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](https://example.com)) and Sidgwick finds it ‘an extravagant paradox.’ 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

* I would like to thank David Blank, Kellie Brownel, Emily Fletcher, Aditi Iyer, Rachana Kamtekar, Sean Kelsey, Seana Shiffrin, and David Velleman for their comments and suggestions.

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

\(^2\) Velleman (2000), 99.

\(^3\) E.g. Penner (1991), and Penner and Rowe (1994).