally, there is reason to think that attributing attitudes such as beliefs is something that we do only deliberately, with our slow and conscious reasoning system, rather than automatically, unconsciously, and subpersonally. Ian Apperly and colleges found that adults do not automatically engage in false belief reasoning (Apperly et al. 2006) and that true belief reasoning is not automatic either (Back and Apperly 2010).

There is also reason to be suspicious of the accuracy of propositional attitude attributions in general. Given the underdermination problem, a proposal that some behavior was caused by belief *b* and desire *d* is going to be a hypothesis. And humans are subject to a confirmation bias when it comes to testing hypotheses. We use positive test strategies, or even worse, we easily become convinced about something just in virtue of already having a hypothesis about it. Then, if someone acts against our hypothesis, our selective memories will sometimes lead us to either forget that hypothesis or to interpret the behavior to confirm our hypothesis (Olson, Roese, and Zanna 1996). In addition, there is empirical evidence that attributions of belief are not accurate, given that adult attributions are prone to egocentric errors (Keysar and Henly 2002). Together, these concerns challenge the *prima*

facie evidence that belief attribution is a common and successful method of predicting behavior.

Instead of associating folk psychology with the ability to attribute propositional attitudes, Pluralistic Folk Psychology examines the practices of folk psychology – such as predicting, explaining, and justifying behavior – and then examines how we engage in those practices. This results in a list of methods we use to achieve our goals. See Table 1 for a rudimentary list of predictive and explanatory types.

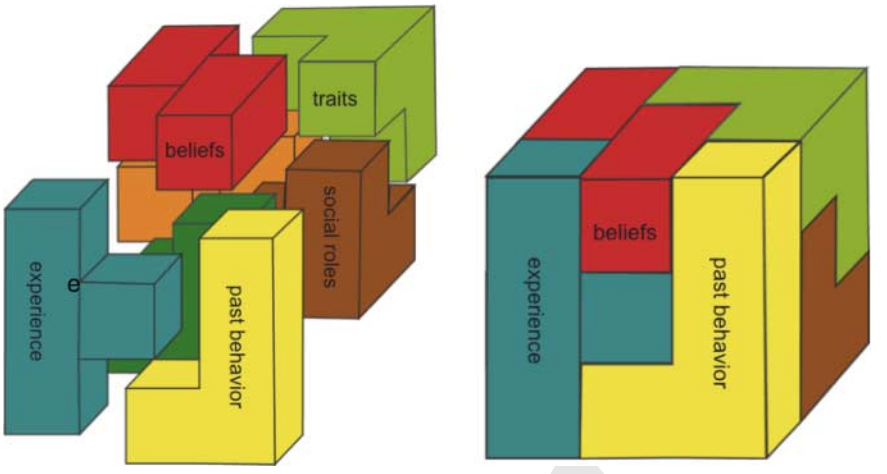
And though there is quite a bit of overlap between predictive and explanatory types of methods or strategies, that should not be taken to mean that there is symmetry in our predictive and explanatory practices. As we saw, we can predict things we cannot explain. But we can also explain things we cannot predict, as historians and pundits are wont to do.

Though it is not entirely clear how these methods interact with one another in any one instance of folk psychological activity, I propose that we in fact solve this interaction problem by building models of those individuals we know well. For all of us, our folk psychology begins as infants interacting with our mothers or other primary caregivers. As we meet more people, we build additional models. And it is from these models that we can form generalizations – about types of people: what women do, what teachers do, what those who believe in God do, and what those who want to get rich do – and about types of situations: what to do in a restaurant, in a theatre, in a dodgy neighborhood, when meeting a new colleague and so on. Rather than beginning with generalizations of behavior, as the standard folk psychology story goes, we begin by understanding individuals as persons. It is from our early experience with individual characters that we later form generalizations about people, though we also form generalizations when we learn things about people from our mother's knee, directly through stories about normal and abnormal behavior, through statements about types of people, and indirectly by observing our caregivers' responses to others. But for most human infants, the root of all folk psychology is the early understanding of the person who takes care of them.

As a child's social domain expands, she builds additional psychological profiles and uses them to understand the people she comes to know well, and these new profiles provide additional information to use in acquiring a more general understanding. When she meets new people, she understands them first via a combination of automatic processes that let her make initial judgments about a person's traits and status, and general model matching to the models she has constructed from past experience. As children gain more

Table 1. Some ways we understand others.

Predictive type	Explanatory type
By reference to self	By reference to self
Stereotypes or social roles	Stereotypes or social roles
Situation or environment	Situation or environment
Inductive generalizations over past behavior	Inductive generalizations over past behavior
Non-propositional mental states such as emotions and sensations	Non-propositional mental states such as emotions and sensations
Teleology or goal	Teleology or goal
Personality trait	Personality trait
Mindreading	Mindreading/reason explanation
Perceptual state	Perceptual state
	Enabling factors
	Causal history



Colour online, B/W in print

Figure 1. Pieces of understanding come together to model an individual.

experience with people, they gain a greater array of particular character knowledge as well as generalizations about normal behavior. This leads adults to have both individual knowledge of persons and general knowledge of character types and normal behavior. We build models from these pieces of our understanding (Figure 1).

With this brief description of how we understand others on the Pluralist Folk Psychology account, we can now apply the account to see how we can seek to understand ourselves. What is crucial to note as we proceed is that as we form models of others, we also form models of ourselves. These models are aspirational as well as descriptive, and they constrain our behavior. It is the comparison with the model that leads to the "that's not me" feeling when we act out of character.

Elements of the self-model

When I really understand myself, what is it that I understand? I suggested already that in the contemporary discussion of self-knowledge the contents under consideration are thoughts and sensations. But to be more precise, we might distinguish five different kinds of content, as presented in Table 2.

Notice, first, that all of these are instances of representations with content. We do understand that we think, feel, desire, and perceive certain things. But it seems that we also understand ourselves in different ways, too. There are at least two other aspects to self-knowledge that do not appear to be part of the traditional analytic approach: knowing how and knowing personality traits. Let us look at each in turn.

Not only do I know certain propositions to be true, I also know how to do lots of things. I know how to make pesto, how to ride a bike, and how to do a headstand. I also know that I can do these things, and hence I am able to represent them propositionally. But there are many other things we know how to do that we might not have metacognitive representational knowledge about. These things that we know how to do come into sharp relief when we leave our normal environment and interact in a place where cultural norms are different. I know how far to stand from people when I talk to them, how to greet people in shops, how to show gratitude for a good meal – at least I know how to do these things in North America. But in different cultures, the expectations are different, and

Table 2. Content of self-understanding in traditional analytic approach to self-knowledge.

Beliefs	Armstrong (1981), Evans (1982), Gallois (1996), ~~Martin (1998)~~, Moran (1997), Neta (2011), Shoemaker (1994), Williamson (2000), Wright (1989)
I believe that the store is now open	
Phenomenal states/sensations/ feelings	Alston (1989), Bar-On (2011), Brewer (1995), Chalmers (2003), Descartes (1985), Gertler (2001), Hill (1991, 1993), Howell (2006), Jackson (1973), Neta (2011), Shoemaker (1994, 1996), Williamson (2000), Wright (1989)
I feel hungry; hot; joyful	
Desires	
I want to go skiing; I desire that it snow for the weekend	Gallois (1996), Shoemaker (1994), Bar-On and Long (2003)
Information	
I know where the food is	Armstrong (1981), Boghossian (1997), Dretske (1994)
Perceptions	
The apple appears red	Dretske (1994), Hume (1978)

AQ14

unless I am given explicit instruction, or spend a lot of time in the strange culture, I run into situation after situation where things do not go as smoothly as I hope. It takes a lot of work to feed myself, or to walk through a busy train station, because I have to explicitly consider interactions that are habitual in my home environment.

One might worry that this is not really a sort of self-understanding at all, but instead is a skill that we have or lack. Instead of having know-how, perhaps we have metacognitive beliefs about our skills that we represent propositionally, like, *I don't understand the norms of a Tokyo subway station*. But understanding that one does not know how to go on from here is an understanding that may not be metacognitive. Consider the feeling that you can jump over a gap while running toward the gap. There is a need for speed, and correct judgments in this case are going to be selected for in order to preserve the organism. I know how to make good decisions about my abilities that need not involve explicit reasoning or mathematical calculations of distance. Horses and dogs make the same sorts of decisions, and hence have this sort of self-understanding. So, we might add to the list of ways we can know ourselves knowing how to act, or how to go on.

Another thing to notice about this list is that it does not include personality traits. Traits are descriptions of an individual in terms of personality or social role. One can be a woman, a mother, a philosopher, a yoga teacher, a reader, a gringo, or a goth. And individuals can have traits such as: kind, soft-spoken, brass, friendly, shy, and rude. Also, people may be irritating or endearing, boring or intriguing, optimistic or pessimistic, anxious or calm. Psychologists who study traits think that they play a significant role in our understanding of self and other. The psychologist Simine Vazire studies our understanding of self and other in terms of traits, and she draws a distinction between internal traits, or aspects of the self that refer to thoughts and feelings, and external traits that are better identified by looking at externally observable patterns of behavior. Examples of internal traits would include being optimistic or calm, whereas external traits include being charming or boisterous (Vazire 2010).

One might think at first that internal traits are in fact part of the traditional approaches to self-knowledge, because the list includes sensations and emotions. If I know I am anxious, I know my sensation. But traits are usually taken to be general properties of persons, not just descriptions of occurrent mental experiences. A calm person can occasionally be anxious –

and because she is generally calm she may handle her anxiety worse than someone who is used to experiencing it.

Internal traits such as generalizations about sensations or behaviors, identity associations, and external traits about behaviors, reputation, and social role are also part of our self-understanding matrix. Adding these to the types typically discussed will fill out the picture of self-understanding and allow us to do more when we analyze what it means to know oneself.

Like our understanding of other people, we understand ourselves in a variety of ways. In fact, when looking through the list of understanding types generated from the Pluralistic Folk Psychology account of predicting and explaining other behaviors, it is easy to see that almost all those methods can be translated into understanding the self (see Table 3).

But there are at least two clear differences between these two types of understanding. First of all, self-reference is not going to be a form of self-understanding, because it *relies* on self-understanding. Self-reference is a way we understand other people by thinking they are like us, sharing preferences, reactions to situations, and so forth. Our self-understanding has a correlate which we can call other-reference, whereby we understand ourselves in comparison with other people. We compare ourselves with others, find both similarities and differences, and then use those similarities and differences to orient ourselves within a social context. For example, when we see another person carefully washing his car, polishing the chrome, and making everything shine, that can bring to light one's own lack of concern for a clean and shiny car. Seeing a family pray before a meal can also highlight one's own lack of religious commitment. Comparisons might be odious, but they are a natural part of how we come to see ourselves as similar and different to others.

Second, stereotype activation works differently for self and other. Stereotypes are sets of properties that are associated with a particular social group. As a kind of knowledge structure, stereotype activation is a case of basic cognitive functioning that is used to make quick predictions of behavior. Because our understanding of the self does not typically involve the need to make predictions, stereotype activation will have a different role to play in self-understanding. Rather than using stereotype activation to predict behavior, there is evidence that stereotype activation directed at the self has automatic effects on our own behavior, as in the case of stereotype threat. When a person is placed in a situation for which there is a

Table 3. Types of understanding self and other.

Understanding type	Understanding others	Understanding self
By reference to self	x	
By reference to other		x
Stereotypes or social roles	x	x
Situation or environment	x	x
Past behavior	x	x
Past experience	x	x
Emotions	x	x
Teleology	x	x
Personality trait	x	x
Propositional attitudes	x	x
Percepts	x	x
Moods		x
Sensations		x

negative stereotype about that person's group, there is a decrease in performance (Steele and Aronson 1995). This is the case even when the individual explicitly disavows the stereotype. There is a corresponding increase in performance for situations in which a positive stereotype is activated. In one illustrative study, Asian-American women who were primed to think about being Asian before taking a math test performed better than controls, whereas Asian-American women who were primed to think about their gender did worse than controls (Ambady et al. 2001).

Because our self-conceptions can change our performance, there is a power in self-understanding. We see that the models that we construct of ourselves will change in different contexts, and our behaviors will likewise adjust given the model and the situation. In the last section of the paper, I will return to this point and suggest that this way of thinking about how we conceive ourselves, as building models that highlight different aspects of an individual, may have metaphysical implications into the nature of the self as well. But before we get to that topic, we should examine whether these sorts of self-understanding amount to knowledge. To what extent do we actually *know* these elements of the self? We will turn to that topic next.

Accuracy of our self-conception

We sometimes do not know why we do what we did. When Uruguay footballer Luiz Suarez bit an Italian defender during a World Cup match in 2014, he merely said, "These things happen on the pitch". It was his third known bite during a football match, so it was not completely unexpected that he bit Chiellini. Suarez is a biter, and while he can predict that he will probably bite again someday, it is still puzzling (even to him) why he bites. He cannot offer an explanation.

This example demonstrates that asymmetry between psychological prediction and explanation occurs even with the self, especially where our habitual actions are concerned. Social psychology has long been rife with cases of humans failing to predict and explain their own behavior. At least as far back as the infamous Milgram authority studies, scientists have found that people will do things that they would not have expected. In addition, confirmation biases cause us to think an action more likely if we consider the reasons for it. For example, in one study, college students were asked to predict if they would act in a friendly or unfriendly way to a new student. The experimental subjects were asked to consider their reasons to act one way or another before making the prediction, but the control group was not given this instruction. The experimental group was both more confident in their response and *less accurate* in their prediction (Wilson and LaFleur 1995). This and similar studies suggest that by considering our reasons – our beliefs and desires – we can actually decrease the accuracy of our predictions. The problem arises because we use a positive test strategy to decide whether a theory is correct, focusing on the information that makes the hypothesis seem more likely. Thus, in developing a theory, we are likely to accept the first plausible explanation, and act on that.

And just as famously, we are also sometimes impaired in our ability to offer accurate explanations for our behavior. In their pantyhose study, the psychologists Nisbett and Wilson (1977) found that humans attribute thoughts to themselves as reasons, even though it is apparent that those thoughts are nothing more than post hoc confabulations. Subjects in a shopping mall were invited to take a market survey, and were presented with four identical pairs of pantyhose that they were asked to choose among. Most subjects chose the rightmost pantyhose. When asked to explain their choice, subjects confabulated responses, claiming that they thought the chosen pair was the softest or had the best color.

No one explained their choice based on the location of the hose. And since the pantyhose were identical, the reasons stated were unlikely to be the cause of behavior. Rather, location was the relevant variable.

When hearing about these results for the first time, one might be surprised about the lack of accuracy in at least some of our explanations. But digging a bit deeper, the idea that we are not authorities on what we ourselves think is easily accommodated within our common-sense folk psychology. If we attributed to others authority with respect to their own explanations, then we would expect first person avowals to be viewed as sacred, and never to be challenged. Since we clearly do not treat people's testimony about their own psychological states to be sacrosanct, it seems that being wrong about oneself is part and parcel of our other-considering folk psychology. Otherwise, the following exchanges would not strike us as acceptable:

(a) A: "I really love him".
B: "Are you sure?"
(b) A: "Owww . . . my knee"
B: "But there's a scrape on your elbow".
A: "I didn't even notice that!"
(c) A: "That's horrible!"
B: "No, you know you like it!"
(d) A: "I'm a really good person who cares about others".
B: "No, you're actually quite selfish".

These exchanges cover a range of self-understanding types: emotions, sensations, judgments, and traits. Timothy Williamson challenges the accuracy of our phenomenal experiences, arguing that a person who feels cold in the morning and warm at noon must go through a phenomenal shift from feeling cold to feeling warm (2000). At some point in this transformation she is barely cold, and truly believes that she feels cold. But at the next moment when she is slightly less cold, she believes falsely that she feels cold. This argument is brought forth as a challenge to the view that we have authority on all our phenomenal states.

But this thought experiment might distract from the reality of our interoceptive abilities. There are individual differences when it comes to the ability to sense the internal workings of the body, such as heartbeat or the state of the stomach. When interoceptive abilities are measured by asking subjects to count their heart beats, and then comparing the self-ascribed heart rate with an objective measure, scientists find that there is a high variability, and that there are correlations between anxiety disorders and high interoceptive ability (Mor and Winquist 2002). This finding is consistent with an embodied account of feelings, according to which those who are better in touch with their feelings are better able to feel their body. Hunched shoulders, nervous gut, and gritted teeth might be one's natural way of being in the world, and yet individuals who do not notice these bodily sensations cannot use them to realize that they are in a state of low-level stress.

Our interoceptive abilities may not be fixed, but rather mutable with effort, as suggested in a recent study looking at the ability of subjects to sense subtle tactile stimuli. Compared with a control group, experienced meditators who practice a form of concentration meditation focused on their body are better at identifying the touches (Fox et al. 2012). Though there are individual differences when it comes to accuracy about certain kinds of sensations, we can improve our skills in this domain.

Another area in which our self-understanding is sometimes compromised is in the identification of personality traits. While we do understand ourselves in terms of traits, thinking of oneself as generous, kind, or outgoing, there has been scientific investigation into the self-attribution of traits by psychologists that suggests that we do not understand ourselves as well as we think we do (Vazire and Carlson 2011). Others may understand us better than we understand ourselves, because we miss important aspects of our own personalities when listing traits. We are usually motivated to see ourselves in a positive way, and this motivation leads to problems when it comes to accuracy (Dunning 2005). But even when we see ourselves more negatively, we still seek to confirm this view, looking for evidence of our failures instead of our successes (Swann 1997). Vazire finds that we are better at knowing our internal traits and that our external traits are often opaque to us. But she thinks that we can improve our self-trait estimation through meta-perception, or taking the perspective of others perceiving us. We are more accurate in detecting the impression we make on others (Carlson, Furr, and Vazire 2010).

And it is not just explaining our behavior in various ways that can be compromised. Self-prediction also does not always work out the way we expect. People's judgments about what will make them happy frequently turn out to be wrong, as psychologist Daniel Gilbert has documented. When people get what they want, be it a raise, a fancy house or car, a new job, or when people achieve some professional goal, they do not become happier than before they got what they wanted; after a brief boost, we all settle back to our baseline (Gilbert 2006). What we think we want is often not something that will bring us any lasting benefit.

We also have difficulty predicting what we will do in unusual circumstances – an example of the fundamental attribution error at work. I might think that as a good utilitarian I would kill the one to save the many, but if I am actually placed in the situation, I may find myself unable to pull the trigger. No less a figure than Charles Darwin experienced this same disjoint between self-prediction and actual behavior. In describing one interaction with a snake at a zoo, Darwin writes,

> I put my face to the thick glass-plate in front of a puff adder in the Zoological Gardens, with the firm determination of not starting back if the snake struck at me; but as soon as the blow was struck, my resolution went for nothing, and I jumped a yard or two backward with astonishing rapidity. My will and reason were powerless against the imagination of a danger which had never been experienced. (1886, 40)

AQ5

Our predictions and our explanations of our own behavior are sometimes impaired. In some cases we can improve our ability to see ourselves. And in other cases our self-understanding is constitutive of our self, as in the effects of stereotype threat, looping effects, and self-fulfilling prophecies. When understanding self, and other, we are letting normative, and often aspirational, considerations come into play. Even our belief and desire ascriptions can be seen as prescriptive. In some cases we ascribe ourselves beliefs or desires in a more aspirational way; we sometimes "fake it till we make it". In such cases we are setting up self-fulfilling prophecies for the self, as in the case of the marathon runner who tells herself that she can finish in four hours in order to better her previous time. In other cases, we purposefully avoid information that will throw us off our game so that we *do not* form a belief that will cause problems. Remember Han Solo's admonition of C-3P0 while navigating the asteroid field to "Never tell me the odds".

The unwillingness to believe things that we do not want to accept has repercussions for the community as well. Consider an individual who helps take care of a symptomatic

relative with Ebola. She does not want to believe that she is going to get sick herself, and so she might tell herself that she is not at risk. With this belief fully entrenched in herself, she will use the belief to drive her behavior, as well as her answers to questions about Ebola exposure. For self-model preservation reasons, the individual risks getting other people sick by believing that she cannot be sick herself.

In this section, my goal was to show that we do not, and should not, think that self-understanding is a purely veridical endeavor. A complete account of self-understanding from the perspective of PFP requires significant work detailing all the types of understanding, and the degree of accuracy for each of these. For our purposes now, it is enough to raise the concerns about accuracy or lack thereof.

Self-understanding, knowledge, and creation

When we expand the contents of self-knowledge from thoughts and sensations to include the plurality of other ways in which we understand ourselves, we see that aspects of our self-understanding are fluid, constructive, pragmatic, and often inaccurate. We shape our self-understanding based on our goals, as well as the thoughts and behaviors of others. Our ways of thinking about the self will differ in different situations. When we look at the nature of self-understanding outside the limited domain of thoughts and sensations, the target of this understanding appears to be moving, and one that is created through the process of understanding. The transition from thinking about the self merely in terms of an objective thing that we need to find, rather than as a fluid thing we are constantly creating, also raises the question about the accuracy of self-knowledge. This suggests both that there are parts of the self to get right or wrong, other parts of the self that we create though our ways of thinking about the self, and that these two aspects might be intertwined and the interactions between them quite mercurial.

Reputation control is a creative aspect of self, but also an aspirational one, for we need to live up to the ways we want others to see us. Stereotype threat is also a creative aspect of self, but a destructive one when it causes us to fail to live up to our natural abilities. Part of our self-conception consists of the stories we tell ourselves and others about how we got from A to B, consistent with narrative views of the nature of self (Bruner 1983; Schectman 1996). And other aspects are more widely observable, dealing with observable patterns of behavior or physiological properties.

The main difference between our understanding of self and our understanding of other stems from the "direction" in which we draw comparisons. When understanding others, we draw from what we think we know about the self. And when understanding the self, we draw from what we think we know about others. In both cases we build models, but what this approach shows us is that there is a co-constitutive relationship between our self-conceptions and our other-conceptions. As an infant at the breast, we humans begin to form models of both self and other. We learn how to interact with the human we are initially most intimate with, and from those interactions start to understand the self. As the caregiver forms expectations of her infant, and the infant comes to socially engage with the caregiver, exchanging smiles and vocalizations, expectations are reinforced. Our self-model is formed in relation to our model of others, and the one model feeds back to change the other.

If the self is a fluid thing that we are constantly creating through our actions and self-constituting thoughts, it is a creation we do not make alone. Others also create us, and we create them. This co-creative aspect of self and other understanding suggests another way of understanding the Greek advice to know ourselves. To know oneself one needs to know

how one is perceived by others, to take on the other's perspective and see oneself reflected back through her eyes. And this is something that psychologists think we can do. So we might interpret the advice as telling us to work on seeing others' perspective when engaged with another, in conversation, in argument, in judgment, and in love. Understood thusly, *know thyself* is a dictate toward empathy. Given their concern with the social aspect of human flourishing, I think the Greeks would approve.

Acknowledgements

Thanks to Fleur Jongepier and Derek Strijbos for inviting me to think about self-knowledge from the perspective of my theory of Pluralistic Folk Psychology. Also thanks to Jacob Beck, Ian Wright, Brian Huss, Olivia Sultanescu, and Johannes Roessler for their conversations and comments on the paper.

Disclosure statement

No potential conflict of interest was reported by the author.

Notes on contributors

AQ1

References

AQ6

Alston, W. 1989. *Epistemic Justification*. Ithaca, NY: Cornell University Press.

Ambady, N., M. Shih, A. Kim, and T. L. Pittinsky. 2001. "Stereotype Susceptibility in Children: Effects of Identity Activation on Quantitative Performance." *Psychological Science* 12 (5): 385–390.

Andrews, K. 2012. *Do Apes Read Minds: Toward a New Folk Psychology*. Cambridge, MA: MIT Press.

Apperly, I. 2010. *Mindreaders: The Cognitive Basis of Theory of Mind*. New York, NY: Psychology Press.

Apperly, I., K. Riggs, A. Simpson, D. Samson, and C. Chiavarino. 2006. "Is Belief Reasoning Automatic?" *Psychological Science* 17 (10): 841–844.

Apperly, I., and E. Robinson. 1998. "Children's Mental Representation of Referential Relations." *Cognition* 67: 287–309.

Apperly, I., and E. Robinson. 2003. "When can Children Handle Referential Opacity? Evidence for Systematic Variation in 5- and 6-year-old Children's Reasoning About Beliefs and Belief Reports." *Journal of Experimental Child Psychology* 85 (4): 297–311.

Armstrong, D. 1981. *The Nature of Mind and Other Essays*. Ithaca, NY: Cornell University Press.

Back, E., and I. Apperly. 2010. "Two Sources of Evidence on the Non-automaticity of True and False Belief Ascription." *Cognition* 115 (1): 54–70.

Bar-On, D. 2011. "Neo-Expressivism: Avowals' Security and Privileged Self-knowledge." In *Self-Knowledge*, edited by A. Hatzimoysis. Oxford: Oxford Publishing.

AQ7

Bar-On, D., and D. Long. 2003. "Knowing Selves: Expression, Truth, and Knowledge." In *Privileged Access: Philosophical Accounts of Self-knowledge*, edited by B. Gertler. Aldershot: Ashgate Publishing.

AQ8

Boghossian, P. 1997. "Content and Self-Knowledge." *Philosophical Topics* 17: 5–26.

Brewer, B. 1995. "Bodily Awareness and the Self." In *The Body and the Self*, edited by J. L. Bermúdez, A. Marcel, and N. Eilan, 291–309. Cambridge, MA: MIT Press.

Bruner, J. S. 1983. *In Search of Mind: Essays in Autobiography*. New York: Harper & Row.

Capelli, C. A., N. Nakagawa, and C. M. Madden. 1990. "How Children Understand Sarcasm: The Role of Context and Intonation." *Child Development* 61 (6): 1824–1841.

Carlson, E. N., R. M. Furr, and S. Vazire. 2010. "Do We Know the First Impressions We Make? Evidence for Idiographic Meta-accuracy and Calibration of First Impressions." *Social Psychological and Personality Science* 1 (1): 94–98.

Chalmers, D. 2003. "The Content and Epistemology of Phenomenal Belief." In *Consciousness: New Philosophical Essays*, edited by Q. Smith and A. Jokic, 220–271. Oxford: Oxford University Press.

Chandler, M., M. Boyes, and L. Ball. 1990. "Relativism and Stations of Epistemic Doubt." *Journal of Experimental Child Psychology* 50 (3): 370–395.

Darwin, C. 1886. *The Expression of the Emotions in Man and Animals*. New York: Appleton and Company.

Descartes, R. 1985. *Meditations on First Philosophy.* Translated by Cottingham, Stoothoff and Murdoch. *The Philosophical Writings of Descartes Volume II*. Cambridge: Cambridge University Press.

Dretske, F. 1994. "Introspection." *Proceedings of the Aristotelian Society* 94: 263–278.

Dunning, D. 2005. *Self-insight: Roadblocks and Detours on the Path to Knowing Thyself*. New York: Psychology Press.

Evans, G. 1982. *The Varieties of Reference*. Edited by J. McDowell. Oxford: Oxford University Press.

Fox, K. C., P. Zakarauskas, M. Dixon, M. Ellamil, E. Thompson, and K. Christoff. 2012. "Meditation Experience Predicts Introspective Accuracy." *PloS one* 7 (9): e45370.

Gallois, A. 1996. *The Mind Within, the World Without*. Cambridge: Cambridge University Press.

Gertler, B. 2001. "Introspecting Phenomenal States." *Philosophy and Phenomenological Research* 63: 305–328.

Gilbert, D. 2006. *Stumbling on Happiness*. New York: Random House.

Hill, C. 1991. *Sensations: A Defense of Type Materialism*. Cambridge: Cambridge University Press.

Hill, C. 1993. "Qualitative Characteristics, Type Materialism and the Circularity of Analytic Functionalism." *Behavioral and Brain Sciences* 16 (1): 50.

Howell, R. 2006. "Self-Knowledge and Self-Reference." *Philosophy and Phenomenological Research* 72: 44–70.

Hume, D. 1978. *A Treatise of Human Nature*. Edited by L. A. Selby-Bigge; revised by P. H. Nidditch. Oxford: Oxford University Press.

Jackson, F. 1973. "Is There a Good Argument Against the Incorrigibility Thesis?" *Australasian Journal of Philosophy* 51: 51–62.

Keysar, B., and A. S. Henly. 2002. "Speakers' Overestimation of Their Effectiveness." *Psychological Science* 13 (3): 207–212.

Mor, N., and J. Winquist. 2002. "Self-focused Attention and Negative Affect: A Meta-analysis." *Psychological Bulletin* 128 (4): 638–662.

Moran, R. 1997. "Self-Knowledge: Discovery, Resolution, and Undoing." *European Journal of Philosophy* 5: 141–161.

Moshman, D. 2004. "From Inference to Reasoning: The Construction of Rationality." *Thinking & Reasoning* 10 (2): 221–239.

Neta, R. 2011. "The Nature and Reach of Privileged Access." In *Self-Knowledge*, edited by A. Hayzimyosis. Oxford: Oxford University Press.

Nisbett, R. E., and T. D. Wilson. 1977. "Telling More than We Can Know: Verbal Reports on Mental Processes." *Psychological Review* 84 (3): 231.

Olson, J. M., N. J. Roese, and M. Zanna. 1996. "Expectancies." In *Social Psychology: Handbook of Basic Principles*, edited by E. T. Higins and A. W. Kruglanski. New York: Guilford Press.

Schectman, M. 1996. *The Constitution of Selves*. Ithaca: Cornell University Press.

Shoemaker, S. 1994. "Self-Knowledge and 'Inner Sense': Lecture I: The Object Perception Model." *Philosophy and Phenomenological Research* 54: 249–314.

Shoemaker, S. 1996. *The First-Person Perspective and Other Essays*. Cambridge: Cambridge University Press.

Steele, C. M., and J. Aronson. 1995. "Stereotype Threat and the Intellectual Test Performance of African Americans." *Journal of Personality and Social Psychology* 69 (5): 797–811.

Swann, W. B. 1997. "The Trouble with Change: Self-verification and Allegiance to the Self." *Psychological Science* 8 (3): 177–180.

Vazire, S. 2010. "Who Knows What about a Person? The Self-other Knowledge Asymmetry (SOKA) Model." *Journal of Personality and Social Psychology* 98 (2): 281.

Vazire, S., and E. N. Carlson. 2011. "Others Sometimes Know Us Better Than We Know Ourselves." *Current Directions in Psychological Science* 20 (2): 104–108.

Wellman, H. M. 1990. *The Child's Theory of Mind*. Cambridge, MA: MIT Press.

AQ9 AQ10 AQ11 AQ12 AQ13

Wellman, H. M., D. Cross, and J. Watson. 2001. "Meta-analysis of Theory-of-mind Development: The Truth about False Belief." *Child Development* 72: 655–684.
Williamson, T. 2000. *Knowledge and Its Limits*. Oxford: Oxford University Press.
Wilson, T. D., and S. J. LaFleur. 1995. "Knowing What You'll Do: Effects of Analyzing Reasons on Self-prediction." *Journal of Personality and Social Psychology* 68: 21–35.
Wright, C. 1989. "Wittgenstein's Later Philosophy of Mind: Sensation, Privacy, and Intention." *Journal of Philosophy* 86: 622–634.
Zawidzki, T. W. 2013. *Mindshaping: A New Framework for Understanding Human Social Cognition*. Cambridge, MA: MIT Press.