

"Plato on Akrasia and Knowing Your Own Mind"

Section 1: Introduction

The standard picture of the development of Plato's views on akrasia depicts an increase in subtlety and psychological realism from the early to the middle and late dialogues. In the Protagoras, Plato analyzes away apparent instances of akrasia by claiming that the person acts on a desire for what he thinks of as best all things considered or best overall, but has overestimated the value of the wrong course of action because its rewards are close in time. Aristotle complains that this "conflicts with the phainomena" (Nicomachean Ethics 1145b28) and Sidgwick finds it "an extravagant paradox."¹ In the Republic, Plato partitions the soul and thus recognizes the existence of non-rational motivations that do not aim at what is best for the whole person overall. These desires do not originate with a judgment about what is best for the person overall and can persist even in the face of a judgment that another course of action is overall better. Conflict of desire is thus possible. In the Republic, on the usual story, Plato also thinks that it is possible for the non-rational motivation, say a spirited desire or emotion, to win out in the

competition with a desire for what is overall best. The person may act on it and thus akratic action is also possible.

Plato's middle and late position does have noteworthy advantages. It appears to be more realistic in that it provides a psychological account that respects the phenomenology of deliberative conflict. It seems, at least in some cases, to the deliberating agent that she, at the same time, possesses two distinct desires that lead her toward incompatible actions and that these desires have two distinct ends, say, her overall good and a tempting pleasure. Second, the middle and late account is more sophisticated in that it recognizes the complexity of human motivation. We might not agree with some contemporary philosophers of action who hold that a realistic account of human motivation must allow for "disaffected, refractory, silly, satanic, or punk" agents, that is, those, who desire things under negative characterizations, for example, the purely self-destructive.² (Or at least we might want to deny that these are rational agents.) Nevertheless, we might prefer an account that allows for a range of positive characterizations under which things can be desired besides that of the overall best.

In rough outline, these are some of the main reasons that Plato's middle-period position has seemed to display a clear gain in psychological realism and subtlety. And I do think that Plato's middle- and late-period views on akrasia embody important psychological and philosophical advances. But here I want to call attention to and discuss one aspect of Plato's earlier views whose interest and subtlety have not, I think, been sufficiently appreciated. On the Protagoras account of apparent akratic action, although I think that I am being overcome by pleasure, I am, in fact, pursuing my overall good; it is just the case that my judgment is mistaken because I have overvalued short-term goods. What is, I think, especially interesting about this explanation is that on it the person's mind is opaque to herself. She thinks that when she chose the wrong action she was pursuing pleasure, but in fact she was acting on a (mistaken) judgment about and desire for what is overall best. In a deep sense, the person does not know her own mind.

Let me begin with three preliminary observations. First, Plato's theory in the Protagoras may be reminiscent of the claim that he makes about desire and the good in the Gorgias and the Meno. Is the Protagoras theory simply

an outcome of the claims in the Gorgias and Meno? Although there are surely important similarities among the three dialogues, the Protagoras position, at least on first examination, is sufficiently distinct to merit its own investigation. To see this, consider two immediate apparent differences from the Gorgias. First, on one interpretation of the Gorgias (or at least on one interpretation of Gorg. 466A-468E), Plato there claims that all that we want or desire is the actual good or best.³ So if something is not actually best for us overall, we cannot have the attitude of desire towards it. Plato certainly does not endorse this view in the Protagoras. Indeed, the evidence suggests that the Protagoras attributes the same type of motivational attitude both to the person who goes for what is actually good and to the person who mistakenly goes for what is only apparently good. Further, it is reasonable to think that this motivational state is simply a desire.⁴ In any case, for convenience, in the rest of the paper I shall attribute a desire for what he mistakenly believes to be good to the allegedly akratic agent. But this attribution is not essential to the issues I shall discuss and we could reformulate these worries if we were to hold that

this agent's motivational state is not correctly characterized as a desire.

A second point to note is this. Consider the following passage from the Gorgias.

So it is pursuing what is good that we walk whenever we walk; that is, because we think that it is better to walk. And conversely, whenever we stand still, we stand still for the sake of the same thing, what is good . . . and do we not also put a person to death, if we do, or banish him and confiscate his property because we think that doing these things is better for us than not doing them?⁵ (Gorg. 468B1-6)

When a person goes wrong, e.g., by killing someone or confiscating his wealth when it is not best to do so, Plato attributes a relevant favorable attitude to the agent: he acts thinking it is best to do so or because he thinks it best to do so (oiomenoi beltion einai). As in the Protagoras account, we have isolated some attitude of the agent involving a positive characterization of some object or state of affairs that is essential to the agent going wrong. But in the Gorgias, the agent has true beliefs about what

he thinks best, more precisely, he is aware that he is experiencing an episode of thinking that something is best and that the object of his attitude is something he thinks is best. Although he suffers from a false belief about what is best, this agent does know his own mind in a way that the allegedly akratic agent of the Protagoras does not. That agent, as we saw, misidentifies the object of his desire: he thinks that what he has and acts on is a desire for pleasure, but it is in fact a desire for the overall best. So the agent who goes wrong in the Gorgias knows his own mind in a way that the agent of the Protagoras does not.

A third and final preliminary point. It is sometimes said that Plato's views about our awareness (or lack of it) of our psychological states is shocking for the modern philosopher and is thus difficult to understand "because of our acceptance of the notion that we have incorrigible knowledge of our psychological states."⁶ The thesis of incorrigibility is then attributed either to Cartesian philosophy of mind or to Fregean philosophy of language and the cure is found in contemporary theories of reference. This approach has produced some valuable insights, but I would like to suggest a broader context for these issues. Cartesian incorrigibility about our

psychological states has hardly been a dominant force in contemporary philosophy of mind (certainly not since the early 20th century, if then). For example, functionalist accounts of desire will not, in general, be friendly to the incorrigibility claim. Nor does one need to be a functionalist to be skeptical about the incorrigibility of our awareness of our desires. Even if one has a non-functional understanding of a mental representation entering a "desire box" in the mind, this does not entail that its possessor is aware of it or can correctly identify it.

Moreover, the acceptance of the possibility of a systematic failure to recognize correctly the contents of our desires is hardly limited to Anglo-American philosophy of mind and action of the last 100 or so years. There is, for example, a rich tradition in psychoanalytic theory of positing desires which are unconscious or are misinterpreted by the agent. Relatedly, non-analytic philosophers such as Marcuse and Adorno, think that one of the pernicious effects of modern capitalist societies is precisely that they engender in their members such misinterpretation of their own desires.⁷ Further, many psychological theories that offer reductionist theories of motivation will have the consequence that I am sometimes

not aware of the object of my desire. So related issues may arise for views as disparate as psychological hedonism and Nietzsche's will to power. Finally, recent experimental psychological studies have suggested, or so it is claimed at any rate, that "there may be little or no direct introspective awareness of the higher order cognitive processes."⁸

I do not want to suggest that all these views that deny Cartesian incorrigibility come down to the same theory, since obviously they do not (and it may be the case that Descartes himself did not hold a general incorrigibility thesis about psychological states).⁹ Nor do I want to suggest that because such views are widespread they are entirely unproblematic. Indeed, much work is again being done on the issue of self-awareness and self-knowledge and on our authority with respect to our own psychological states. Some have argued that some form of self-knowledge and first-person authority are necessary for the possibility of rational deliberation, others that our very understanding of what it is to be an agent requires that there are attitudes that we can take to our own psychological states that we cannot take to those of others. There is much more to be said about all of this than I can

discuss in this article, but I hope at least to raise some new issues.

So the structure of the rest of the paper is as follows. In the next section, I draw on some recent work on self-knowledge to sketch several worries to which our lack of awareness of our own minds might give rise. In the third section, I shall turn to the details of Plato's solution to the puzzle of apparent akratic action in the Protagoras. There I argue that although it is possible that the standard interpretation of Plato's solution captures the intent of the argument, it is not the only solution to the puzzle that is compatible with Plato's broader commitments in the early dialogues. I shall sketch an alternate solution to the puzzle of apparent akratic action that is, I think at any rate, compatible with Plato's broader commitments and is philosophically preferable to the standard interpretation. I shall also suggest that reflection on this alternate interpretation may well lead one in the direction of Plato's views in the middle period, e.g., in the Republic. In the fourth section, I draw on the worries to which our lack of awareness of our own minds might give rise to help illuminate the situation of the apparent akratic in the Protagoras. Finally, I conclude with a few brief remarks

about what the lines of thought that we have explored may suggest about Plato's conception of rationality.

Section 2: Issues Raised by Recent Work on Self-Knowledge

So now let us turn to the worries to which our lack of awareness of our own minds might give rise. To begin, we are creatures that have beliefs and desires and are moved to act on our beliefs and desires. For this reason, we are able to explain our actions in terms of our beliefs and desires. To explain my reaching for a glass, I appeal to my desire for some water and my belief that the stuff in the glass is water. Note that this quite satisfying explanation of my action invokes my desire for some water and my belief about what is in the glass, that is, it invokes first-order beliefs and desires. It is these first-order beliefs and desires themselves that bring about my action and it is in terms of them that we can explain the action and see its point or rationality. But this simple story of rational action at a time must be made more complex to accommodate the fact that the world changes and so do my beliefs and desires. As a rational creature, I must (a) adapt my beliefs and desires in the light of new information, and (b) decide what sorts of further inquiry to engage in. To do each of

these successfully, it seems that the person must have an awareness of his initial beliefs and desires.

An essential part of rationality, for creatures with the conceptual capacities of human beings, is the appropriate adjustment of beliefs and desires in the light of new information about the world, and, as a necessary means to such adjustment, the conducting of appropriate tests and reasoning. For someone to know what sorts of tests and reasoning are called for it is essential that he know what his current beliefs are--only so can he know which of them are called into question by new information, and what questions about the world his tests and reasoning should be focused on. Similar points apply to desires; one cannot rationally revise one's desire system in the light of experience without having knowledge of what desires one currently has. This seems to require that beliefs and desires be self-intimating.¹⁰

We can take a further step. We do not merely think of ourselves as rational creatures, but as rational agents. Consider, as a point of contrast, Hobbesian deliberation. For Hobbes, deliberation is simply a psychic episode in

which various appetites and emotions battle against one another until one finally wins out.

When in the mind of man, appetites and aversions, hopes and fears, concerning one and the same thing arise alternately, and diverse good and evil consequences of the doing, or omitting the thing propounded, come successively into our thoughts, so that sometimes we have an appetite to it, sometimes an aversion from it, sometimes hope to be able to do it, sometimes despair or fear to attempt it, the whole sum of desires, aversions, hopes and fears, continued till the thing be either done or thought impossible, is that we call DELIBERATION. . . In deliberation, the last appetite or aversion, immediately adhering to the action, or to the omission thereof, is that we call the WILL, the act (not the faculty) of willing.¹¹

But this is a very weak conception of agency that leaves the person nothing more than a spectator of this conflict of desires. We might well think that if we are really to be genuine agents, we must display certain kinds of activity with respect to our beliefs and desires. (Nor is this is merely a matter of having second-order knowledge of one's

beliefs and desires, since this is open to the Hobbesian deliberator--indeed it is assumed by Hobbes to be part of the picture.) I must, for example, notice conflicts or tensions among my beliefs and desires, consider how to resolve them by considering consequences of various courses of action, call to mind other relevant beliefs and desires and so on.¹²

The line of thought here moves in the following direction. In order to do what a genuine agent should do, I must have knowledge of my own beliefs and desires. What comes first is the fact that I have knowledge of, or special access to, my beliefs and desires. Given this ability, I can go on to engage in the sorts of activities described above. But some have suggested that the line of thought should run in the opposite direction: to put it crudely, it is not because I know my own mind so well that I can make up my mind; it is because I make up my mind that I have special access to it.

When I avow a belief, I am not treating it as just an empirical psychological fact about me; and to speak of a transcendental stance toward it is meant to register the

fact that . . . as a commitment, it is not something I am assailed by, but rather is mine to maintain or revoke.¹³

A final worry to which our lack of awareness of our own minds might give rise is this. If it is right that there is an intimate link between having self-knowledge and being an agent, then, as Sydney Shoemaker suggests, a failure of self-knowledge may undermine our sense that we are a single agent.

Let us refer to the person performing the bulk of the actions realized in the movements of this body as 'the agent' and let us refer to the person who professes introspective ignorance as to what is going on as 'the agnostic.' If anything would make it reasonable to say that the agent and the agnostic are one and the same person, it would be the fact that what the agent is doing fits with and is 'rationalized' by, beliefs and desires that can be independently ascribed to the agnostic.¹⁴

Section 3: Plato's Solution to the Puzzle of Apparent Akratic Action

Keeping these issues in mind, let us turn to the Protagoras passage itself (351B-357E). This passage begins with a simple descriptive account of what some people say happens in certain cases.

They maintain that many people are unwilling to do what is best, even though they know what it is and are able to do it, but do other things instead . . . they say that those who act that way do so because they are overcome by pleasure or pain or are conquered [by spirit or love or fear]. (Prot. 352D6-352E2)

But Plato proceeds to make it clear that these people are giving a description from the inside of an alleged case of akrasia. And what Plato must explain is this experience of theirs (touto pathos 352E6, cf. 353A4-5), that is, not just something they describe, but something they themselves undergo (cf. Prot. 353C4-5).

So what is the experience of theirs as they describe it?

X knows that A is best overall for X and better overall for X than B, knows that he can do either A or B, and does B under the influence of pleasure, pain etc.¹⁵

So this is not a case in which the person is pursuing pleasure, but simply has no view about whether in doing so he is also pursuing the good. As described, this is a stronger case in which he thinks that he is pursuing the pleasant and rejecting the good. (Reductionist accounts of motivation are not committed to this possibility merely in virtue of being reductionist. Even if all my desires are the expression of, e.g., the will to power and I am unaware of this fact, the typical person will usually not have an explicit belief about whether this is the case and may often not even have any implicit belief about the issue.) But this is, Plato thinks, the wrong description of the case and he offers a better one. On Plato's account, when the pleasures are temporally close, they seem greater than they are (and greater than they seem when they are at a greater temporal distance.)¹⁶ This "power of appearance" accounts for our choice (Prot. 356C8-E4) and our later regret, since when the pleasures are no longer in the short-run future, our judgment is no longer distorted.

So what exactly is going on? Here is a standard way of telling the story: I do not wish to take a position on whether this interpretation is correct, but I do want to suggest that it is not necessitated by the text and that Plato will find it hard to stick to this story. It runs as follows. At t1, the person judges that A is best and better than B and so desires to do A. At t2, the person judges that B is best and better than A, so desires to do B and does B.¹⁷ The judgment that B is best is false, but is explained by the person's overestimating the near-term pleasures involved in B. At t3, the person judges that A is better than B and regrets choosing and doing B. On this interpretation, there is no synchronic knowledge or belief akrasia, that is, at the time of action the agent does not believe or know that something other than what he is doing is best. Diachronic belief akrasia is possible in that such switches can occur, on this story, when the person only has belief about what is best at t1. Diachronic knowledge akrasia is not possible, since the person will not suffer such switches if he has knowledge at t1.

Now this is, I think, a coherent story and it is possible that it is the one intended in the Protagoras. But it does have certain problematic features. To begin, note that this

cannot be the full story. To try to satisfy the initial description of the problematic experience, at t_3 the person must also judge that at t_2 he believed that A was best and better than B. This is a false belief, but it is nevertheless one that the person holds. Next, it dismisses the apparent phenomenology of the experience to be explained. The person feels a conflict and we need to give a satisfactory explanation of this. This issue is especially pressing, since the gap in time between t_2 and t_3 can, it seems, be very small. (A committed vegan can feel regret immediately after ordering a café latte and before drinking it.) Since this is so, we need an explanation of two things.

First, we need an explanation of why the person goes wrong with respect to what his beliefs and desires were just a short time ago. (A) We might suggest that this change is rapid, but this hardly seems sufficient. There are many cases in which I am aware of rapid belief changes. (E.g. while playing a game of speed chess, I am aware of many options and certainly seem able to be very confident that I changed my mind during deliberation and can be confident about what many of these changes were.) (B) We might see this as a case of person rationalizing his behavior *ex post facto*, but becoming committed to a false

belief in the course of doing so. But should we not expect the person to find that a plausible reason for his behavior is that at the time of acting he thought that it was in fact the best option? It seems that rationalization should produce the opposite result from what the theory requires. (C)

There may be other interesting options here, but let us consider the "power of appearance (hê tou phainomenou dunamis, 356D4):" there are issues about exactly what the power of appearance is and how it works. Nevertheless, the power of appearance's basic feature is to cause certain pleasures to appear greater than they are. If its tendency is to make B (falsely) appear best to the person, it is hard to see why it should also cause forgetfulness of this.¹⁸ At any rate, we would need an account of how it does so.

The second thing the standard interpretation needs to explain is this. The person can say "A is best" just before and just after acting. Why cannot the utterance continue through the time of choosing and acting? It is not obviously sufficient to say that this is impossible, since the person at the time of choice believes that B is better and chooses B as best. This reply relies on the assumption (which may seem so compelling that it is not made explicit) that the person's continuing to say sincerely (aloud or mentally) that

"A is best" is excluded by the fact that he believes that B is better. But our confidence in this assumption should be diminished once we have allowed that a person can be mistaken about his very recent beliefs. What excludes the possibility that a person might sincerely say one thing, but really believe another? In particular, what excludes the possibility that a person might sincerely say "A is best" while in fact believing that B is best? And what excludes the closely related possibility that the person might think that he believes that "A is best" while in fact believing that B is best? Why cannot the person be mistaken about the beliefs that he has at the moment? This failure of self-awareness is not merely an idle possibility, but one that Plato needs to take seriously, since (a) as we have seen, the standard story is committed to such failures over a short period of time, and (b) in other nearby dialogues Plato is committed to related sorts of failure of self-awareness. There are two apparent examples in the Gorgias, for example, some with respect to beliefs and others with respect to desires in which the person is mistaken about his current mental states.¹⁹

The standard interpretation may be Plato's position in the Protagoras. And we might not be surprised that it is

unsatisfactory, since Plato, at least by the time of the Republic (if not earlier), came to find it unsatisfactory. But we may be able to see how to construct a better story that is not inconsistent, I think, with the basic ideas of the Protagoras and the other early dialogues and points to the Republic position or, at any rate, beyond the Protagoras.

Let us begin by considering the claim we took note of in section 2 that I know my own mind because I make it up. Suppose that I have often undergone the experience that I think of akratic action, e.g., I have often found myself giving in and going for a café latte. Further, suppose that I decide to give in this time and act on my desire for pleasure. Here, on Plato's theory, is a fundamental limit on my ability to make up my mind (and by making it up to know it). I cannot bring it about that I act on a desire for pleasure that is not aimed at what is overall best.

Plato now has two options. First, he can say that in this case I might form a choice or decision at t_1 , remember at t_3 vividly and with great confidence that I so acted (or went along), but that at t_2 I in fact acted on a different judgment of what is best.²⁰ We have seen some of the drawbacks of this story. But Plato has another option (whether or not he took it in the Protagoras.) Note that

the above limits on my being the author of my mental states shows that my conscious reflection and decision about what I am going to aim at is, in fact, insulated from my desire for the good upon which I act. My reflection and decision cannot bring it about that my desire aims at what is pleasant as opposed to what is good.

What is it that can make this the case? One plausible explanation of why this is so might posit a mechanism for forming and sustaining desires that is independent of my conscious thoughts in the following ways.

- (1) The generated desire always aims at the good and cannot have this object changed by conscious reflection.²¹
- (2) The object of my desire is fixed by the mechanism and the operation of the mechanism is not fully open to conscious reflection in such a way that I can always say correctly what I am desiring.

Both of these are claims to which Plato in the early dialogues should be sympathetic. With respect to (1), Plato does think that desires always aims at the good. He does not clearly recognize the possibility that a person might just give up and try to go for the less good or worse option. But it seems clear that he would think that even if such an intervention is possible, desire would still aim at

the good. With respect to (2), we have seen that Plato accepts that we can fail to recognize the real object of our desire over a very short time interval. If this is so, there does not seem to be a principled reason for rejecting the possibility of mistake at the moment of desiring.

If Plato were to accept (1) and (2), then there would be no need for the agent to stop saying "A is best" at the time of action.²² He can still say this, but he is wrong about what he really believes and is doing. So even if this is not the account that Plato actually gives in the Protagoras, is it a possible story given the rest of his commitments? A significant advantage of this account is that it does not need to posit the odd nearly instantaneous forgetting required on the standard interpretation. But I note briefly a few immediate concerns about this second story.

(1) Is this not *akrasia*? Is not the agent acting against his judgment of what is best? This is perhaps not a fatal worry. We find a conflict between an apparent belief and a real belief elsewhere in the early dialogues, e.g. in the Gorgias. What we need is some reason to privilege the belief that B is best and the desire-like item related to it. Perhaps one reason for doing so is that the belief that B is

best and the desire for B as best are the outcome of the calculation that the agent makes at the time of choice and action.²³

(2) On this story, the person would be saying "A is best," but would really believe that B is best and would be acting on a desire for B as best. Do we have to attribute to the person a desire for A at the time of action? If so, it seems that we have two sources of desire and this would seem to put us well on the road to something like the Republic position. It is, however, not phenomenologically so implausible to think that at the time of acting I have lost my desire for A, while I still say that it is best. This story also has some plausibility, if we think that desires are formed as the outcome of calculation about what is best.

(3) One might object that it is only in a shallow way that the agent fails to know her own mind in the interpretation that I have sketched. Suppose that John is the head of the spy ring and I desire to arrest the spy chief, but falsely believe that Jane, and not John, is the spy chief. One might say that, in some way, I do desire to arrest John and that my ignorance of this is a kind of failure to know my own mind. But such a failure of self-knowledge is a superficial kind of failure, if it is a failure of self-

knowledge at all. What I am unaware of is, rather, one might think, the fact that John is the spy chief which is a fact about the world and not my mind. Similarly, this objection runs, what the apparent akratic of the Protagoras really fails to believe is that pleasure is the good which is a putative fact about the world and not his own mind. But this misdescribes the case. It is essential to Plato's solution that at t2 the agent's desire is directed at the good as such or is under the guise of the good. The agent's conscious unawareness of this a failure to understand her own mind. (For similar reasons, neither is it the case that the agent's ignorance is simply of the fact that she believes that pleasure is the good.)

(4) Finally, we should note an intriguing passage near the end of Plato's argument.

Is it not the power of appearance that causes us to wander, often causing us to take things topsy-turvy and to regret our actions and choices with respect to things large and small? But the art of measuring, would have made this appearance ineffective [ἄκυρὸν μὲν ἂν ἐπιήσει τὸ τὸ φάντασμα], and by showing us the truth would have

brought our soul into the repose of abiding in the truth and would have saved our lives.²⁴ (Prot. 356D4-E2)

What is intriguing here is the claim that the measuring art makes "this appearance ineffective (akuron)."²⁴ It is not clear how much weight we should place on this phrase, but making the appearance akuron is not the way that one would expect Plato to say that the appearance has simply been destroyed, that is, that it no longer exists. akuron is surprisingly rare in Plato, the only other reference in the early dialogues is Crito 50B4.²⁵ There the personified Laws of Athens suggest that Socrates running away would render the verdict of the court akuron. Neither Socrates' disobedience (nor even generalized disobedience) would take the verdict off the books or render it legally void, what it would do is make it the case that the verdict, although present and persisting, would not determine what actually happens. If the parallel holds, what this passage suggests is that even in a person with the measuring art an appearance could be present, it would just not determine how the person acts. But if Plato were to allow this, we seem to get two potentially conflicting sources of

judgments and to be on the way to the middle-period position.²⁶

Section 4: The Implications of the Failure of Self-Knowledge

The literature on the importance of self-knowledge we noted in section 2, gives us, I think, a better understanding of the defects of the mistaken agent in the Protagoras. (He is mistaken both in that he makes the wrong choice and in that he thinks he is subject to akrasia, but is not. For ease of reference, I shall call him the "akratic" in scare quotes.) This literature suggested three sorts of worries to which a lack of self-awareness could give rise.

I. The failure to have knowledge of one's own beliefs and desires tends to undermine the efficacy and rationality of the agent's deliberation and associated deliberative processes. Here I note three ways in which this helps illuminate the situation of the self-described "akratic" in the Protagoras.

(1) The "akratic" thinks that his knowledge about what is best (or belief about what is best) is sometimes overcome by his fears, desires, spirited emotions and so on. So he thinks that the fundamental problem that he has does not

lie in the information that he has or in his awareness of this information, but rather in the forcefulness of his non-rational motivations. This is especially so, since the "akratic" thinks that akrasia operates not only against true belief, but also against knowledge. If it operated only against true belief, then at least the possibility would remain open that an improvement in his cognitive state to knowledge would prevent akrasia precisely in virtue of this epistemic gain. Thus the "akratic" will be subject to two important impairments. First, he will not be motivated to seek further information or to improve his epistemic state, since he thinks that this will not solve the problem (cf. Prot. 357E2-4). Note that a person who was aware of the correct psychological theory in Plato's view could handle cases of apparent akrasia more successfully, although this would require an odd sort of distancing from one's own mental states. Granting that knowledge of the measuring art renders cases of apparent akrasia impossible, it is still possible that one might both accept Plato's psychological theory and falsely believe that one possesses the measuring art. In such a case, an instance of apparent akrasia would allow the agent to infer that he really did lack knowledge, although he might not be able to avow that his apparent

desire for pleasure was really for the good or to find any direct epistemic reason for doubting his own knowledge claims.

The second problem that such an "akratic" will have is that in light of his belief that improved information will not help, he will either take no direct steps to try to prevent further instances of akrasia or will attempt to boost the motivational force of his good desires and diminish those of his "bad" desires. Such steps are likely to be ineffective and will, in any case, be a distraction from what he needs to do.

(2) The "akratic" in such a case will be unable to bring his true beliefs about what is good appropriately to bear on the situation at hand.²⁷ Rehearsal of the reasons why a particular case of action is better overall will not have the appropriate result. Here, however, this failure of integration does not seem simply to be a result of the person's ignorance. (Indeed, because the person misconceives the object of the akratic desire it seems quite possible that he will not even try to integrate it with his beliefs about what is best.) Rather, perhaps the most plausible way to explain the situation is that the generation of a desire with a particular object is not (at least fully)

available to conscious control and to integration with the rest of the person's beliefs. This will at least be the case for agents who do not possess the measuring art. But as we shall see shortly, it is not so clear that Plato should be confident that agents possessing the measuring art are not subject to similar defects.

(3) Finally, it is not clear whether the fact that the akratic desire is directed at the good or that the nearness of the pleasure influences the judgment of goodness are things that the agent can ever avow as opposed to acknowledging on the basis of a psychological theory. If they can be avowed, then it is especially hard, I think, to provide a plausible mechanism that would result in this being forgotten and replaced by a false belief as it should be on the standard story. If the short-run calculation is conscious, how could it be instantly forgotten? But if they cannot be avowed, we need an explanation both of this fact and of the reason why they show the (semi-) rational consistency that they do. As I suggest below, related problems give us reason to think that it may be plausible to split up the agent.

II. As we have also seen, failures of self-awareness threaten to undermine the idea that the person is an

agent. I shall note three points that are especially relevant to the Protagoras. But before doing so, let us remember that Plato's basic view about the direction of our desires to the good imposes in itself some real limitations on our agency. No matter what we may think we are doing (or intend to do), we cannot, on Plato's view, succeed in forming a desire for something other than what we think is good.

First, as we have just seen, the "akratic's" second-order beliefs and desires will be subject to various kinds of malformation and thus the "akratic's" activity with respect to his own beliefs and desires is likely to be defective. The "akratic's" agency will be undermined insofar as he is moved to action by beliefs and desires that he does not and perhaps cannot understand. This is true at least for the "akratic" who is not aware of the correct diagnosis of his situation.

But serious problems still face the "akratic" who has some degree of theoretical insight into his situation. Some degree of awareness of his real situation is not sufficient to eliminate the power of appearance, and in the "akratic" beliefs about what is best overall will, at least in some important cases, not interact rationally with the beliefs

involved in the desire for the near-run good. If the beliefs about and desire for the near-run good are to be thwarted, the person will have to rely on various strategies of self-manipulation. E.g. in addition to avoiding occasions of temptation, he might try to make vivid in his mind any short-run bad consequences, dwell upon the long-run bad consequences and in doing so perhaps exaggerate them (cf. Tim. 70D8-71B5). But in doing so, his aim may not be to provide further information that will appropriately and rationally interact with that involved in the genesis of the "akratic" desire, but rather to produce the right causal consequences.

Finally, we have so far proceeded on the assumption that the person can and should be confident that his more persisting judgments are correct or at least better founded. But it is not clear how confident the person should be about this. The psychological theory that the person has come to accept suggests that his introspective beliefs are quite unreliable and that the considerations embodied in the "akratic" desire are relatively isolated from the rest of the person's beliefs. Perhaps we can develop reasons for thinking that the considerations diachronically more available are not just psychologically more stable, but are

epistemically more reliable. But the existence of a relatively stable independent point of view for evaluating actions should give us some concern.

III. The third and final problem concerns the unity of the agent. Although Plato is committed to the unity of the agent in the Protagoras, I think that some of the difficulties surrounding Plato's views call this unity into question.

First, on the criteria that Shoemaker gives, the unity of the person is called into question since what the agent (the one going for the worse option thinking it best) is doing is not rationalized by the beliefs of the agnostic (the one who falsely believes that he is acting akratically.) The pursuit of the wrong option is precisely what the agnostic thinks is an instance of the overcoming of reason by desire or emotion. Such disruption of the unity of the person is not an inevitable result of any and all cases of akrasia, but is present here because the person's lack of awareness of the real object of his desire. Nor is it the case that partitioning of the person inevitably produces such serious instances of a failure to know one's own mind. In the Republic, for example, the person (or the Reasoning part of the person) is well aware that the motivations that can

prompt to genuine akrasia are not aimed at the overall good.²⁸

But even part from Shoemaker's criteria, we find, I think, reasons internal to the Protagoras' theory to think that it points towards the splitting of the self. As we have seen the most plausible explanation for (1) the failure of the "akratic" desire to be sensitive to other relevant information, and especially, (2) the failure of the person to recognize the real object of his desires suggests some significant degree of insulation of the process of belief and desire formation from conscious reflection and thought. If it is to be at all plausible that the person thinks, shortly after making the wrong choice, that he did not act on a desire for the best, but rather on one for pleasure it should be the case that the calculation influenced by the nearness of the reward is not fully available to consciousness. The items do not interact, and perhaps cannot be brought to interact, in the ways that we expect items within one mind to interact.

Once we take the further step of seeing how each set of beliefs and desires hang together and interact appropriately, we are on the way to the sort of the division of the self that we find in the Phaedo and the Republic.

Before turning to my conclusion, let me note three points about this development. In particular, there are three unsatisfactory aspects of the Protagoras position, reflection on which, may have led Plato in the direction of the Phaedo and the Republic.

(1) On the Protagoras account, the object of "akratic" desire is still the overall good, although the ability to process and take in new information in whatever mechanism produces the desire is very limited. Once we allow desire to be generated by such a simple mechanism, it is more plausible to think that it might respond to one desirable feature without an attempt to reach any sort of overall judgment, much less one of goodness.

(2) On the Protagoras account, the main reporting consciousness has a less committed attitude towards the overall good than that embodied in the desire. After all, the main reporting consciousness thinks it is possible for it itself to act against the good and thinks that it does so. The Phaedo-Republic picture allows for recognition of such motivations, that is, desire for something other than the overall good, but separates them off from reason's own activity.

(3) Finally, as we saw, the Protagoras is committed to the idea that knowledge of the measuring art is sufficient to render the power of appearance ineffective. (As we also saw, it was not entirely clear whether this means that no appearances are left.) But once the "akratic" desires and the mechanisms generating them are granted some autonomy from conscious reflection and control, it is difficult to see why this should be so.

I shall conclude with a brief suggestion about what these reflections can help tell us about Plato's conception of practical rationality. Care must be taken here: Plato draws no sharp distinction between theoretical and practical rationality and, indeed, we would need a more nuanced notion than simple rationality. Nevertheless, Plato thinks that reason can be deployed in finding out how to act or how to live. This will involve both the use of reason to determine what is good and to attain what is good. The measuring art, if it includes or is supplemented by knowledge of what the appropriate unit of measurement is, may embrace both of these. But what our reflections have suggested is that at least in those who do not fully possess the measuring art, a grasp of what is good and of how to attain it is not sufficient for guiding one's life by

reason as best as one can. What is also required is something like a grasp of a true psychological theory, that is, some grasp of what goes on in the soul and how what goes on in the soul will appear to the agent. This is, I think, a lesson to which modern theories of practical rationality might pay greater heed.

Bargh, John and Chartrand, Tanya. 1999. "The Unbearable Automaticity of Being." American Psychologist 54, 462-79.

Bobonich, Chris. 2002. Plato's Utopia Recast. Oxford University Press.

-- (forthcoming). "Socrates and Eudaimonia." The Cambridge Companion to Socrates, ed. D. Morrison (Cambridge University Press, 2011), pp. 293-332.

Cooper, J. (ed.) 1997. Plato: Complete Works. Hackett.

Devereux, Daniel T. 1995. "Socrates' Kantian Conception of Virtue." Journal of the History of Philosophy 3, 381-408.

Geuss, Raymond. 1981. The Idea of a Critical Theory: Habermas & the Frankfurt School. Cambridge University Press.

Hobbes, T. (ed. E. Curley). 1994. Leviathan. Hackett.

Irwin, T. 1977. Plato's Moral Theory. Oxford University Press.

Lamb, W.R.M. 1977. Plato II: Laches, Protagoras, Meno, Euthydemus. Harvard University Press.

Moran, Richard. 2001. Authority and Estrangement: An Essay on Self-Knowledge. Princeton University Press.

Newman, Lex. 2000. "Descartes' Epistemology." The Stanford Encyclopedia of Philosophy (Fall 2000 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2000/entries/descartes-epistemology/>.

Nisbett, R. and Ross, L. 1980. Human Inference: Strategies and Shortcomings of Social Judgment. Prentice-Hall.

Nisbett, Richard E. and Wilson, Timothy D. 1977. "Telling More Than We Can Know: Verbal Reports on Mental Processes." Psychological Review 3, 231-259.

Penner, Terry. 1991. "Desire and Power in Socrates: The Argument of Gorgias 466A-468E that Orators and Tyrants Have No Power in the City." Apeiron 24, 147-202.

— 1996. "Knowledge vs. True Belief in the Socratic Psychology of Action." Apeiron 29, 199-230.

— 1997. "Socrates on the Strength of Knowledge: Protagoras 351B-357E." Archiv für Geschichte der Philosophie 79, 117-149.

Penner, Terry and Rowe, C.J. 1994. "The Desire for the Good: Is the Meno Inconsistent with the Gorgias?" Phronesis 29, 1-25.

Reshotko, Naomi. 1991. "The Socratic Theory of Motivation." Apeiron 25, 145-170.

Shoemaker, Sydney. 1988. "On Knowing One's Own Mind." Philosophical Perspectives 2, 183-209.

— 1990. "First-Person Access." Philosophical Perspectives 4, 187-214.

— 1994. "Self-Knowledge and 'Inner Sense': Lecture II: The Broad Perceptual Model." Philosophy and Phenomenological Research 2, 271-290.

Sidgwick, Henry. 1988. Outlines of the History of Ethics. Hackett.

Taylor, C.C.W. 1976. Plato Protagoras. Clarendon Press.

Velleman, J. David. 2000. The Possibility of Practical Reason. Clarendon Press.

Wilson, Timothy D. 2002. Strangers to Ourselves. Harvard University Press.

¹ Sidgwick (1988), p. 24.

² Velleman (2000), p. 99.

³ E.g. Penner (1991) and Penner and Rowe (1994).

⁴ E.g. "pursuit" (diôkein) is restricted to the actual good at Gorg. 468B3 and seems to be a state at least involving a desiderative component. At Prot. 354C4 "pursuit" (diôkein) is said to be the state that people find themselves in with respect to the perceived pleasant. At Prot. 358C6-359A1, what one willing (ethelein) to go for or goes for willingly (hekôn) is what one thinks best; and here "to go for" (ienai, erchesthai) seems to be equivalent to "pursue."

⁵ Translations of the Gorgias draw on, with modification, that of Zeyl in Cooper (1997); translations of the Protagoras draw on, with modification, that of Lamb (1977), Lombardo and Bell in Cooper (1997) and Taylor (1976).

⁶ Reshotko (1991), p. 151 n. 8.

⁷ Geuss (1981), p. 81.

⁸ Nisbett and Wilson (1977), p. 231. For more recent work, see Bargh and Chartrand (1999) and Wilson (2002).

⁹ Newman (2000).

¹⁰ Shoemaker (1990), pp. 206-7.

¹¹ Hobbes (1994), p. 33 (i.e. Part 1, chapter 6, para. 50-3), capitalization and emphasis in the original, but I have slightly modified the punctuation.

¹² Shoemaker (1988), pp. 186-7.

¹³ Moran (2001), p. 89. I am indebted here to conversations with Krista Lawlor.

¹⁴ Shoemaker (1994), p. 280.

¹⁵ The most interesting and problematic cases of akratic action in contemporary accounts are those in which the person acts voluntarily or intentionally. For some discussion, see Bobonich (2002), pp. 532-3 n. 37. The issue in the Protagoras is first presented as one about the strength of knowledge (epistêmê). But Socrates also thinks

that it is not possible to act contrary to what you simply believe best at the time of action (Prot. 358B6-E6).

¹⁶ It is sometimes suggested that in order to explain the change in judgment at t2 we need to posit the onslaught of a non-rational desire for B sometime between t1 and t2. But I do not think that this is necessary. It is not unreasonable to think that the proximity of a smaller good can focus the person's attention on it and thus lead to a miscalculation, see Nisbett and Ross (1980), pp. 49-51 on proximity effects.

¹⁷ Cf. Penner (1996) and (1997). Penner, to the extent that I can understand his views, usually thinks that the person does not desire B (it merely seems good to him) because desire is always directed at the actual best, e.g. "if [an action] does not result in one's real good . . . one does not want to do the action", Penner and Rowe (1994), p. 8. emphasis in original. I am not sure how to reconcile this with Penner's (1997), p. 139 claim that "no one will want to go towards things they fear, since they think those things bad." On the general issue, see nn. 3 and 4. We can, however, allow a desire for B as part of the standard story as in the text above.

¹⁸ I suggested (n. 16) that it is not clear that we need to invoke non-rational motivations to explain why B appears more pleasant at t2. But even if such non-rational motivations do exist, they have as a primary feature a tendency to cause a best judgment. We would need to invoke some further feature to explain why they would also cause forgetfulness of that best judgment. We might, of course, hold that some form of repression causes the forgetfulness, but we would then need an account of the mechanism of repression and on many theories of

repression this will require psychological subsystems with sophisticated abilities.

¹⁹ E.g. Gorg. 473E-474B, 466A-468E. I leave open here what the exact relation is between these claims in the Gorgias and the interpretation I am considering for the Protagoras.

²⁰ Although Plato does not sharpen choice or decision into technical notions in the early dialogues, I think that the hypothetical event described is sufficiently intuitively clear.

²¹ This is intended to be compatible with both of the following claims being true of any agent at the time of action: (a) "S desires what is actually best", and (b) "S desires what seems best to S." But a similar issue, I think, would arise if we held that all desire is for the actual good. On a traditional understanding (e.g. Irwin (1977, 78-82), Plato in the early or Socratic dialogues thinks all desire aims at the agent's own final good, i.e. his optimal happiness, in such a way that the agent always desires what he believes best overall for himself and does not desire anything he believes incompatible with this. My alternate explanation of the Protagoras' rejection of the possibility of akrasia, as well as what I call the "standard interpretation" (with the qualification in n. 17), accepts this traditional understanding. There has been much recent discussion of the issue which I cannot enter into here, (e.g. Devereux (1995), Brickhouse and Smith and Reshotko (this volume)), but see Bobonich (forthcoming). I do not think that the recent criticisms have provided sufficient reason for rejecting the traditional understanding of Plato's psychological eudaimonism, cf., e.g., nn. 16 and 18. In any case, none of the criticisms are intended to deny that the desire formed under the "power of

appearance" is for the best or is formed under the guise of the good.

²² Perhaps he might even be able to say "I am going for pleasure, not for what is best."

²³ This belief is also the one acted upon, but being acted upon does not seem to be sufficient to privilege the belief that Polus presumably acts upon in the Gorgias.

²⁴ Strictly speaking, this appearance is a literal visual appearance of the size of physical objects. But Socrates' use of the this analogy suggests that a similar point should hold concerning appearances about the size of pleasures.

²⁵ There are only five other instances in Plato, Laws 715D3, 929E6, 954E6 and Tht. 169E2 and 178D9.

²⁶ But it is worth noting that this possibility does not require the existence of two distinct judgment-makers: all that it requires is that the content of a previous judgment still, in some way, remain "live" after a new one has been formed.

²⁷ Might one have knowledge of what to do without possessing the measuring art? (Perhaps Plato would hold that one either has full knowledge of the measuring art or one has no knowledge at all of what to do, although this is not obviously entailed by his discussion in the Protagoras.)

²⁸ The person's Reasoning part will not, it seems, have first-person access to the states of the lower parts. I would like to thank David Blank, Kellie Brownell, Corinne Gartner, Rachana Kamtekar, Sean Kelsey, Krista Lawlor, Seana Shiffrin, and David Velleman for their comments and suggestions.