Ancient Philosophy 35 (2015)

Two Interpretations of Socratic Intellectualism

Thomas A. Blackson

The ancients thought that 'reason has desires', but what they had in mind is not obvious. The likely alternatives turn on what they thought about beliefs. They may have thought that some beliefs are motivating and that all motivation in terms of reason stems from belief, or they may have thought that no beliefs are motivating and that all motivation stems from desire.[1]

These possibilities allow for competing interpretations of Socratic intellectualism. For convenience, I call them the D and B interpretations.[2] According to the D interpretation, the human psyche contains a standing desire for the real good. Further, all motivation in human beings ultimately stems from this desire.[3] According to the B interpretation, there is no such desire in the human psyche. Instead, all motivation in human beings ultimately stems from beliefs of a certain sort and thus has its source in epistemic cognition and in reason.

Which interpretation is correct? The D interpretation is now much better known,[4] but the famous passage in the *Protagoras* in which Socrates considers

[1] Michael Frede is the classic source for this interpretative framework for understanding desire and reason in the ancients. 'The assumption is that at least some desires, like the desire to know the truth or to obtain what is thought of as good, are desires of reason itself, rather than desires reason merely endorses. It may also be part of this aspect of the notion of reason that reason itself not only has desires, but that the objects of its desires to some extent are fixed, so that it becomes part of what it is to be endowed with reason to have certain preferences, at however high a level of generality these might be fixed… Plato and Aristotle departed from [the Socratic] view by introducing desires which are irrational in the sense that they do not have their origin in reason, but in an irrational part, or irrational parts, of the soul which has a certain degree of autonomy. Thus what one feels or desires may be independent of what one believes. But, though, Plato and Aristotle, unlike Socrates, are willing to grant this, they still hold on to the view that some desires are desires of reason. It is unclear whether this, upon further analysis, turns out to be more than the claim that there are thoughts or beliefs of such a kind that the mere having of the thought or belief on its own is a sufficient motive to act' (Frede 1996, 6-7; cf. Frede 1986, 96; Frede 1992, xxx; Frede 2000, 8).

[2] These names are not part of the current literature. I introduce them, unimaginatively, as abbreviations for the 'Desire' and 'Belief' interpretations of Socratic intellectualism.

[3] Rowe 2007, 23 states the main lines of the interpretation: 'Briefly, and at bottom, it consists in the claims (a) that all human agents always and only desire the good; (b) that what they desire is the real good, not the apparent good; and (c) that we what we do on any occasion is determined by this desire together with whatever beliefs we have about what will in fact contribute to our real good. Hence the label "intellectualist": we only ever do what we think will be good for us.' Rowe develops his interpretation in collaboration with Terry Penner. See Penner 1991, Penner and Rowe 1994, Rowe 2002, and Penner and Rowe 2005, 216-230 (see also Reshotko 2006).

[4] Taylor 2000, 62-64 states a version of the **D** interpretation: 'The basis of the theory is the combination of the conception of goodness as that property which guarantees overall success in life with

whether knowledge is a leader and ruler, as opposed to something that can be dragged around as a slave, can easily seem to favor the B interpretation. Moreover, historians have traditionally thought that if any set of passages in the Platonic dialogues expresses the views of the historical Socrates about the nature of motivation in human beings, it is these passages in the *Protagoras*.[5] The B interpretation, however, has not emerged as the consensus interpretation of Socratic intellectualism in the *Protagoras*.

This would be disconcerting were it not for the interpretative assumptions that have framed the discussion. To some of the most prominent historians of ancient philosophy, it has seemed unlikely that Socrates or Plato would have abandoned the view that all motivation is a matter of desire and thus would have traded this idea for the view that motivation in human beings always stems from beliefs of a certain sort and thus has its source in epistemic cognition and in reason.[6] And so, with respect to the question of Socratic intellectualism in the *Protagoras*, the B interpretation has been at a disadvantage. Because the psychology in the B inter-

the substantive thesis that what in fact guarantees that success is knowledge of what is best for the agent. This in turn rests on a single comprehensive theory of human motivation, namely, that the agent's conception of what is overall best for him- or herself (i.e. what best promotes *eudaimonia*, overall success in life) is sufficient to motivate action with a view to its own realization. This motivation involves desire as well as belief; Socrates maintains (*Meno* 77c, 78b) that everyone desires good things, which in context has to be interpreted as the strong thesis that the desire for the good is a standing motive, which requires to be focused in one direction or another via a conception of the overall good. Given that focus, desire is locked onto the target which is picked out by the conception, without the possibility of interference by conflicting desires. Hence all that is required for correct conduct is the correct focus, which has to be a correct conception of the agent's overall good. On this theory motivation is uniform, and uniformly self-interested; every agent always aims at what he or she takes to be best for him- or herself, and failure to achieve that aim is to be explained by failure to grasp it properly, that is, by cognitive defect, not by any defect of motivation. Socrates spells this out in the *Protagoras*, on the assumption, which he attributes to people generally, that the agent's overall interest is to be defined in hedonistic terms… There is considerable disagreement among commentators as to whether Socrates is represented as accepting the hedonistic assumption himself or merely assuming it *ad hominem*…, but there is no doubt that…the view that the agent's conception of the good is the unique focus of motivation (maintained also in the *Meno*) is Socrates' own. This account of goodness as knowledge thus issues directly in one of the claims for which Socrates was notorious in antiquity…, "No one goes wrong intentionally" (*Prot*. 345e).'

[5] For a statement of the traditional view, see Kahn 1996, 73-74. See also Vlastos 1988, 99.

[6] Kahn 1996, 227, 229, 242-243 says that the *Protagoras* seems to represent 'the extreme case of the general tendency of Socratic intellectualism to ignore the emotional and affective components of human psychology, or to reinterpret them in terms of a rational judgment as to what is good or bad'. But Kahn himself believes and insists that such 'a thesis of omnipotent rationalism seems patently false'. He argues that this reading of the *Protagoras* is 'naive' and that 'neither Plato nor Socrates in the *Protagoras* is guilty of ignoring obvious facts of human behavior or denying the complexity of motivation that is conceptualized for the first time in the psychological theory of the *Republic*'. Rowe 2002 says that Kahn is wrong to claim 'that the intellectualist model "implausibly reduces [human motivation] to a judgment concerning what is good"'. Rowe says that 'on any account of ("socratic", or Socratic) intellectualism, human motivation surely must also involve desire—a basic, universal, unthinking desire for the good'. So although Kahn and Rowe disagree about how to understand the *Protagoras*, they both believe it is obvious that motivation in human beings must ultimately stem from desire and that neither Socrates nor Plato could have thought otherwise.

pretation has not been regarded as something that either Socrates or Plato could have seriously entertained, it has not been seen as a viable alternative to the psychology in the D interpretation.

This disadvantage is unwarranted. The psychology in the B interpretation may be implausible by contemporary standards, but relative to the D interpretation, there is nothing uncharitable about it. And when this issue of charity does not tip the balance, the B interpretation is the more likely interpretation of Socratic intellectualism in the *Protagoras*. Given that Socrates is expressing a view about the human psychology that he believes,[7] it is the strong suggestion of his entire discussion with Protagoras about whether knowledge is a ruler and a leader that there are no motivational states in human beings that do not stem from beliefs of a certain sort.

I. The D and B Interpretations

When the D and B interpretations are understood as interpretations of a view of the human psychology that the character believes, they are part of an interpretation of the historical figure. According this interpretation, Socrates thought that human beings are psychological beings and that the human soul is a collection of states and processes that cause action.[8]

[7] This assumption is questionable, as Kahn 1996 clearly demonstrates he understands Plato along unitarian lines and gives an 'ingressive' (59) interpretation to explain the apparent inconsistency between the psychology in the *Protagoras* and the psychology in the *Republic*: 'I suggest that Socrates is here deluding the sophists with a rationalist theory of choice, just as he has deluded them with Laconic philosophy in the interpretation of Simonides' poem, and that the motivation is the same in both cases: to establish the paradox that no one is voluntarily bad, and hence that deliberately bad actions are always motivated by a false view of the good. This result was insinuated in the poetic episode and is now deductively argued for on the basis of the hedonist premiss and the rational model for decision. He is no more committed to the hedonism and the rationalist decision theory than he is to the virtuoso misinterpretation of Simonides' poem. The former, like the latter, is a device for presenting the paradox… If we thus avoid a naïve reading of this extremely subtle argument, we see that neither Plato nor Socrates in the *Protagoras* is guilty of ignoring obvious facts of human behavior or denying the complexity of motivation that is conceptualized for the first time in the psychological theory of the *Republic*' (242-243). Kahn and others may be right to understand the *Protagoras* along such unitarian lines. My argument is directed only to those who are not drawn to this sort of unitarian interpretation of the *Protagoras* and who do not think that 'much of Socrates' reasoning' to use Kahn's words, 'is manipulative and insincere' (242).

[8] Frede 1996, 19 presents this view of Socrates: 'historically the decisive step was taken by Socrates in conceiving of human beings as being run by a mind or reason. And the evidence strongly suggests that Socrates did not take a notion of reason which had been there all along and assume, more or less plausibly, that reason as thus conceived, or as somewhat differently conceived, could fulfill the role he envisaged for it, but that he postulated an entity whose precise nature and function was then a matter of considerable philosophical debate… [W]hat Socrates actually did was take a substantial notion of the soul and then try to understand the soul thus substantially conceived of as a mind or reason. By "a substantial notion of the soul" I [mean]…a notion according to which the soul accounts not only for a human being's being alive, but for its doing whatever it does, and which perhaps, though not necessarily, is rather like what we could call the self. This was not a common conception, it seems, even in Socrates' time, but it was widespread and familiar enough under the influence of nontraditional religious beliefs, reflected, for instance, in Pythagoreanism. And it seems to have been

This interpretation of Socrates traces its modern origins to John Burnet. He argues that 'Socrates was known as a man who spoke strangely of the soul.'[9] It is important not to exaggerate the novelty of Socrates' conception of the soul (see Lorenz 2009), but if his way of talking about the soul was at least partly responsible for his reputation for wisdom, as Burnet argues, then Socrates took a seminal step in what became a long-lived philosophical tradition of theorizing about human beings as psychological beings. Socrates thought that a human being can sometimes control his actions and hence can sometimes control the direction his life takes. This thought, in itself, would not have been at all unusual. The innovative step was in the explanation of how a human being controls his actions and thereby controls the direction his life takes. According Socrates, a human being controls his actions, and thereby controls his life, by exerting control over his soul.

It is part of this interpretation that Socrates did not have a detailed view of how the human soul functions. This would be a view about what states and processes are in the soul, which of these states and processes admit control, and which of them do not admit control because they are fixed in the soul. As Michael Frede has said, Socrates 'postulated an entity whose precise nature and function was then a matter of considerable philosophical debate'. This is important for understanding Socratic intellectualism in the *Protagoras*. How the states and processes function in the human soul to produce action was a matter of debate, and the D and B interpretations are different views of the psychology that Plato has Socrates introduce in the *Protagoras*.

II. The D Interpretation

In the psychology in the D interpretation, there are beliefs, there are desires, and neither is reducible to the other.[10] Further, one of the desires takes a special form. This desire is for the real good. This desire is not something human beings control. It is an invariant part of the human soul, and all motivation stems from this fixed desire for the real good.

such a substantial notion of the soul which Socrates took and interpreted as consisting in a mind or reason.'

[9] Burnet 1916, 161. When Burnet says that Socrates was 'known' to speak strangely about the soul, he relies on the passage in Aristophanes' *Clouds* where the denizens of the 'thought-factory (φροντιστήριον) are derisively called "wise ψυχαί"' (157). Claus 1981 criticizes Burnet's work on early uses of ψυχή, but Claus nevertheless comes to essentially the same conclusion about Aristophanes's use of ψυχή in connection with Socrates: that it is part of a 'parody of a rational notion of ψυχή' (159; see also Havelock 1972 and Handley 1956). For general discussion of the soul in early Greek thought, see Burnet 1916, 141-160, Furley 1956, Claus 1981, Bremmer 1983, Lorenz 2009, and Huffman 2009.

[10] One might distinguish beliefs and desires in a rough way in terms of 'direction of fit'. Agents change some psychological states to fit the world. For other psychological states, they change the world to fit the state. Given this much, one might say that the former psychological states are beliefs, that the latter are desires, and that no psychological state has both directions of fit. Penner and Rowe, as far as I know, do not engage in this sort of analysis of belief and desire. Instead, they appear to rely on what they take as the ordinary understanding of belief and desire.

This conception of the human psychology provides a straightforward way for practical cognition to achieve its aim of making the circumstances good for the agent, but it would be a mistake to conclude that the only way for practical cognition to achieve its aim is for the agent to have a standing desire for the real good. It is the aim of practical cognition to change the current situation so that it instantiates features that are good for the agent, and practical cognition achieves its aim by getting the agent to value these features. Different psychologies make this happen in different ways, and the psychology in the D interpretation is just one possibility.

In the D interpretation, practical cognition achieves its aim as follows. The desire for the real good is a standing desire. The real good is the goal, and the desire marks the acceptance of this goal and induces planning so that the agent forms beliefs about how to achieve the goal in the circumstances. The right action follows, given true beliefs, and this is the hallmark of Socratic intellectualism. Control over the soul is control over belief. Given true beliefs, whatever plan the agent adopts, and so intends to execute, is a plan to achieve the real good.[11] Thus, in the causal history of every action, there is a belief about the real good. As Rowe 2007, 23 says, in an intellectualist psychology 'we only ever do what we *think* will be good for us'.[12]

To see that the mechanism in the D interpretation is not the only possibility, it is helpful to imagine a non-intellectualist psychology. In this psychology, the desire for the real good is not a structural feature of the soul. Instead, the soul has a mechanism for proposing goals and adopting them by default. For example, when the agent is in the physiological state that constitutes being hungry, the mechanism proposes eating as a goal. This goal is accepted by default. The agent does not have a standing desire for the real good, and he does not form the belief that eating is the real good for him in the circumstances. Instead, a desire to eat arises automatically when he is hungry. This desire encodes the acceptance of the goal to eat, and it triggers either a habitual or a planned response. The goal is to eat, and the response consists in a sequence of actions to change the situation so that the agent is eating. Practical cognition thus achieves its aim of making the circumstances good for the agent, but the mechanism is different from the one in the D interpretation. This imagined psychology is coherent, but it is not an intellectualist psychology because there is not a belief about the real good in the casual history of every action.

The imagined psychology is thus not a candidate for the psychology in Socratic intellectualism, but in order to understand that the B interpretation is no more uncharitable than the D interpretation, the point to notice is that the aim of practical cognition is different from the cognitive states and processes that satisfy

[11] Bratman 1987 observes that intentions encode plan adoption (see also Bratman, Israel, and Pollack 1988).

[12] Cf. Vlastos 1988, 99: '[For Socrates in the early dialogues] the intellect is all-powerful in its control of the springs of action; wrong conduct, he believes, can only be due to ignorance of the good.'

this aim. This is important to keep in mind in going forward in the investigation into the *Protagoras* because it allows for the possibility that there is an intellectualist psychology that does not include a standing desire for the real good. If there is such a psychology, and I will argue that the psychology and cognitive design in the B interpretation is an example, then the D interpretation has a competitor.

III. The D+PR Interpretation

Penner and Rowe 2005 have formulated the most well-known version of the D interpretation. In their formulation, the D+PR interpretation, in addition to the desire for the real good, there is a special theory of action individuation and instrumental desire.[13] Penner and Rowe do not believe that this theory is explicit in Plato, but Penner has constructed what they take to be the underlying view. This construction is perhaps yet to be worked out completely, but the general contours of the D+PR interpretation are nevertheless reasonably clear.[14]

In the D+PR interpretation, although human beings have desires in addition to the standing desire for the real good, these desires do not work in the psychology in quite the way one might initially expect. It can seem natural to think that, in a given set of circumstances, a plan to achieve the real good may require the agent to accept various subgoals. Further, it can seem natural to think that the desires for subgoals stem from the acceptance of these goals. This, however, is not quite true in the D+PR interpretation. The agent has a standing desire for the real good. To act, he needs to engage in epistemic cognition to figure out what the real good is in the circumstances.[15] Suppose that he forms the belief that it is *g*. Once he accepts *g* as a subgoal, he forms a desire. What is this desire? According to Penner and Rowe 2005, 221, as part of their explanation for the Socratic thesis that 'no one errs willingly', it is 'the desire to do this action here and now which is

[13] Penner and Rowe 2005 also think that the desire for the real good is the desire for the agent's real good.

[14] The following remarks are representative. 'In the absence of any answer in the Platonic text, Penner has constructed an account which enables us to stick with desire for good as desire for the real good, while allowing for the operation of an executive desire in producing action—notwithstanding the fact that this new executive desire will not be a desire for the actual action done. There will be another, defective, sort of desire—which Plato might have called a "false desire"—that will bring about the action which the agent did' (Penner and Rowe 2005, 221). 'We now take up desire for means. If the preceding arguments suggest reason for saying that what one desires as one's end is one's real happiness rather than one's apparent happiness, we now need a reason for saying, as Socrates says in the *Gorgias*, that when one (voluntarily) does a particular action that does not result in maximizing one's real happiness, one didn't after all want to do that action. …Actually, we cannot offer a detailed answer here: It is far too large a question' (Penner and Rowe 1994, 8-9; see also Penner 2011).

[15] The D and the D+PR interpretations conform to the broadly Humean theory of motivation according to which desire is always necessary for motivation. Further, in these interpretations, motivation is always a matter of the standing or fixed desire for the real good. This desire is the starting-point for all motivation, but to generate a specific motivation, and hence an action, the agent must form a belief about what the real good is in the circumstances. These beliefs may vary from agent to agent. The desire for the real good does not. It is a necessary feature of all agents.

both the really best means to the agent's maximal happiness (maximal good) *and* the actual action done which the agent thinks to be the best means available' (see also Penner and Rowe 1994, 3-9). Hence, because actions are individuated broadly in terms of their consequences, their version of the D interpretation has the following very striking implication: an agent who has false beliefs about the real good does not perform the action his desire encodes.[16] Given the theories of instrumental desire and action individuation in the D+PR interpretation, and given that the agent is mistaken about what the real good is in the circumstances, what the agent does fulfills none of his desires: 'the agent does not want to do the action he or she is doing—the one that will turn out not to maximize the agent's available happiness or good' (Penner and Rowe 2005, 217). He goes wrong. He does something that does not bring about the real good, but he does not 'err willingly'.

The D+PR interpretation is ingenious philosophically, and it is also an important contribution to Platonic scholarship because any adequate interpretation of Socratic intellectualism in the *Protagoras* must be consistent with thesis that 'no one errs willingly'. Socrates famously says that 'no one goes willingly (ἑκών) toward the bad or what he believes to be bad' (*Protagoras* 358c7; cf. 345e). His meaning is not transparent, but the idea appears to be that if someone brings about something bad, then what he has brought about is somehow not what he aimed to bring about. The D+PR interpretation accounts for this general understanding of the Socratic thesis by making what the agent does be something other than what he desires.[17]

IV. The B and B+FD Interpretations

In the B interpretation, there is no standing desire for the real good in the human psychology. Instead, because all motivation ultimately stems from belief, all goals ultimately have their basis in epistemic cognition.[18] In the B interpretation, some beliefs are motivating.[19]

[16] Penner and Rowe 2005, 8n14 attribute to Socrates what they describe as a 'Davidsonian' as opposed to a 'Goldmanian' criterion for the identity of actions. 'The identity of a given particular action is fixed by all the particular properties the action actually has, including the consequences that action has; it is not fixed by the particular descriptions under which the agent does it' (8).

[17] As a variation on the D+PR interpretation, one might let the standing desire for the real good be the agent's only desire. Such an agent would act once he forms the belief that some course of action is the real good in the circumstances. If this belief is false, then what the agent does is not something he desires. This variation on the D+PR interpretation is not identical with the D+PR interpretation (see n14). It seems, however, to be something that Rowe may have contemplated.

[18] This aspect of the B interpretation, although perhaps unusual, does appear to have support among contemporary analytic philosophers. 'In a rational agent, there must also be a purely ratiocinative basis for desire formation. The sole ratiocinative basis for desiring something should be the belief that it is a suitable goal' (Pollock 1995, 270). Cf. Frede 2011, 21: 'Socrates, Plato, Aristotle, the Stoics, and their later followers…all agree that reason, just as it is attracted by truth, is also attracted by, and attached to, the good and tries to attain it.'

[19] In terms of the metaphor of 'direction of fit', some psychological states are 'besires'. They carry both a mind-to-world and a world-to-mind direction of fit. One might understand the claim in

The psychology in the B interpretation is intellectualist: in the causal history of every action, there is a belief about the real good. The only way to generate a specific motivation, and hence an action, is in terms of a belief about the real good. The belief may be false. In this case, the agent does not move toward the real good. Rather, because his belief is false, he moves toward the merely apparent good. Nevertheless, there is a belief about the real good in the causal history of every action. The D and B interpretations are both intellectualist, but each secures intellectualism through a different cognitive architecture and design in the psychology.

This is worth considering in more detail because there can be a temptation to think that the B interpretation relies on an incoherent design for practical cognition. The aim of practical cognition is to make the agent's circumstances good, and if the states and processes that constitute the cognition do not tend to bring about this end, then it is unclear whether the states and processes really are an architecture, or design, for practical cognition. In the B interpretation, if proposing suitable goals in epistemic cognition is ongoing, then practical cognition achieves its aim. A belief that something is a suitable goal results in a plan and intention to carry out a given course of action. The outcome of the course of action provides evidence about the suitability of the goal, and this evidence feeds into the ongoing process of proposing suitable goals. But one might wonder whether the process of proposing suitable goals must be ongoing. The D interpretation has the desire for the real good. It is fixed in the psychology. If nothing similar exists in the psychology in the B interpretation to guarantee that the process of proposing suitable goals is ongoing, then it would seem that the B interpretation is not a coherent design for practical cognition.

In fact, there is something similar: in the B interpretation, the process of proposing suitable goals is itself a fixed part of the psychology. The agent, as part of an on-going process, forms beliefs about what the real good is in the circumstances. There is no antecedent desire that sets this process in motion. This process is a fixed or structural part of the psychology. There must be some structural parts in every psychology. The D interpretation posits the desire for the real good as a structural part, and it explains the ongoing epistemic process of proposing suitable goals in terms of this desire. In the D interpretation, the epistemic process of forming beliefs about what the real good is in the circumstances is grounded in the antecedent desire for the real good. This is a standing desire in the psychology, and it causes the agent to form beliefs about what the real good is in the circumstances. The B interpretation does not have any standing desires. Instead, it fixes the epistemic process of forming beliefs about the real good as a standing or fixed part of the psychology. So the psychology in the B interpretation does appear coherent.

the B interpretation that some beliefs are motivating to be the claim that some beliefs have the directions of fit that define besires. Altham 1986 seems to have coined the term 'besire'. For some recent discussion of besires in connection with contemporary philosophical problems in the analytic tradition, see Zangwill 2008.

The primary objection to this coherence would seem to be based on the broadly Humean conception of reason as the slave of the passions. Under the influence of this idea, one might argue that only the general reasoning process of belief formation and retraction can be built into the cognitive architecture of a rational agent, not any specific process to solve a particular problem. This is a powerful philosophical consideration,[20] but obviously it has much less weight in a historical investigation. The Humean conception of reason would seem to be modern in origin and thus, in the absence of evidence, should not be read into the ancients. No one has provided any such evidence.. So the B interpretation should not be dismissed out of hand. In the B interpretation, there is no fixed desire for the real good. Instead, the process of forming and retracting beliefs about the real good is itself a fixed part of the human psychology.

Indeed, in a certain way, the psychologies in the D and the B interpretations are very similar. The fixed desire for the real good is the starting-point for action in the D interpretation. This desire triggers the epistemic process of forming beliefs about what the real good is in the circumstances. These beliefs trigger instrumental desires. These desires trigger planning. In the B interpretation, the starting-point is an epistemic process. As a structural feature of the psychology, the agent forms beliefs about what the real good is in the circumstances.

Further, if desires exist as functional states, there is a subclass of the B interpretation, the B+FD interpretation, that even more closely resembles the D interpretation (cf. Lorenz 2006, 28). If desires are states that function in a certain way in the psychology, then by forming and retracting beliefs about what the real good is, a human being is forming and retracting desires for various states of affairs. But the B+FD interpretation is still a B interpretation. In the B+FD interpretation, there is no standing desire for the real good. Further, all desires are identical to beliefs. In forming a belief that something is a suitable goal, a human being forms a desire, but there is no psychological state other than a belief that something is a suitable goal that functions as a motivational state. In the B and B+FD interpretations, all motivation in human beings ultimately stems from beliefs of a certain sort. In the D and D+PR interpretations, all motivation ultimately stems from the desire for the real good. This desire does not stem from and is not identical with any belief.[21]

[20] Empirical work has cast doubt on this assumption (see, e.g., Cosmides 1985 and Cosmides 1989).

[21] Note that the D interpretation cannot be supplemented with a functional analysis of desire. First of all, the point of the D interpretation is to insist that there really are desires in the human psychology, that these desires are not beliefs, and that in human beings all motivation ultimately has its source in the desire for the real good. Moreover, no belief can do the job. The only candidates are beliefs that something in particular is the real good, and the claim in the D interpretation is not that there is something in particular such that all human beings desire it as the real good. Note also that the B+FD interpretation cannot be supplemented with a fixed belief that there is something in particular that is the real good. It is intrinsic to the B interpretations that the process of belief formation and retraction about the real good is a basic part of epistemic cognition. The essential idea is that, contrary to the broadly Human conception of reason, epistemic cognition in human beings is not limited to the

Is either the B interpretation or the B+FD interpretation consistent with the Socratic thesis that 'no one errs willingly'? In the D+PR interpretation, when someone goes wrong, he does not go wrong 'willingly' because he does not desire to do what he in fact does do.[22] This way of understanding the Socratic thesis is also available to the B+FD interpretation. To account for the Socratic thesis, the Penner-Rowe theories of action and instrumental desire do the work. And it is clear that the B+FD interpretation may be modified similarly. Suppose that the agent forms the belief that *g* is the real good. Given the functional analysis, because this belief is motivating, the agent has a desire in virtue of having this belief. What is this desire? Given the Penner-Rowe theories of action and desire, it is 'the desire to do this action here and now which is *both* the really best means to the agent's maximal happiness (maximal good) *and* the actual action done which the agent thinks to be the best means available'. So, it should be evident that both the D interpretation and the B interpretation can be supplemented so that they are consistent with the Socratic thesis that 'no one errs willingly', as Penner and Rowe understand the content of this thesis.[23]

This is important. Given that the D and the B interpretation can both be so supplemented, it follows that the Socratic thesis that 'no one errs willingly' is really a secondary issue in the investigation. The immediate question is whether the evidence of the *Protagoras* decides between the D and B interpretations. The D and B interpretations are the basic forms of the competing interpretations of Socratic intellectualism. The complicating factor of the Socratic thesis that 'no one errs willingly', and whether the Penner-Rowe modification provides the best way to understand this thesis, may be set aside. The D and B interpretations can both be made more specific so that an agent does something other than what he desires when he goes wrong, if this turns out to be the best way to understand the Socratic thesis that 'no one errs willingly'.

V. Textual Evidence

The *Protagoras* occupies a unique place in the history of philosophical thought

general process of forming and retracting beliefs in response to evidence. A fixed belief that something in particular is the real good eliminates the need for anything more than the general process and so is inconsistent with the B interpretations.

[22] One might think it is better to say that the agent desires to do what he does, and so does it willingly, but does not go wrong willing because it does not follow that he desires to do what he does under the description of going wrong. This alternative to Penner and Rowe's analysis depends on difficult issues involving referential opacity in propositional attitude contexts. See Penner 2011 for some discussion of these issues in connection with Socrates and Plato.

[23] The B interpretation may also be supplemented with desires in another way. Instead of having desires exist functionally, it is possible to have them arise in the psychology to encode goal adoption. In this version of the B interpretation, like all versions of the B interpretation, there is no standing desire for the real good. The epistemic process of forming and retracting beliefs about the real good is a fixed, structural part of the psychology. When the agent settles on a belief about what the real good is in the circumstances, a desire for what he believes is the real good arises in the psychology. This desire is not identical to any belief, but it is strictly dependent on the antecedent belief about what the real good is in the circumstances. So belief 'rules' in this version of the B interpretation.

about cognition, reason, and motivation in human beings. Historians have thought that the discussion with Protagoras in some way reflects the views of the historical Socrates on reason and motivation in human beings. Moreover, the striking image of reason as a 'slave' (352c1) enters the history of philosophy in this passage. The character Socrates seems to reject the conception of reason implicit in this image, and subsequent philosophers typically took sides either for or against Socrates on this issue. For example, in the reaction to the classical tradition of Plato and Aristotle that characterized Hellenistic philosophy, the Stoics seemed to have looked to the *Protagoras* to develop and defend what they understood as the view of the historical Socrates against the innovations Plato introduced in the *Republic* in his Tripartite Theory of the Soul.[24]

In the *Protagoras*, two Socratic theses frame the discussion of whether reason is a ruler. At the outset, in questioning Hippocrates about what he hopes to become by going to Protagoras, Socrates tells Hippocrates that he is a psychological being (313a). Second, Socrates tells Hippocrates that the health of his soul depends on 'teachings' or 'doctrines' (μαθήμασιν, 313c7). A human being controls himself and his life by exerting control over his soul, and a human being exerts control over his soul by exerting control over his beliefs. These theses explain why Socrates is so keen for Hippocrates to understand the import of his decision to seek a sophistical education (313a-314b). At stake is the health of his soul and thus his well-being.

It is against this background that Socrates considers alternative ways that human cognition might work. The first he associates with popular opinion: that in human beings 'knowledge' is not a 'ruler' and that often when knowledge is present what rules is something else, 'sometimes desire, sometimes pleasure, sometimes pain, at other times love, often fear' (352b1-8). The second possibility is the one he himself seems to accept. He says that if someone were to know 'what is good and bad' (352c5), he would not be overcome and hence would act as his knowledge dictates. And subsequently it becomes clear that there is nothing special about the motivating power of knowledge as opposed to mere belief. Socrates says that 'no one who knows or believes there is something better than what he is doing, something possible, will go on doing what he had been doing when he could be doing what is better' (358b7-c1).

The psychology and cognitive architecture Socrates associates with popular opinion is not easy to reconstruct with any certainty, since his description is extremely brief, but the following is a natural possibility. In human beings, according to popular opinion, there is automatic goal proposal and default acceptance. When someone is hungry, he gets the desire to eat. This desire leads to action if the opportunity arises. In addition to goal proposal and default accep-

[24] The philosophical outlook that unites the Hellenistic philosophers is their critical attitude toward what they regarded as the excesses of the prior classical tradition of Plato and Aristotle. On the question of the soul, the Stoics seem to have thought that Plato and Aristotle went wrong in their departure from the view Socrates seems to have held. For a clear statement of the Stoic reversion to Socratic intellectualism, see Cicero's *Academica* i 39.

tance, there is an overriding mechanism to stop desires from issuing in action. When someone believes that something better is possible, the default acceptance of the proposed goal of eating can be overridden and thus the desire to eat can be eliminated. This overriding mechanism, however, does not always work properly. Sometimes the belief that there is something better fails to eliminate the desire. A compulsive eater provides an example. He may believe or even know that there is a better option but have the desire to eat nonetheless. He may even act on the basis of this desire. This would not be rational. The belief should dispel the desire, but popular opinion supposes that the desire is not always dispelled. Knowledge is not always a 'ruler' and a 'leader' in the human psychology and cognitive architecture. It can be dragged around as a slave.

Socrates rejects this psychology as a description of 'human nature' (ἀνθρώπου φύσει, 358d1), but his rejection alone does not uniquely determine an alternative and hence does not decide between the D and B interpretations. Belief 'rules' in both interpretations, since both are intellectualist. Belief 'rules' in the B and the B+FD interpretations, since belief is the source of all motivation. In the D and the D+PR interpretation, the standing desire for the real good is causally prior to belief. So belief does not 'rule' by being first, but the agent nevertheless always acts in terms of his belief. Hence, Socrates' rejection of the psychology he associates with popular opinion does not decide between the possible interpretations of his intellectualism.

But Socrates' argument against popular opinion is much more telling. The structure of the argument, although not completely clear, seems to take the form of an inference to the best explanation. The phenomenon to be explained is the experience of 'being overcome by pleasure' (352e6-353a1). To make the case against popular opinion, Socrates shows that the experience of being overcome is not best explained in terms of the conception of human cognition in which knowledge is not a ruler but can be dragged around. To show this, Socrates argues that the explanation popular opinion provides is 'ridiculous' (355d1). If popular opinion were correct, then, given the premise that pleasure is the good,[25] the experience of being overcome by pleasure would be one in which a human

[25] It is controversial whether the premise is *ad hominem* or whether it is also something the character Socrates believes is true. (For some discussion, and a map of some of the literature, see Russell 2005, 239-248.) Given the dialectical and elenctic character of the question-and-answer method, it follows that the premise is *ad hominem*. Socrates, however, might also believe that the premise is true. It is a premise in what seems to be his only argument for the conclusion that reason rules. If he does belief this premise, it is necessary to know what he believes. And the crucial evidence is at 358a5-b2, where Socrates emphasizes that when he asks whether the pleasant is good, he is asking about something one might call 'pleasant' (ἡδύ), 'delightful' (τερπνόν), or 'enjoyable' (χαρτόν). This strongly suggests that the premise is a way to express the natural idea that '*S* is pleased that *P*' and '*S* is happy that *P*' are two ways to say the same thing. The aim of practical cognition is to make the circumstances 'good' for the agent, to make the agent 'pleased' with the circumstances, and to make the agent 'happy' with the circumstances. This deflationary reading is all that is required for the argument against popular opinion. And given this much, the premise is relatively uncontroversial and something Socrates could easily believe. It is not the proposition that the good is sensory pleasure.

being 'does what is bad, knowing that it is bad, it not being necessary to do it, having been overcome by the good' (355d1-3). According to Socrates, it is more plausible to explain the experience of being overcome by pleasure in terms of the psychological state of 'ignorance' (ἀμαθία, 357d1). And all parties to the argument subsequently agree that ignorance is a matter of 'having a false belief and being deceived about matters of importance' (358c4-5, τὸ ψευδῆ ἔχειν δόξαν καὶ ἐψεῦσθαι περὶ τῶν πραγμάτων τῶν πολλοῦ ἀξίων).

The soundness of Socrates' argument is obviously uncertain (see Wolfsdorf 2006b), and as usual Socrates can be understood to argue dialectically, but the crucial point for deciding between the D and B interpretations is clear: Socrates locates the motivation in being overcome in a false belief. This is straightforward evidence for one of the B interpretations. Popular opinion assumes that there is a source of motivation in human beings other than beliefs, but Socrates argues that popular opinion is wrong about all the examples it cites. These are examples where knowledge appears as a slave and seems to be ruled and dragged around by other things, 'sometimes desire, sometimes pleasure, sometimes pain, at other times love, often fear' (352b7-8). Hence, given that Socrates is being sincere, one naturally understands him to believe that knowledge rules because there is *no* source of motivation in human beings other than belief. In particular, there is absolutely nothing in his remarks to suggest that he thinks that all motivation ultimately stems from a standing desire for the real good and that this motivation gets misdirected by false beliefs about the good.

There is more evidence for the B interpretations in what Frede 1992, xxix has described as a 'clue' to why Socrates thinks that intellectualism is true. In 358d6-7, Socrates characterizes fear as a belief of a certain kind: he says that 'it is an expectation of something bad'. He does not just say that fear is always accompanied by this expectation. He says that it *is* this expectation. And fear is one of the things that popular opinion says can 'rule' a human being. If fear is a belief, and if the other things Socrates mentions on behalf of popular opinion are also beliefs, then it is obvious why popular opinion is wrong when it says that in human beings belief is sometimes powerless in the face of fear and other such things. The motivation in the experience of being overcome is a belief. Contrary to popular opinion, there are not two kinds of thing that are in competition for 'ruling' and 'leading' in a human psychology, desires and beliefs. There are only beliefs.

The B interpretations are thus a more natural fit for the *Protagoras* than the D interpretations. The leading idea in the 'love of wisdom' (φιλοσοφία) in the traditionally early dialogues is that a human being controls his soul, and hence the direction his life takes, by exerting control over what he believes. In the *Protagoras*, Socrates no doubt has this idea in mind when he asks Protagoras whether 'knowledge is a fine thing capable of ruling a person, and if someone were to know what is good and bad…intelligence (φρόνησιν) would be sufficient to save a person' (352c3-7). Knowledge and intelligence are sufficient because 'being overcome' is having a false belief. The analysis of fear strongly suggests

that Socrates thinks that there are no motivational states that are not beliefs. As a logical possibility, he could think that knowledge is a ruler and a leader because all action is a function of *both* beliefs about the real good *and* a fixed desire for the real good. But there is in fact no hint of this view in the *Protagoras*.

One might argue that the hint comes from other dialogues, such as the *Gorgias* and the *Meno*, where some have said that the character endorses a D interpretation,[26] but this argument will require some very questionable premises. The first is obviously that Socrates endorses a D interpretation in these dialogues. But even if this were granted for the sake of argument, there would still be no reason to believe that Socrates has a D interpretation in mind in the *Protagoras*. For this to follow, there would have to be reason to believe both that the historical Socrates had a consistent, detailed theory of the soul and that Plato intended to use the character Socrates in all three dialogues to express this theory. And clearly this cannot be established independently of the evidence in the dialogues themselves. Hence, because the discussion in the *Protagoras* is evidence for the B interpretations, not the D interpretations, it follows that if the character endorses a D interpretation elsewhere in the traditionally early period, then there is reason to believe that the historical figure did not have a consistent and detailed theory of the soul. It would not follow that Socrates had inconsistent beliefs about the soul. He might have committed himself only to intellectualism. It would then be left to Plato to work out the details. And given the complexity of the issue, it would not be surprising if he were unsure about how this should be done.

Alternatively, one might argue that the *Protagoras* is neutral between the D and B interpretations. The argument, in this case, would be that the discussion is focused narrowly, that the only concern is to establish intellectualism, and that in establishing intellectualism Socrates expresses no view about the particular cognitive mechanism that underwrites his intellectualism. For the mechanism, according to the argument, one must look to traditionally subsequent dialogues, such as the *Gorgias* and the *Meno*, where he endorses a D interpretation.

This argument will also require some questionable premises. The first, again, is that Socrates has a D interpretation in mind in the *Gorgias* and the *Meno*. But if even this were granted, it would remain clear that the *Protagoras* is evidence for the B interpretations, not the D interpretations. Socrates asks Protagoras whether he agrees with him that 'intelligence would be sufficient to save a person' (352c6-7). Contrary to the popular opinion that knowledge can be dragged around, Socrates locates the motivation in the experience of 'being overcome by pleasure' in a false belief. He says that 'to control oneself is nothing other than wisdom' (358c3). With respect to the question of whether there is something Prodicus calls dread or fear (d5), Socrates says that it is identical to a belief whatever one calls it. The whole tenor of the discussion in the *Protagoras* is that

[26] See Penner 1991, Penner and Rowe 1994, and Penner and Rowe 2005. For a detailed and strongly negative assessment of some of the argument Penner and Rowe present, see Wolfsdorf 2006a.

knowledge and wisdom are all important for the good life because in human beings action is always a matter of belief. Contrary to the D interpretations, there is simply no indication in the *Protagoras* that belief is important because it focuses a fixed desire for the real good. It is just not there. The textual evidence for Socratic intellectualism in the *Protagoras* favors the B interpretations over the D interpretations.

VI. Conclusion

Socratic intellectualism may be false, but there is nothing uncharitable about the B interpretations of Socratic intellectualism relative to the D interpretations. Hence, prior to the textual evidence, there is no reason to think that either is more likely than the other. And on a level playing field, when the D and B interpretations are part of an interpretation of the historical figure, the B interpretations emerge as the best interpretations of Socratic intellectualism in the *Protagoras*. In the discussion with Protagoras about whether knowledge is a ruler, Socrates seems to think that intellectualism is true because human beings are psychological beings in which all motivation ultimately stems from beliefs of a certain sort. He gives no indication that there are any desires that do not stem from beliefs. In particular, he gives no indication that in every human being there is a standing desire for the real good. If Socrates endorses a D interpretation in other dialogues that traditionally are thought to predate the *Republic*, something that may or may not be true, then this would be a reason to believe that the historical Socrates committed himself only to intellectualism, not to a particular cognitive mechanism to underwrite his intellectualism. It would be a reason to believe that Plato explored different ways to work out the details in his attempt to understand Socratic intellectualism and the Socratic claim that human beings are psychological beings. If Plato did explore different ways to work out the details, it would not be too surprising. Socrates' doctrines were puzzling, and it is widely thought that Plato in the *Republic* rejects Socratic intellectualism for the Tripartite Theory of the Soul.[27]

Arizona State University
School of Historical, Philosophical and Religious Studies
Philosophy Faculty
Tempe AZ 85287-3902

BIBLIOGRAPHY

Altham, J.E.J. 1986. 'The Legacy of Emotivism' 275-288 in G. MacDonald and C. Wright edd. *Fact*

[27] Terry Penner's talk on Socratic intellectualism at Arizona State University in the late 1990's and our subsequent discussion during a hike in the Superstition Wilderness Area helped me better understand many of the issues I have discussed. Since then I have twice conducted seminars on various aspects of reason and experience in the ancients. My students in these seminars helped me work out my views more clearly. In addition, I received helpful comments on a draft from the editor of *Ancient Philosophy* and from an anonymous reader for the journal. These comments were among the best I have received on a journal submission.

Science and Morality. Oxford: Blackwell.
Bratman, Michael E. 1987. *Intention, Plans, and Practical Reason*. Cambridge: MIT Press.
Bratman, Michael E., David J. Israel, and Martha E. Pollack. 1988. 'Plans and Resource-Bounded Practical Reasoning' *Computational Intelligence* 4: 349-355.
Bremmer, Jan. 1983. *The Early Greek Concept of the Soul*. Princeton: Princeton University Press.
Burnet, John. 1916. 'The Socratic Doctrine of the Soul' Second Annual Philosophical Lecture. Read to the British Academy, January 26, 1916. Reprinted 126-162 in *Essays and Addresses*. Books for Libraries Press, 1968.
Claus, David B. 1981. *Toward the Soul. An Inquiry into the Meaning of* ψυχή *before Plato*. New Haven: Yale University Press.
Cosmides, Leda. 1985. *Deduction or Darwinian algorithms: an explanation of the 'elusive' content effect on the Wason selection task*. PhD Thesis. Harvard University.
Cosmides, Leda. 1989. 'The logic of social exchange: has natural selection shaped how human reason? Studies in the Wason selection task' *Cognition* 31: 187-276.
Frede, Michael. 1986. 'The Stoic Doctrine of the Affections of the Soul' 93-110 in M. Schofield and G. Striker edd. *The Norms of Nature*. Cambridge: Cambridge University Press.
Frede, Michael. 1992. 'Introduction' vii-xxxii in *Plato*. Protagoras. Translated with Notes, by Stanley Lombardo and Karen Bell. Indianapolis: Hackett.
Frede, Michael. 1996. 'Introduction' 1-28 in M. Frede and G. Striker edd. *Rationality in Greek Thought*. Oxford: Oxford University Press.
Frede, Michael. 2000. 'The Philosopher' 1-18 in J. Brunschwig and G.E.R. Lloyd edd. *Greek Thought. A Guide to Classical Knowledge*. Cambridge: Harvard University Press.
Frede, Michael. 2011. *A Free Will. Origins of the Notion in Ancient Thought*. Berkeley: University of California Press.
Furley, David. 1956. 'The Early History of the Concept of the Soul' *Institute of Classical Studies* 3: 1-18.
Handley, E.W. 1956. 'Words for "Soul," "Heart" and "Mind" in Aristophanes' *Rheinisches Museum* 19: 205-255.
Havelock, Eric A. 1972. 'The Socratic Self as It Is Parodied in Aristophanes' Clouds' *Yale Classical Studies* 22: 1-18.
Huffman, Carl. 2009. 'The Pythagorean conception of the soul from Pythagoras to Philolaus' 21-44 in Dorothea Frede and Reis Burkhard edd. *Body and Soul in Ancient Philosophy*. Berlin: Walter de Gruyter.
Kahn, Charles. 1996. *Plato and the Socratic Dialogue*. Cambridge: Cambridge University Press.
Kahn, Charles. 2002. 'Response to Christopher Rowe' *Journal of the International Plato Society*. 2: https://www3.nd.edu/~plato/plato2issue/kahn.htm.
Lorenz, Hendrik. 2006. *The Brute Within. Appetitive Desire in Plato and Aristotle*. Oxford: Oxford University Press.
Lorenz, Hendrik. 2009. 'Ancient Theories of the Soul' *Stanford Encyclopedia of Philosophy*. Ed Zalta ed.
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.
Penner, Terry. 2011. 'Socratic Ethics and the Socratic Psychology of Action: A Philosophical Framework' 260-292 in D.R. Morrison ed. *The Cambridge Companion to Socrates*. Cambridge: Cambridge University Press.
Penner, Terry and Christopher Rowe. 1994. 'The Desire for Good: Is the *Meno* Inconsistent with the *Gorgias*?' *Phronesis* 38: 1-25.
Penner, Terry and Christopher Rowe. 2005. *Plato's* Lysis. Cambridge: Cambridge University Press.
Pollock, John. 1995. *Cognitive Carpentry: A Blueprint for How to Build a Person*. Cambridge: MIT Press.
Reshotko, Naomi. 2006. *Socratic Virtue. Making the Best of the Neither-Good-nor-Bad*. Cambridge: Cambridge University Press.
Rowe, Christopher. 2002. '"Just How Socratic Are Plato's 'Socratic' Dialogues?" A response to

Charles Kahn, Plato and the Socratic Dialogue: The Philosophical use of Literary Form' *Journal of the International Plato Society*. 2: https://www3.nd.edu/~plato/plato2issue/rowe2.htm.

Rowe, Christopher. 2007. 'A Problem in the *Gorgias*' 19-40 in Christopher Bobonich and Pierre Destree edd. *Akrasia in Greek Philosophy*. Leiden: Brill.

Russell, Daniel. 2005. *Plato on Pleasure and the Good life*. Oxford: Oxford University Press.

Taylor, C.C.W. 2000. *Socrates. A Very Short Introduction*. Oxford: Oxford University Press.

Vlastos, Gregory. 1988. 'Socrates' *Proceedings of the British Academy* 74: 89-111.

Wolfsdorf, David. 2006. 'Desire for good in *Meno* 77B2-78B6' *Classical Quarterly* 56: 77-92.

Wolfsdorf, David. 2006. 'The Ridiculousness of Being Overcome by Pleasure: *Protagoras* 352b1–358d4' *Oxford Studies in Ancient Philosophy* 31: 113-136.

Zangwill, Nick. 2008. 'Besires and the Motivation Debate' *Theoria* 74: 50-59.