Socratic Motivational Intellectualism[1]

FREYA MÖBUS

WHAT IS MOTIVATIONAL INTELLECTUALISM?

Socrates' view about human actions in Plato's early dialogues has often been called 'intellectualist' because, in his account, we always do what we *believe* is the best (most beneficial) thing we can do for ourselves, given all available options.[2]

The main textual evidence for Socrates' intellectualism about human actions comes from the *Protagoras*:

> No one who knows (*eidōs*) or believes (*oiomenos*) there is something else 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. . . . No one goes willingly toward the bad or what he believes (*oietai*) to be bad; neither is it in human nature, so it seems, to want (*ethelein*) to go toward what one believes (*oietai*) to be bad instead of to the good. And when one is forced to choose between one of two bad things, no one will choose the greater if he is able to choose the lesser.[3]
>
> – *Prt.* 358b–d

In the *Protagoras*, Socrates seems to claim that our actions always follow a belief about the best thing to do; that is, a belief about what is in one's own best interest at the time of action.[4] This idea, we could say, makes Socrates an intellectualist regarding *actions*. But one might wonder why Socrates is considered to be an intellectualist regarding *motivations* as the title of this chapter suggests. What exactly are 'motivations'? Unfortunately, interpreters of Socrates generally do not define this term. Without a more specific definition, 'motivations' can refer to anything that moves us to act – desires, goals, emotions, facts, beliefs, reasons and so on. Our discussion here will not hinge on one particular understanding of 'motivation'. Instead, we will understand Socratic motivational intellectualism as the view that for any given intentional action our motivations – whatever moves us to act – are tied to the intellect, specifically to a belief about the best (most beneficial) thing we can do for ourselves.[5] Without such a belief, no action follows. How exactly motivations are 'tied to' beliefs about what is best is, as we will see, one of the major points of disagreement among interpreters.

Motivational intellectualism is often considered to be at the centre of Socrates' intellectualist account of actions, according to which:[6]

(1) We never act against our present judgment about what is best to do. Since we *always* do what we believe is the best thing for us to do, Socrates' motivational intellectualism implies the denial of *synchronic belief akrasia* (*Prt.* 358b–d).
(2) All wrongdoing is due to ignorance. Since we always do what we believe is best for us, any action that is in fact not best for us results from a false belief. Moral wrongdoing or injustice is never in our own best interest (*Ap.* 30c–d, *Cr.* 49a–b, *Grg.* 469b). Thus, all moral wrongdoing is due to ignorance.

In addition to these two claims, some interpreters have added a further, more controversial claim:

(3) Our beliefs about what is best generate action-causing desires. Desires that are not generated by such a belief, that is, 'non-rational desires', cannot motivate actions.[7]

Since motivational intellectualism – the idea that we always act according to our beliefs about what is best – can explain these Socratic claims, many interpreters place it at the core of Socratic philosophy.[8] Some have even argued that it is this theory of human motivation – and not, for example, Plato's theory of forms – that distinguishes Socrates' philosophy from Plato's, and that suggests a division of Plato's dialogues into different groups: 'intellectualist' (or pre-*Republic*) and 'non-intellectualist' or 'anti-intellectualist' (post-*Republic*) dialogues (Rowe 2007: 21–25).

In addition to the idea that 'we always do what we believe is best for us to do', Socrates' account of motivation includes another major idea, the desire thesis:

(4) We all desire the good (*Prt.* 358d, *Meno* 77d–e, *Euthd.* 278e, *Grg.* 466a–468e).

The desire thesis, in turn, is often considered to be the foundation for the claim that[9]

(5) No one does wrong willingly (*Prt.* 345e, 358d–e; *Grg.* 509e). When we do what is wrong (i.e., not in our own best interest), we act against what we want.

Despite their centrality for Socratic philosophy, interpreters disagree on the exact interpretation of these five claims and of Socrates' account of motivation. This chapter surveys the interpretative landscape so that the reader may more easily navigate the primary and secondary literature and decide for themselves whether Socrates is a motivational intellectualist and, if so, in what sense.

The answer to this question matters not only for those who are academically invested in Socratic philosophy, but also for all who have joined Socrates on his mission of living a good life. For Socrates, we live well if we act well; wrongdoing leads to a life lived poorly and miserably. Understanding why we act as we do, promoting right-doing, and preventing wrong-doing, are thus of utmost importance for anyone who wants to live well.

Further, one's stance on Socrates' motivational intellectualism has broader implications for understanding his philosophy of human nature. Do we always do what we believe is in our own best interest? Is it entirely up to us to figure out what is in our own best interest, or do we come equipped with some motivational orientation toward what is truly good for us? In other words, do we all come equipped with an inherent desire for our real good? We will start our discussion with this question.

WE ALL DESIRE THE GOOD

When discussing Socrates' theory of motivation and action, interpreters often refer to the 'desire thesis' (Barney 2010): We all desire the good. In this section, we will disentangle the different ways of understanding this thesis.

The claim that we all desire the good is initially puzzling. Many would rightly wonder: 'What do you mean by *the good*?' Katja Vogt (2017: 16–20) helpfully explains that we can take 'the good' to refer to at least three different kinds of things:

(a) Long-term goods: Some might think that 'the good' refers to the good life (e.g., 'the good is the examined life') or happiness. Long-term goods are those that we try to achieve over the course of our life.
(b) Mid-range goods: Others might think of 'the good' in terms of good things to have. Socrates might say that 'wisdom is the good', while Callicles might object that 'pleasure is the good'. Mid-range goods are those intermediate goods that we want in our lives (e.g., health, wealth, friends, wisdom).
(c) Small-scale goods: Still others might think that 'the good' refers to the good we aim for in individual actions. For example, 'eating these vegetables is the good (thing for me to do)'. Small-scale goods are those that we consider the best thing for us to do in any given situation (e.g., to eat these vegetables).

These different goods can be in a means-end relation to one another. For example, I want to eat these vegetables (small-scale good) in order to stay healthy (mid-range good), and I want to stay healthy (mid-range good) in order to live a good life (long-term good). Alternatively, one might think that, for example, being healthy, having friends and being financially well-off are specifications of what a good life looks like.

In Socrates' desire thesis, then, does 'the good' refer to small-scale, mid-range, or long-term goods? Interpreters of Socrates commonly do not make these distinctions and clarify what kind of good they have in mind. To better understand Socrates' desire thesis, I propose that we apply Vogt's concepts to the following passages: 'no one wants to go toward what one believes to be bad' (*Prt.* 358d); 'we desire the good' (*Grg.* 468b); and 'we all wish to do well' (*Euthd.* 278e).

We opened this chapter with a passage from the *Protagoras* in which Socrates says that in any given situation, we always want to do what we believe is the best (or the 'good') thing for us to do (*Prt.* 358d). If we put this into Vogt's terms, motivational intellectualism – as formulated in the *Protagoras* – refers to our desire for small-scale goods. In this reading, the desire thesis is a reformulation of Socrates' motivational intellectualism.

In the *Gorgias*, Socrates discusses small-scale goods in relation to mid-range goods: We sit or walk or embark on sea voyages for the sake of some further good that we hope to accomplish:[10]

> It's because we pursue the good (*to agathon*) that we walk whenever we walk; we suppose that it's better to walk. And conversely, whenever we stand still, we stand still for the sake of the same thing, the good (*tou agathou*). . . . Hence, it is for the sake of the good (*tou agathou*) that those who do all these things do them.
>
> – *Grg.* 468b

Here, 'the good' refers to mid-range goods: We do everything we do for the sake of some further good such as wisdom, health or wealth (*Grg.* 467e).

Turning to the *Euthydemus*, we see Socrates holds that we all desire the same long-term good, namely happiness or 'doing well':

> Do all men wish (*boulometha*) to do well (*eu prattein*)? Or is this question one of the ridiculous ones I was afraid of just now? I suppose it is stupid even to raise such a question, since there could hardly be a man who would not wish to do well.
>
> – *Euthd.* 278e

Later in the text, Socrates states that 'we all want to be happy (*eudaimones*)' (*Euthd.* 282a). We 'do well' and become happy, Socrates explains, by acquiring 'many good things' and using them well (*Euthd.* 278e–281b).

Given the textual evidence presented above, Socrates' desire thesis – we all desire the good – can be understood in three different ways: We all desire happiness or 'doing well' (long-term good); we all desire intermediate goods like health and friends that we want in our lives (mid-range goods); we all desire the good thing we aim for in individual actions (small-scale goods). For Socrates, that we desire 'the good' is a fact about 'human nature' (*Prt.* 358d) that we do not need to argue for or justify (*Euthd.* 278e). No one wants to be miserable, and no one wants what is bad for oneself.

Of course, I may incorrectly conclude that certain things are good when in fact they are not good or are perhaps even bad for me. But as Socrates explains in *Meno* 77c–e, we must be careful not to conclude that in such instances I desire something bad. I still desire what I believe is good for me:

> Those who do not know things to be bad do not desire what is bad, but they desire those things that they believe (*ōonto*) to be good but that are in fact bad. It follows that those who have no knowledge of these things and believe (*oiomenoi*) them to be good clearly desire good things.
>
> – *M.* 77d–e

The *Meno* echoes Socrates' claim in *Protagoras* 358d that no one wants what they believe is bad for themselves. In the *Meno*, Socrates explains further that no one wants bad things because they make us miserable and unhappy. And no one wants to be miserable and unhappy (*M.* 77e–78a). The *Meno* is often considered to be the locus classicus for the so-called *apparentist* interpretation of Socrates' claim that we all desire 'the good': We all desire *apparently* good things, that is, things that we *think* are good even when they are in fact bad for us.[11]

In *Gorgias* 466a–468e, Socrates considers agents who go terribly wrong in determining what is best for them to do: tyrants. When explaining tyrants' actions – unjustly confiscating someone's property, exiling them, and putting them in prison or even executing them – Socrates eventually convinces his interlocutor, Polus, that while these agents do what they believe is best, they do not do what they *want*:

> Socrates: If a person who is a tyrant or an orator puts somebody to death or exiles him or confiscates his property because he believes (*oiomai*) that doing so is better for himself when actually it's worse, this person, I take it, is doing what seems (*dokei*) good to him, isn't he?
> Polus: Yes.
> Socrates: And is he also doing what he wants, if these things are actually bad? Why don't you answer?
> Polus: All right, I don't think he's doing what he wants. (*Grg.* 468d)

Socrates even goes on to claim that 'no one does what is unjust (*adikein*) because he wants to, but that all who do so do it unwillingly (*akontas*)' (*Grg.* 509e). Those who act unjustly do not do what they want, because in doing what is unjust, they harm themselves. For Socrates, 'doing what is unjust is the worst thing there is' for a person (*Grg.* 469b) because injustice harms one's soul and makes one miserable (*Grg.* 470e).

Notice that in the *Gorgias* Socrates seems to make a claim about desire that is quite different from the *Meno*. In the *Meno*, it seems that we always want what we (subjectively) take to be good. In the *Gorgias*, Socrates seems to say that we all always want what is in fact (objectively) good for us. The *Meno* suggests the '**apparent good**' interpretation of desire: We want what we think is good. In this interpretation, tyrants do what they want. The *Gorgias*, in contrast, seems to suggest the '**real good**' interpretation: We only want what is *really* or *truly* good for us. And in this interpretation, the tyrants do not do what they want. How can we resolve this seeming inconsistency?

Before we dive into the secondary literature on this question, let me outline why our answer matters in the context of Socratic motivational intellectualism. If all our desires are for apparent goods – if indeed we desire only what we think is good – then what we desire depends entirely on our beliefs. Motivationally, we would be born as blank slates. But if some of our desires are for the real good, regardless of what we believe is good, this would indicate that we come equipped with an inherent orientation toward the real good. Human nature would, to some extent, push us in the right direction. The apparent and real good interpretations lead to two different accounts of Socratic motivational intellectualism, and the former is more intellectualist than the latter, or so I will argue. According to the apparent good interpretation, *all* of our desires derive from beliefs. But according to the real good interpretation, our desire for the real good is inborn and precedes any of our beliefs. The central question thus becomes: *Are we born with an inherent desire for the real good?*

Let us first look at interpretations that maintain Socratic psychology includes an inherent desire for the real good:

> We, humans, are hardwired to seek our own good. What we want is, ultimately, to do well for ourselves. The striving for this condition of doing well, which Socrates calls 'the good', is something that every human soul comes equipped with.
>
> – Segvic 2006: 180

According to this line of interpretation, we humans are born with an inherent desire for our real long-term good: true happiness, or truly doing well and living a good life. Terry Penner developed this line of thought in detail, sometimes in collaboration with Christopher Rowe. According to Penner and Rowe, in each situation, we must figure out which action would most promote our happiness, that is, which action would be the best thing for us to do. If all goes well, we conclude correctly that a particular action is best for us to do. This true belief then transforms our overarching, general desire for true happiness into a particular desire to execute a specific action – 'the generalized desire . . . *becomes* the executive desire' (Penner 1991: 153). The combination of our inherent desire for the real good and our belief that action *x* is the best thing to do brings about an action-causing desire to do *x* – and so we do *x*.

But agents often deliberate incorrectly and are thus mistaken about their long-term, mid-range or small-scale goods. The tyrant, for instance, might incorrectly believe that living well consists in accumulating money and power and that executing political opponents is the best means toward this end. In such cases, the agent's belief about what

is best for him to do brings about a 'defective' or 'false desire' (Penner and Rowe 2005: 221), which then leads to an action (e.g., executing political opponents). But since this action is not in fact the best means to the agent's overall real good, strictly speaking he does not do what he wants. In fact, according to Penner and Rowe, any action that does not *maximise* the agent's happiness is not truly a wanted action: one does what one wants to do only if one's action leads to 'the greatest amount of happiness attainable starting from where one is at right now' (Penner 1991: 182). Rowe sums up his interpretation as follows:

> Briefly, and at bottom, it [i.e., Socrates' intellectualist theory of motivation] 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 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 2007: 23

We *do* what we think is good, but we *desire* what is truly good. For Penner and Rowe, we want only our real long-term good – true happiness – and that which would most promote our true happiness; that is, truly good small-scale and mid-range goods. The desire for the real long-term good, in this interpretation, is our first and only desire: We are born 'hardwired' to pursue the real good, and all subsequent desires for mid-range and small-scale goods that arise out of deliberation are simply particularisations of our general desire for the real good. According to Penner and Rowe (1994), Socrates in *Meno* 77d–78b does not in fact claim that we desire apparent goods. Instead, they argue that the *Meno* supports the same idea as the *Gorgias*: In everything we do, we desire what is really good.

Rachana Kamtekar (2006) agrees with Penner and Rowe that we have an inherent desire for our own real long-term good; that is, for true happiness. She also agrees that the *Gorgias* suggests that, strictly speaking, we only 'want' actions if they truly promote our long-term good. But Kamtekar differs from Penner and Rowe in maintaining that we also desire our apparent good; that is, what we think is good, as the *Meno* seems to suggest. In other words, our desire for the real good is not our only desire. To harmonise the 'apparent good' with the 'real good' interpretation, Kamtekar distinguishes between two kinds of conative attitudes: *boulēsis*, or the desire for the real good, which Kamtekar translates as 'wanting', and *epithumia*, or the desire for our apparent good, which she translates as 'desire'.[12] We *want* (*boulesthai*) the real good 'rather than what we think good (*Grg.* 468b–d)', but we *desire* (*epithumein*) the apparent good, that is, 'things that we think are good, which are sometimes in fact bad (*M.* 77d–e)' (Kamtekar 2006: 127). The tyrant, for example, 'desires' to murder his enemies; but he does not 'want' it.

Our 'wanting' what is truly good for us is an inherent motivational orientation. Thus, Kamtekar explains, no agent is entirely conatively cut off from the truth (Kamtekar 2006: 150–151). This conative connection to our real good explains why getting what we desire does not always lead to lasting fulfilment. Our 'striving comes to rest' only once we achieve our real good, but when we acquire things that are not really good, but only apparently so, 'we will not be stably fulfilled by them' (Kamtekar 2006: 156). When we do something that does not in fact promote our real good, our 'wanting' what is truly good for us remains unfulfilled, and thus we may experience a lingering feeling of dissatisfaction.

In contrast to Penner, Rowe and Kamtekar, Rachel Barney (2010) proposes that the *Gorgias* does not introduce an inherent desire for our real good; instead, the *Gorgias* offers an addition to or clarification of Socrates' claim in the *Meno* that we desire the apparent good. In Barney's interpretation, 'I pursue what seems good *as an attempt* to obtain what really is so' (2010: 53). In other words, agents desire the apparent good, believing that it is the real good (even if it is not):

> As the Appearance thesis says, I always desire what seems good to me. But, as the Reality thesis clarifies, that does *not* mean that I desire objects *under the description* 'what seems good to me', taking my subjective responses to be constitutive of value. Rather, in desiring I do my best to track what is antecedently valuable, insofar as I can detect it. Properly understood, the Desire thesis is really a claim about the priority of cognition to motivation.
>
> – Barney 2010: 38

Barney here explicates the apparentist interpretation that we always desire what we think is good, which goes back to Gerasimos Santas (1964). In this interpretation, we do not have an inherent desire for the real good. Certain things are truly good for us, Barney explains, but we are not hardwired to desire them; we do not have a 'latent teleological orientation' toward them (2010: 37). Instead, *all* desires are causally dependent on evaluative beliefs; for Socrates, 'we cannot desire something without first finding it good' (2010: 43).

Thomas Blackson (2015) also argues that, for Socrates, all desires follow beliefs and that we thus do not have an inherent desire for the good. But whereas Barney proposes that all desires are *caused* by beliefs (2010: 43), Blackson *identifies* desires with beliefs:

> 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.
>
> – Blackson 2015: 31

In Blackson's interpretation, our beliefs about what is good do not bring about the separate psychological state of desire; instead, these beliefs are in themselves motivating once they are combined with the belief that their objects are indeed best to pursue. For example, my belief that 'eating chocolate is good' becomes a desire to eat chocolate once I have concluded that eating chocolate is the best thing for me to do right now. In making all desires identical to beliefs, this interpretation attributes the perhaps most radical form of motivational intellectualism to Socrates.

At this point, let us return to Socrates' claim in *Gorgias* 509e that 'no one does what is unjust (*adikein*) because he wants to, but that all who do so do it unwillingly (*akontas*)'. Socrates repeats this idea with slight variation in the *Protagoras*, saying that 'no one errs (*examartanein*) willingly (*hekonta*) or willingly does anything shameful (*aischra*) or bad (*kaka*)' (345e). These claims are often summarised as '**no one does wrong willingly**'.[13]

Socrates himself does not explain this claim any further, and interpreters are left wondering why and in what sense exactly wrongdoing is 'unwilling'. Especially apparentists are in need of an explanation. Apparentists argue that the tyrant believes executing his enemies is best and he thus *wants* to do so, since all desires are 'causally dependent on' (Barney) or 'identical to' (Blackson) beliefs. Thus, in the apparentist interpretation, the tyrant does what he wants, which seems to contradict Socrates' claim. Below, we will review two arguments for 'no one does wrong willingly': one that premises

motivational intellectualism and another that premises the desire thesis. I will propose that the latter is more convincing and open to both 'apparent good' and 'real good' interpreters.

Some argue that the tyrant acts 'unwillingly' because he acts from ignorance (see, e.g., Santas 1964: 160, n.25). We can reconstruct this argument as follows:

(i) We always do what we believe is the best (most beneficial) thing we can do for ourselves. [Motivational intellectualism]
(ii) Thus, if we do what is in fact bad (i.e., harmful) for us, we must falsely believe that what we are doing is good for us.
(iii) Actions done due to false beliefs (i.e., ignorance) are unwilling; if I act as I do because I made a mistake, then I do not really want to do what I am doing.[14]
(iv) So, no one does what is bad for oneself willingly.
(v) Doing injustice is bad for oneself.
(vi) So, no one does what is unjust willingly. (*Grg.* 509e)[15]

While *Protagoras* 358d lends some support to this argument by presenting 'no one does wrong willingly' and motivational intellectualism as being closely connected, Kamtekar (2017b: 71–72) argues that *Apology* 25c–26a raises doubts about understanding 'unwillingly' as 'acting from ignorance'. In *Apology* 25c–26a, Socrates says that if one does wrong willingly, it must be because one does not know that if one corrupts one's associates, they will harm one in turn. In other words, a certain kind of ignorance – not knowing that corrupting one's associates leads to harm to oneself – makes *willing* wrongdoing possible. The *Apology* thus suggests that not every action done due to ignorance is unwilling, contrary to premise (iii).

Others argue that the tyrant acts 'unwillingly' not because he acts from ignorance but because he does not do what he *really* wants (see e.g., Kamtekar 2017b: 69–128; Brickhouse and Smith 2018: 38 n. 2; 47). We can reconstruct this alternative argument as follows:

(i) We want to do what is truly good (truly happiness promoting) for ourselves. ['Real good' interpretation of the desire thesis]
(ii) 'Willingly' means 'acting in accordance with what one truly wants'; 'unwillingly' means 'acting against what one truly wants'.
(iii) So, no one does what is bad for oneself willingly.
(iv) Doing injustice is bad for oneself.
(v) So, no one does what is unjust willingly. (*Grg.* 509e)

'Real good' interpreters argue that we want to do what is truly good for ourselves because we have an inherent desire for our real good. When we pursue what is in fact bad for us, we act against this desire and – since this desire is part of our natural set-up – we act against 'human nature' (Kamtekar 2017a: 76). Apparentists, on the other hand, could argue that we want to do what is truly good (happiness promoting) for us not because of an inherent desire for the real good but because we believe that happiness is good. Since we believe that happiness is good, we desire happiness, and in desiring happiness, we are aiming at our true happiness or 'real good'; we are trying to get it right (Barney 2010: 53). When we do what is in fact bad for us, we act against what we truly want and thus unwillingly.

EMOTIONS, APPETITES AND NON-RATIONAL DESIRES

When identifying the motivations for our actions, we might think of emotions and appetites. Someone might explain, for example, 'I got a glass of water because I was thirsty' (appetite) or 'I walked away because I was angry' (emotion). But in Socrates' explanation of our actions, emotions and appetites do not seem to play an important role. According to Socratic motivational intellectualism, we act as we do because we *believe* it is best. The only conative element that seems to play a role in motivating our actions is our desire for the good. Thus, interpreters have traditionally paid little attention to emotions and appetites, sometimes even depicting 'Socrates as denying that people feel the urge to do things other than what they believe to be good for them'.[16] It seems to some that Socrates 'does away with' (Aristotle *MM* 1182a15–23) or at least totally 'disregard[s]' (Kahn 1996: 227) emotional or affective factors of human motivation. He seems to propose 'una erradicación total de factores irracionales que pueden interferir en el logro de una condición racional' (Fierro 2012: 60).

One might think that interpretations like these are supported by Socrates' apparent denial of ***akratic*** actions. According to Socratic motivational intellectualism, we *always* do what we believe is the best thing we can do. This entails that we never act against our belief that a certain action is best for us to do. But agents seem to act *akratically* all the time; that is, they seem to act against their better judgment because they are overcome by strong appetites or emotions. Imagine, for example, a pie eater who believes that eating pie is bad for him but eats the pie anyway. How could Socrates explain this action?

Socrates could describe this example differently: In the moment of action, when the agent reaches for a slice of pie, he believes that doing so is the best thing for him to do. Before reaching for the slice, he might have believed that eating pie would be bad for him; he may even return to this belief right after he takes a bite. But in the moment of action, Socrates could maintain, the agent believes that eating pie is the best thing for him to do. In other words, Socrates could deny that at the moment of eating, the agent believes that eating pie is bad for him.

We might still call this an instance of *akrasia* – acting against one's better judgement – but here it is understood diachronically instead of synchronically. With Penner (1990: 45–48), we can call the pie eater someone who exhibits 'diachronic belief *akrasia*':

- ✓ *Diachronic belief akrasia*: believing that *x* is best at t_1; believing that *y* is best at t_2; doing *y* at t_2; and then, at t_3, returning to one's belief that *x* is best.

What Socrates denies is what Penner calls 'synchronic belief *akrasia*', as well as 'knowledge *akrasia*':

- **X** *Synchronic belief akrasia*: believing that *x* is best at t_1, but doing *y* at t_1; in other words, doing *y* while continuing to believe that *x* is best (rejected at *Protagoras* 358c–d).
- **X** *Knowledge akrasia*: knowing that *x* is best and nevertheless doing *y* (rejected at *Protagoras* 352c, 356d–357a, 358b–c).[17]

Socrates' motivational intellectualism – we always do what we believe is the best thing we can do – implies that we never act against what we believe is best in the moment of action. But it leaves open the possibility of changing one's belief over time. Socrates could thus describe what happens to the pie eater as an instance of diachronic belief *akrasia* (see Brickhouse and Smith 2010: 199–210 for a very helpful discussion).

Notice that the Socratic explanation of *akratic* actions leaves room for the experience of mental conflict that we usually associate with acting against our better judgment. The agent might feel torn between eating and not eating pie. This experience could result from a conflict between appetites and desires: The agent's appetite for pie conflicts with his reasoned desire to abstain from eating (Singpurwalla 2006: 250–254). In this interpretation, the agent experiences two conative forces pulling him in different directions at the same time (i.e., synchronically). Alternatively, the agent might feel conflicted because he goes back and forth between two different beliefs about the best thing for him to do – eating or not eating pie; in this case, he alternates over time (i.e., diachronically) between two desires pulling him in different directions (Reshotko 2006: 87).

Agnes Callard (2014) has proposed a different explanation of the *akratic* experience. Callard argues that the *akratic* agent experiences a 'distinctive phenomenology of conflict and psychological strife' (2014: 36) because of his specific kind of ignorance. While all wrongdoers are ignorant – they all act on false beliefs about what is best for them to do – only the *akratic* wrongdoer has an experience (*pathos*) (*Prt.* 352e, 353a, 357c) of his ignorance; *akratic* ignorance is painful. The *akratic* agent's pain, Callard argues, is a symptom of this particular kind of ignorance:

> In akrasia, ignorance is felt as pain. Just as physical pain is the sensing of a bodily injury of which we are at times unaware, so too psychological pain can be the sensing of epistemic injury the person does not fully fathom. When he says that the akratic has an experience (*pathos* / *pathēma*) of his ignorance, Socrates is pointing to the fact that the akratic is the one whose ignorance does not *completely* escape his own notice.
>
> – Callard 2014: 36–37

The *akratic* agent himself might deny that he suffers from ignorance and experiences an 'epistemic injury'. After all, he claims to believe (or even to know) that eating more pie is in fact bad for him. He might explain that he only ate more pie because he was overcome by the pleasant appearance of the pie. But in Callard's interpretation, Socrates would diagnose this *akratic* person as suffering from belief–appearance confusion: In the moment of action, he did not actually *believe* that eating pie is bad for him (if so, he would not have eaten more pie); instead, it only *appeared* to him that eating pie is bad.[18] In fact, the *akratic* agent's appearance of the pie is correct (eating more pie is bad), but he acts on a false belief (eating more pie is good) while mistaking this false belief for a false appearance. *This* is the experience (*pathos*) of *akratic* ignorance.

When describing *akratic* actions as instances of ignorance (i.e., acting on a false belief about what is best to do), Socrates seems to pay no attention to the agent's emotions and appetites. This might seem to confirm some readers' suspicion that Socrates disregards our emotions and appetites. However, emotions like anger, shame and fear; appetites like erotic attraction; and states like pleasure and pain are clearly present in Plato's early dialogues and influence some characters' actions. The events of the *Euthyphro* start when a day-labourer kills a slave in drunkenness and anger (*Eu.* 4c). Anger also plays a central role in Socrates' trial in the *Apology*: The Athenians are angry with Socrates because he embarrassed them (*Ap.* 23c–d), and he appeals to them not to let their anger influence their vote (*Ap.* 34c–d). Then, after his conviction, Socrates explains that he lost his case because he lacked shamelessness and did not want to appeal to the judges' emotions by lamenting in tears and presenting his weeping wife and children (*Ap.* 38d). Shame and fear also dominate the discussion between Crito and Socrates about whether it is right for

Socrates to flee or to stay in prison: Crito suspects that Socrates refuses to flee because he fears that Crito and any others who help him will be punished (*Cr.* 44e–45b). He then tries to shame Socrates into fleeing (*Cr.* 45e–46a). In the *Laches*, Socrates discusses what it means to be courageous in the face of pain, pleasure, appetites and fear (*La.* 191d–e). As he explains in the *Gorgias*, some people do not withstand pain or fear: children, for instance, avoid medical treatment, and criminals try to avoid painful punishment (*Grg.* 479a–c). Socrates himself admits that he is afraid at times, specifically of thinking he knows something when he does not (*Chrm.* 166d) and of conducting investigations incorrectly (*Chrm.* 172e). Socrates also experiences erotic desire when facing the handsome young Charmides for the first time (*Chrm.* 155c–d).

In light of these passages, any interpretation of the Socratic psychology of action must include emotions and appetites and acknowledge that they play some role in generating our actions. But which role exactly is debated. Presently, there are three main actively defended interpretations: Penner's interpretation, which he sometimes defended in collaboration with Rowe, and which was further developed and defended by Naomi Reshotko; Rachel Singpurwalla's; and Thomas Brickhouse and Nicholas Smith's, which was originally inspired by Daniel Devereux. I will focus on the debate between Penner, Rowe and Reshotko on the one hand and Brickhouse and Smith on the other, which has dominated the secondary literature on Socrates' motivational intellectualism in recent years.[19]

The debate between these two accounts of Socratic motivational intellectualism is often presented as a debate over whether or not Socratic psychology includes a type of desire that is sometimes called '**non-rational**' or '**good-independent**'. But as we will see, framing the debate in these terms can be problematic because interpreters use the terms 'non-rational', 'good-independent' and 'desire' in different ways.[20]

The term '**non-rational**' is often used to refer to mental states that lack some kind of relation to beliefs or deliberation and reasoning: 'Non-rational' states are described by interpreters alternatively as not arising from, involving or responding to beliefs or deliberation.[21] Some interpreters take this further, arguing that non-rational states are not responsive to beliefs or reasoning *because* they are not or do not involve beliefs (i.e., if they were beliefs, they could be altered by belief changes).[22]

The term '**good-independent**' captures an alternative aspect of these types of mental states. 'Good-independent' mental states lack some kind of relation to goodness: They are understood as independent of our desire for the good or as independent of our beliefs or deliberations about goodness (either about goodness generally or about our overall, long-term good specifically).[23] But these alternative ways of defining 'good-independent' are not at all synonymous.

Finally, the term '**desire**' is itself used in different ways. Penner and Reshotko use 'desire' very narrowly, referring only to action-causing motivations – desires 'to do something', that is, desires that arise out of our inherent desire for the real good combined with beliefs about what is best to do (Penner 1992a: 128). They distinguish 'desires' from what they call 'itches' and 'hankerings' (Penner 1991: 201, n.45) or 'longings', 'drives', 'urges' and 'raw desires' (Reshotko 2006: 55, 76–77, 84–88). According to both Penner and Reshotko, emotions and appetites are not full-fledged desires; they are mere 'itches' or 'hankerings'. In the Socratic account of motivation, they argue, all 'desires' are particularisations of our inherent desire for the good. These particular (or 'executive') desires are brought about by the belief that a certain action is best. In this sense, all 'desires' are thus 'rational'; there are no non-rational executive desires within Socratic psychology (Penner 1990: 39, 1992a: 128).

Brickhouse and Smith, in contrast, use 'desire' very broadly to refer to feelings of attraction and aversion (Brickhouse and Smith 2010: 72, 2015: 14–15; see also Martinez and Smith 2018: 70). They argue that Socrates identifies three kinds of natural attractions and aversions in the *Charmides* (167e): 'appetite (*epithumia*), which aims at pleasure, wish (*boulēsis*), which aims at what is good, and love (*erōs*), which aims at what is beautiful. Each of these seems to have an aversive alternative, as well: we avoid pain, what is bad, and what is ugly' (2015: 14). In other words, we are naturally attracted to pleasure, goodness and beauty, and we feel aversive toward pain, badness and ugliness. These attractions and aversions are 'non-rational' in the sense that they 'seek their objects in a way that is independent of [i.e., not caused by][24] our reasoning or deliberation about what is really good for us' (2013b: 191). We desire pleasure, beauty and goodness because of inherent natural attractions to each.[25]

If we understand 'desire' broadly as 'attraction' and 'non-rational' as 'not caused by reasoning or deliberation', then Brickhouse and Smith's view is that we have three fundamentally 'non-rational desires' (our inherent desires for pleasure, goodness and beauty), whereas Penner and Reshotko's view is that we have only one such 'non-rational desire' (our inherent desire for the good). But if we understand 'desire' narrowly as 'desire to do something', then Brickhouse and Smith's view is that some of our 'desires' are rational while others are non-rational, and Penner and Reshotko's view is that all of our 'desires' are rational.

Against Brickhouse and Smith's very broad understanding of 'desires' as feelings of attraction and aversion, Singpurwalla has argued that feeling an attraction or aversion is different than 'actually desiring to act on that feeling' (2006: 252). For example, I might 'find a lifestyle of jet-setting and party-hopping attractive, but not really desire it, since I realize it is incompatible with fulfilling desires or goals that I believe make an essential contribution to living the good life' (2006: 252). Thus, Singpurwalla proposes that we distinguish full-fledged desires from attractions and aversions: Attractions and aversions are not themselves desires, but they can lead to desires if they are 'endorsed as true, and so give rise to a belief' (2006: 252). In Singpurwalla's interpretation, 'non-rational desires are evaluative beliefs' based on attractions and aversions (2006: 252).

Since the terms 'non-rational', 'good-independent' and 'desire' are used in very different ways, they invite misunderstanding. I will therefore centre this discussion around the phenomena that these terms were supposed to capture in the first place: emotions and appetites. What roles do emotions and appetites play in the Socratic account of action?

While Penner, Rowe and Reshotko consider emotions and appetites as belonging to the same class of mental states – namely 'itches' or 'hankerings' – and as influencing our actions in the same way, Brickhouse and Smith (2015) argue that we should distinguish between emotions and appetites. In the *Protagoras* (358d5–6), they argue, Socrates seems to endorse a cognitive account of emotions, claiming that emotions are beliefs. If Brickhouse and Smith are right that, for Socrates, emotions are beliefs but appetites are not, then these two states play different roles in generating actions and must be discussed separately. In this chapter, I will mainly focus on appetites. For a detailed discussion of emotions, see Chapter 12 of this volume ('Socrates on Emotion').

Interpreters on both sides of the debate – Brickhouse and Smith on the one side and Penner, Rowe and Reshotko on the other – agree that appetites can influence our actions by influencing our beliefs and deliberations.[26] For example, my craving for chocolate might affect my calculation of what is best; without this craving, I might not conclude that

buying a Snickers bar is the best thing for me to do right now (Reshotko 2006: 84–87). Both sides also agree that appetites are aversive or attractive; they drive or urge us.[27] Disgust, for instance, is aversive. However, they disagree about the extent to which appetites can drive or urge us.

The interpretive disagreement over the role of appetites in Socrates' account of motivation boils down to the following questions:

- Can appetites represent their objects as good or bad and thereby drive us toward or away from specific things?
- Can appetites incline us toward new beliefs about what is best to do?
- Can appetites distort already-formed beliefs about what is best to do?

Brickhouse and Smith answer these questions in the affirmative, while Penner, Rowe and Reshotko answer negatively.

Penner, Rowe and Reshotko have argued that appetites can influence our actions by *informing* our beliefs and deliberations:

> My thirst informs me that drinking something in the near future would be in my best interest, but my beliefs about what kinds of drinks are available, how they taste, how much they cost and how much effort it takes to obtain one of them will all be integrated into my executive desire to grab four quarters and to walk to the vending machine . . . to buy a bottle of grapefruit juice. . . . My non-rational urges do effect my behaviour, but they do not cause any behaviour all by themselves.
>
> – Reshotko 2013: 171

My thirst informs me that I should drink, but it does not tell me what I should drink (e.g., grapefruit juice or water) or how to get that drink (e.g., buying juice at the vending machine or getting tap water from the kitchen). Appetites like thirst do not tell us which things are good and worth pursuing or bad and worth avoiding. They cannot move us toward or away from specific things.[28] They thus cannot make us believe that any particular thing would be best to pursue. In other words, they cannot cause beliefs about what is best to do.[29]

According to Penner and Reshotko, my thirst is only one piece of information that I consider when I deliberate about what would be best for me to do. I may have any number of different appetites at any given moment – I might be thirsty, hungry and sleepy – and there may be many different ways of satisfying those appetites, but I 'always do just *one* particular action. How is it determined which one?' (Penner 1991: 202, n.45). Socratic motivational intellectualism provides a clear answer: We either drink, eat or sleep, depending on our belief about the best thing to do at the moment of action.

Once an agent has determined that doing *x* is best, he forms an (executive) desire to do *x*. While the agent might also feel an appetite, the satisfaction of which would require him to do *y*, this appetite ('non-rational desire') cannot trump his desire to do *x* ('rational desire').[30] However, the agent's appetite might change his calculation about what is best to do. For example, if he suddenly feels very hungry, he might determine that it is overall best to eat something first and then run errands. In other words, Penner, Rowe and Reshotko argue that appetites can inform us, but they cannot distort already-formed beliefs about what is best to do, nor can they compete with 'rational' desires (i.e., desires that arise from combining such beliefs with our inherent desire for the real good).

According to Brickhouse and Smith, on the other hand, appetites can distort already-formed beliefs about what is best. They argue that it is primarily this claim that distinguishes their interpretation from others:

> There is 'a fairly strong scholarly consensus that, according to the Socrates in Plato's Socratic dialogues, appetites, such as hunger and lust, and passions, such as anger and hatred, are not capable of altering an agent's judgment about what it is best for her to do. . . . The primary impetus behind *SMP* [Brickhouse and Smith's *Socratic Moral Psychology*] is to challenge *this particular consensus* and to provide an alternative account according to which Socrates recognizes the possibility that appetites and passions can, under certain conditions, not only affect, but even severely impair, judgment about what is best'.
>
> – Brickhouse and Smith 2012b: 325–326

According to Brickhouse and Smith, appetites can 'severely impair' beliefs about what is best because they present their objects as good and worth pursuing, or as bad and worth avoiding; this drives the agent toward or away from specific things and inclines him to believe that pursuing or avoiding those things is the best thing to do.[31] For example, my craving for chocolate might present a specific object (that chocolate bar in front of me) and a specific action (buying that chocolate bar in front of me) as good, which, by default, leads me to believe that buying the chocolate bar is the best thing for me to do, unless some other belief-forming process interferes (e.g., unless I consider some contrary evidence that convinces me that I should not satisfy this appetite).

In Brickhouse and Smith's interpretation, appetites can motivate actions under the condition that the agent believes satisfying his appetite is the best thing for him to do:

> Only if the soul ended up judging that what appeared to be good (because presented as such by an appetite, for example) was actually the best choice one could make in a given situation, would one become fully motivated to act. The stronger the appetite or passion, the more compelling the appearance of good would be. . . . But in this way, as we have shown, appetites and passions could never motivate us independently of what we believe is good for us, for although our beliefs about what is good for us might be unstable (particularly in those persons susceptible to strong appetites or passions), we will always act in the ways we presently believe are best for us.
>
> – Brickhouse and Smith 2010: 200; see also 2010: 62, 79, 107, 108; 2012b: 337; 2015: 16

In Brickhouse and Smith's account, therefore, Socrates is still a motivational intellectualist because our motivation for action is tied to a belief about what is best. Our *actions* always follow from such a belief: We *act* as we do because we believe it is best. However, Brickhouse and Smith argue that Socrates is not an intellectualist about *desires*: We do not *desire* what we desire because we believe it is best.[32] Instead, we desire what we desire (pleasure, beauty and goodness) because of an inherent natural attraction. In Penner, Rowe and Reshotko's interpretation, by contrast, Socrates is an intellectualist about both actions *and* desires; our actions and executive desires (i.e., our desires to do something) follow from our beliefs about what is best. Brickhouse and Smith might say that Penner, Rowe and Reshotko's interpretation got things backwards: According to Brickhouse and Smith, executive desires usually do not follow from beliefs, but beliefs can follow from desires. In other words, beliefs usually do not generate desires, but desires can generate beliefs.

Despite their disagreements, Brickhouse and Smith agree with Penner, Rowe and Reshotko that *appetites* 'do not cause any behaviour all by themselves' (Reshotko 2013: 171), that is, without a belief that this is the best thing to do. But when it comes to *emotions*, Brickhouse and Smith propose a fundamentally different account.

Brickhouse and Smith (2015) argue that emotions are evaluative beliefs brought about by natural attractions or aversions. Fear, for instance, is a belief that results from our aversion to pain (2015: 15). Our natural aversion to pain presents certain actions as best for us to do, thereby inclining us to believe that these actions are indeed best (2015: 19). According to Brickhouse and Smith, this resulting belief is fear. In this interpretation, emotions thus differ from appetites in a crucial way: Since emotions *are* beliefs about what is best to do, emotions can generate actions on their own.[33] An appetite, however, must generate or call to mind a belief – the belief that pursuing the object of the appetite is best – in order to bring about an action. Appetites can prompt us to believe that their objects are good and worth pursuing; emotions are beliefs that their objects are good and worth pursuing. This is in contrast to Penner, Rowe and Reshotko, who claim that neither emotions nor appetites can orient one toward an external object in the absence of other beliefs and the desire for the good.

We saw that Socrates does not disregard emotions and appetites. Emotions and appetites can play some role in generating our actions, even though all actions are ultimately explained in terms of beliefs (we act as we do because we *believe* it is best). Above, I reviewed the two dominant interpretations of Socratic motivational intellectualism and the roles they assign to these mental states. We will see in the next section that these two different interpretations of Socratic motivational intellectualism align with different interpretations of the Socratic response to wrongdoing and specifically of Socrates' stance on punishment. Penner, Rowe and Reshotko have argued that philosophical conversation is the only reliable means for correcting wrongdoers: Since emotions and appetites cannot cause actions (or beliefs about what is best, which then generate actions), they do not need correction through non-argumentative means such as punishment. Brickhouse and Smith, by contrast, have argued that Socrates believes some wrongdoers require punishment to correct their misguided appetites.

WRONGDOING AND PUNISHMENT

According to Socratic motivational intellectualism, we always do what we believe is in our own best interest at the moment of action. Moral wrongdoing ('acting unjustly') is never in our own best interest (*Cr.* 49a–b, *Grg.* 469b). Socrates believes that wrongdoing is the worst thing there is for the wrongdoer (*Grg.* 509b) because it harms the soul (*Cr.* 47d–48a, *Ap.* 30d), and those with harmed souls lead miserable lives (*Grg.* 511c–512b). So, wrongdoers do wrong because they have concluded falsely that their actions are in their own best interest (*Prt.* 357c–358d, *Grg.* 466d–468e). For Socrates, wrongdoing always reflects an intellectual failure.

Since agents do wrong because they are ignorant – they falsely believe that they benefit from wrongdoing – correction efforts must focus on making them less ignorant. One plausible way to make agents less ignorant is via philosophical conversations. Through conversations with Socrates, wrongdoers will become better deliberators; they will get rid of their false beliefs and acquire new true beliefs, and they are thereby more likely to abstain from future wrongdoing. The question at the centre of this discussion is whether philosophical conversations are appropriate for correcting *all* wrongdoers. Are philosophical

conversations the *only* means of correction that Socrates endorses, or does Socrates' intellectualist account of motivation and action leave room for other means of correction?

To answer these questions, one must turn to Socrates' apparent approval of legal punishment in the *Gorgias*:

> Wrongdoing should not be kept hidden but brought into the open, so that [the wrongdoer] gets punished and gets healthy; he should force himself . . . and present himself courageously as to a doctor for cauterization and surgery, pursuing the good and admirable thing without taking into account the pain. And if he is so unjust that he deserves flogging, he should present himself to be beaten; if he deserves imprisonment, to be imprisoned; if a fine, to pay it; if exile, to be exiled; and if death, to die.
>
> – *Grg.* 480c–d

In this passage, Socrates seems to suggest that wrongdoers ought to be punished – fined, imprisoned, exiled and even flogged or put to death.[34] Elsewhere in the *Gorgias*, he explains that these punishments benefit the wrongdoer by improving his soul (*Grg.* 477a–b): The punished wrongdoer gets rid of the 'most serious bad thing', namely 'injustice, ignorance, cowardice, and the like', which lead to misery and a life lived poorly. It is thus in the wrongdoer's own best interest to disclose his wrongdoing and be punished.

The challenge interpreters face is: 'if it [i.e., wrongdoing] is all supposed to be a matter of intellectual error, what use is it to *punish* anyone? . . . How can making people suffer – fining, imprisoning, flogging, exiling, executing them – how can any of *that* make them *think* better?' (Rowe 2007: 28). Interpreters have answered this question in three different ways. Below, we will look at their answers in more detail.

Answer #1: Punishment cannot make us think better. Socrates does not approve of legal punishment.

Some interpreters have argued that, for Socrates, all wrongdoers can be improved *only* through philosophical conversations and thus Socrates does not in fact approve of legal punishment. We can reconstruct their argument as follows:

(i) In Socrates' intellectualist account of motivation and action, 'the only factor that is ever relevant to changing someone's conduct . . . is changing his beliefs' (Penner 2011: 289). If we want to correct wrongdoers, we must improve their beliefs.
(ii) Punishment cannot improve beliefs. For Socrates in Plato's early dialogues, '*only philosophical dialogue* can improve one's fellow citizens' (Penner 2000: 164; see also Penner 2018 and Rowe 2007).
(iii) Thus, Socrates cannot approve of punishment.

In this interpretation, our only hope to improve wrongdoers is changing their beliefs via extended philosophical conversations:

> 'If only we could *discuss* things for long enough, if only we could *understand* what is best,' Socrates seems to say, 'all would be well, and all conduct would be virtuous!' For Socrates, when people act badly or viciously or even just out of moral weakness, that will be merely a result of intellectual mistake.
>
> – Penner 2000: 165

Philosophical conversations are the only reliable means of correction, because the only thing that can make us do wrong is a false belief about what is best, and such intellectual

errors, these interpreters maintain, can only be corrected in an intellectual way: 'nothing apart from talking and reasoning with us will be necessary, because there is nothing apart from what we think and believe that is even in principle capable of causing us to go wrong' (Rowe 2006: 166).

Rowe contrasts this Socratic account of correction with Plato's. For Plato, our emotions and appetites can cause us to do wrong and act against our better judgment (in other words, Plato allows for synchronic belief *akrasia*; for a helpful discussion of this contrast between Socrates and Plato, see Brickhouse and Smith 2010: 199–210). Correction efforts may therefore need to target either a wrongdoer's ignorance *or* his misguided emotions and appetites, depending on the cause of wrongdoing. For Plato, 'our *desires as well as our reason needs persuasion, education, direction. That* is where punishment comes in, as a suitably irrational way of dealing with irrational drives' (Rowe 2007: 29). But for Socrates, by contrast, 'one can only go wrong through ignorance' (Rowe 2007: 24). Thus, in this interpretation, wrongdoers only need philosophical conversations to improve, and Socrates 'does not endorse flogging, imprisonment, or any other vulgar kind of punishment' (Rowe 2007: 36).

To explain *Gorgias* 480c–d, where Socrates seems to approve of legal punishment – imprisonment, paying a fine, exile and even bodily punishment – Rowe proposes that Socrates uses an ordinary notion of punishment simply to make it easier for his interlocutor, Polus, to understand his argument (Rowe 2007: 34). Nevertheless, so the argument goes, Socrates does not approve of any kind of 'punishment' beyond teaching and philosophical conversations, the only 'punishments' he believes can improve wrongdoers (Rowe 2007: 32, 34–35; see also Penner 2000, 2011, 2018; Edwards 2016[35]).[36]

Answer #2: Punishment cannot make us think better, but certain forms of punishment can deprive us of the means that facilitate wrongdoing. Socrates approves of these kinds of legal punishment.

Shaw (2015) has argued that Socrates endorses only certain forms of punishment – the death penalty, exile and confiscation of property – because these punishments deprive the wrongdoer of the means that facilitate his wrongdoing. Paying a fine, for example, deprives one of money; exile deprives one of friends; and the death penalty deprives one of life. Money, friends and being alive are means that enable an agent to do wrong. In other words, Socrates approves of punishment only for the sake of incapacitation or the deprivation of the means to do wrong. But since certain bodily punishments like flogging cannot be justified in this way, according to Shaw, Socrates does not approve of them.

Shaw argues that Socrates' apparent endorsement of flogging in *Gorgias* 480c–d does not provide sufficient grounds for concluding that Socrates approves of this type of bodily punishment, because his approval of flogging is conditional: A wrongdoer should be flogged *if* he is so unjust that he deserves to be flogged. But this condition, Shaw argues, might never be met (2015: 79; see also Moss 2007: 232 n.8).

Answer #3: Punishment can make us think better by improving our misguided appetites. Socrates approves of legal punishment.

Brickhouse and Smith (2010: 102–110; 2018: 45–47) have argued that we must distinguish between two different kinds of wrongdoers. Some wrongdoers commit crimes due to mere ignorance; they might miscalculate costs and benefits or lack certain

information. Teaching and instruction are appropriate for correcting these wrongdoers. In the *Apology*, Socrates argues that he himself belongs to this group of wrongdoers, if indeed he harmed anyone (*Apology* 26a). But other wrongdoers, Brickhouse and Smith argue, commit crimes due to strong appetites. These wrongdoers are ignorant 'in a different sense' (2010: 123): Their appetites cause false beliefs and make them disinclined to follow reason. For such wrongdoers, 'calm conversation . . . would not be effective' (2015: 26). These wrongdoers cannot hope to improve 'unless they undergo punishment' (2010: 123).

When we act to satisfy our appetites, Brickhouse and Smith argue (2010: 117–124; 2018: 47), our appetites become stronger. In some wrongdoers, appetites have become so strong that they severely impair the ability to listen to reason. For these wrongdoers, appetites habitually cause them to believe that the objects of their appetites are good and worth pursuing; acting on those appetites then further strengthens them. According to Brickhouse and Smith, punishment can break this vicious cycle by changing wrongdoers' calculations about what is beneficial to them: The punished wrongdoer comes to believe that the pleasure from satisfying appetites is not worth the pain from punishment (2010: 124). Since punishment is unpleasant or even painful, it gives appetitive wrongdoers a convincing reason to avoid future wrongdoing; and when they abstain from wrongdoing and do not act on their appetites, their appetites are not 'filled up' and become 'weaker' (2010: 123–124; 2018: 51). Weakened appetites are less likely to cause the agent to believe that the objects of those appetites are good and worth pursuing. Once appetites are weakened, a wrongdoer can see the objects of those appetites for what they really are – merely apparent goods (2010: 124).

In this interpretation, painful punishment brings about a *conative* improvement by weakening the wrongdoer's excessively strong appetites. This conative improvement in turn may bring about an *epistemic* improvement: The weakened appetites are less likely to cause false beliefs about what is the best thing to do. Brickhouse and Smith thus conclude that Socrates is serious when he approves of legal punishment.

The advantage of Brickhouse and Smith's interpretation is that it takes the textual evidence in the *Gorgias* at face value. However, critics have argued that, even setting aside the question of whether painful punishment is consistent with Socrates' psychology of action, we still have reason to reject Socrates' apparent approval of painful punishment.

Critics have worried that painful punishment will not improve wrongdoers by making them less ignorant; instead, it could make them worse and more ignorant, either by enforcing what Socrates holds to be a false belief, namely that 'pain is bad' (Kamtekar 2016: 6, n. 13), or by creating a new false belief, namely that 'getting caught is bad' (as one might infer from Shaw 2015: 76). The agent would then become a sophisticated wrongdoer: someone who merely tries to avoid getting caught. For responses to these objections, see Möbus (2023).

Further, a critic could argue that Socrates' alleged approval of punishment is incompatible with his belief that no one does wrong willingly. In the *Apology*, Socrates argues that only wrongdoing done willingly should be legally punished. Unwilling wrongdoers should be taken aside and instructed privately. But if indeed all wrongdoing is unwilling, then it seems that no wrongdoer should be legally punished:

(i) Only willing wrongdoing should be legally punished (*Ap.* 26a).
(ii) No one does wrong willingly (*Prt.* 345e, 358d–e; *Grg.* 509e).
(iii) So, no one should be legally punished.

But several interpreters (Kamtekar 2017b: 72; Brickhouse and Smith 2018) have argued that in the *Apology*, Socrates claims that some people actually do wrong willingly, namely, those who exhibit a certain kind of ignorance – those who do not know that if they harm their associates, they will be harmed in turn (*Ap.* 25c–d). How can Socrates claim both that all wrongdoing is unwilling and that some wrongdoers do wrong willingly and should be legally punished?

Brickhouse and Smith (2018) have proposed that all wrongdoing is involuntary (they translate *akōn* as 'involuntary' instead of 'unwillingly') because all wrongdoing is self-harm – it harms one's soul – and no one wants to harm oneself.[37] But they argue that this is perfectly compatible with the idea that the same agent could voluntarily harm others. An action 'can be voluntary in one sense and yet involuntary in another' (2018: 52): A wrongdoer can voluntarily harm his victim while at the same time involuntarily harming himself (Brickhouse and Smith 2018: 50–51). This is why Socrates can claim both that no one does wrong voluntarily *and* that some wrongdoers do wrong voluntarily and thus should be legally punished.

Against all three interpretations presented above, Freya Möbus (2023) has argued that Socrates in the *Gorgias* approves of painful bodily punishment like flogging (*pace* Penner and Rowe; *pace* Shaw) and that we can explain the efficacy of bodily punishment without introducing non-rational desires (*pace* Brickhouse and Smith). Möbus proposes that experiencing bodily punishment can benefit the wrongdoer in two ways: It can epistemically improve the wrongdoer by prompting him to form the new true belief that wrongdoing is bad for him, and it can prevent a further epistemic worsening of the wrongdoer by deterring him from future wrongdoing, at least sometimes. Bodily punishment can improve certain wrongdoers under certain circumstances, Möbus explains, because experiencing bodily pain and feeling that wrongdoing is bad is more persuasive than mere philosophical arguments. In Möbus' account, flogging can be educationally effective precisely because it is painful (*pace* Shaw).

IS SOCRATES AN INTELLECTUALIST ABOUT HUMAN MOTIVATION?

In this chapter, we discussed Socratic motivational intellectualism – the claim that we always do what we believe is the best thing for us to do, and we saw that many place motivational intellectualism at the centre of Socrates' account of human action. Some, however, have pushed back on this interpretation.

Kamtekar (2017a: 72–73, 2017b: 1–4, 69–128) has argued that we should not put so much weight on Socrates' alleged motivational intellectualism because we have only *one* piece of textual evidence for it, *Protagoras* 358c–d. Further, this one piece of evidence, Kamtekar argues, stands on shaky ground: Socrates says that '*if* the pleasant is the good', no one would do something if he believed there was something better that he could do. Since Socrates presents motivational intellectualism as premised on hedonism (i.e., 'the pleasant is the good'), our interpretation of his account of motivation hinges on the question of whether Socrates is a hedonist. While Kamtekar has argued that Socrates rejects hedonism (pointing us, for example, to *Grg.* 493a–495a), others have proposed that he only rejects a certain kind of hedonism (Rudebusch 1999, Moss 2014).

The foundation of Socrates' account of motivation is not intellectualism, Kamtekar proposes, but rather the desire thesis – we all desire our own real good. The desire for our real good, Kamtekar argues further, 'may be manifested in different ways: certainly by

our pursuit of what we believe to be best, but also by our pursuit of pleasant things and fine things' (2017b: 3). In other words, we do not *always* and *only* do what we *believe* is best for us to do; sometimes we do what merely seems pleasant, for example. If Kamtekar is right, Socrates is not an intellectualist about human motivation after all.

The conversation about Socrates' account of motivation remains ongoing, and the implications for his philosophy of human nature have yet to be determined: Are we indeed beings who *always* do what we believe is in our own best interest? Do we come equipped with an inherent desire for what is truly good for us? For Socrates, our answers to these questions matter because if we act well, we live well; so, to live a good life, we must understand why we act as we do.

NOTES

1. I am very grateful to editors Rusty Jones, Nicholas Smith and Ravi Sharma for the invitation to write this chapter and for their extraordinarily helpful comments on earlier versions of it. I would also like to thank Naomi Reshotko and Antonio Chu, whose detailed feedback helped me present several arguments more clearly.

2. By 'Plato's early dialogues', I here mean the *Euthyphro*, *Apology*, *Crito*, *Charmides*, *Laches*, *Lysis*, *Euthydemus*, *Meno*, *Protagoras*, *Ion*, *Hippias Minor* and *Major*, *Gorgias* and *Republic I*.

3. All translations used in this chapter are from Cooper (ed.), *Plato: Complete Works* (1997). I have occasionally made small changes to the translation for certain passages, in which case I provide the Greek in brackets.

4. The view that I always do what I believe is best *for me* is often referred to as psychological egoism. Many understand Socrates' motivational intellectualism in this way (see e.g., Brickhouse and Smith 2013b: 185; Penner 1992a: 128), specifically his remarks in the *Protagoras* (see e.g., Reshotko 2013: 179; Taylor 2019: 60–62). Others, however, have pushed back against the claim that Socrates is a psychological egoist (see e.g., Jones and Sharma 2017: 132–133) and have offered alternative interpretations of the *Protagoras* passage (see e.g., Ahbel-Rappe 2012: 332–335). One might worry that if Socratic agents are psychological egoists, then they further their own good at all costs, even at the expense of the well-being of others. But some interpreters have argued that, for Socrates, the ethical good and our prudential good never come apart (Brickhouse and Smith 2010: 44–46): What I ethically ought to do is always also what is in my own best interest; likewise, wrongdoing is never good for me.

5. I hope to offer an understanding of 'motivational intellectualism' that is shared by many interpreters. Parts of the definition are thus intentionally left vague to avoid biases for or against certain views (I therefore propose, e.g., that motivations 'are tied to' beliefs instead of 'are', 'follow', 'are caused by' or 'derive from').

6. For a very helpful outline of Socrates' motivational intellectualism and its centrality for his ethics, see Taylor (2019: 60–62). Kamtekar (2017b: 1–3) has questioned what she calls the 'mainstream account', according to which motivational intellectualism is the 'theoretical basis' or 'foundation' for the 'Socratic intellectualist package'. I will return to her view at the end of this chapter.

7. Among those who have argued that, in the Socratic account, non-rational desires do not motivate actions are Irwin (1977: 78–82), Penner (in various articles; see e.g., 1992a) and Rowe (see especially his 2012a).

8. See e.g., Penner (2000), Rowe (2007), Brickhouse and Smith (2010: 199–210) and Taylor (2019: 60–62).
9. Kamtekar (2017b: 69–128).
10. While in this passage Socrates refers to wisdom, health and wealth as 'goods', he describes what I call 'small-scale goods' – for example, sitting, walking and embarking on sea voyages – as 'neither good nor bad'; they 'sometimes partake of what's good, sometimes of what's bad, and sometimes of neither' (*Grg.* 467e).
11. Penner and Rowe (1994) and Reshotko (2006: Chapter 2) argue against this widely accepted interpretation of the *Meno* passage.
12. Segvic (2009), Devereux (1995: 398ff.), Santas (1964: 152 n.15) and others also distinguish between *boulēsis* and *epithumia*.
13. Santas (1964) argues that we should not 'lump together' these passages. We must distinguish, Santas argues, between prudential and moral wrongdoing and thus between Socrates' prudential paradox that 'no one does what is bad for oneself (i.e., prudentially wrong or bad) willingly' and his moral paradox that 'no one does what is unjust (i.e., morally wrong or bad) willingly'. According to Santas, Socrates takes the prudential paradox to be a fact about human nature that serves as a premise in his argument for the moral paradox.
14. I thank Rachel Barney for a helpful discussion of this argument and this premise in particular.
15. For a similar reconstruction, see Kamtekar (2017b: 70–71).
16. Reshotko (2013: 170) attributes this view to Irwin (1977, 1995), among others.
17. Socrates believes that knowledge *akrasia* of any kind – diachronic or synchronic – is impossible (Penner 1990: 47). Diachronic knowledge *akrasia* is impossible because knowledge is stable; it does not vacillate. Synchronic knowledge *akrasia* is impossible because we never act against what we 'know or believe' is best for us in the moment of action (*Prt.* 358b–d).
18. Callard proposes 'simulacrum' instead of 'appearance' as a translation of 'phantasma' in *Prt.* 356d. 'A simulacrum is a representation not believed to be veridical by the one who has it' (2014: 52).
19. Rowe's 2012a discussion of Brickhouse and Smith's *Socratic Moral Psychology* and their 'Reply to Rowe' (2012b), as well as Reshotko's discussion of the different interpretations (2013: 170–172), are particularly helpful for understanding the debate. For a helpful discussion of similarities and differences between Devereux's (1995) interpretation and Brickhouse and Smith's, see Brickhouse and Smith (2013b: 193–194). For a critical discussion of Brickhouse and Smith's *Socratic Moral Psychology* by Jones (2012), Butler (2012) and Devereux (2012), as well as Brickhouse and Smith's response to these critics (2012a), see *Analytic Philosophy* (53.2).
20. As Kamtekar notes as well: 'the terms we use – desire, belief, appetite, nonrational – are so familiar that we do not usually stop to ask what we are saying' (2012: 259). She identifies two ways of understanding 'non-rational': 'good-independent motivations (that is, motivations that do not represent their objects as good), and uncritical (but quite possibly good-directed) motivations' (2012: 257).
21. See, for instance, Singpurwalla (2006: 243, n.1): 'a non-rational desire is a desire that arises independently of reasoning and so has the potential to come into conflict with our

reasoned conception of the good'. For the idea that non-rational desires are not responsive to belief and deliberation, see Penner (1992a: 128–129). See also footnote 22 below.

22. Singpurwalla (2006) argues that non-rational desires are beliefs and that they can thus be changed by reasoning: 'The fact that Socrates conceives of irrational desires as evaluative *beliefs*, as opposed to desires that are independent of any evaluation of the object of desire or appearances, opens up an interesting possibility, namely, for Socrates irrational desires are resistant to reason but not invariably or essentially so. Irrational desires are still beliefs, and the aim of beliefs is to represent the world; thus any evaluative belief is sensitive to evidence' (2006: 253). If non-rational desires did not involve beliefs, it would be a 'mystery' how they could affect or be affected by our beliefs (2006: 251).
23. For the idea that our non-rational desires are independent of our desire for the good, see, for instance, Reshotko (2006: 54–55, 74). For the idea that our non-rational desires are independent of our beliefs or deliberations about goodness, see, for instance, Rowe: 'Desire, taken by itself, i.e., until beliefs are "plugged into" it, will *always* be non-rational; . . . desires only become desires "to do something" when they are combined with beliefs, i.e., about what is best for the agent; and that is also the moment at which they become "rational" desires' (2012a: 309); Singpurwalla: non-rational desires arise independently of 'reasoned beliefs about value' (2006: 249); Irwin: rational or good-dependent desires (Irwin seems to use these terms interchangeably, 1977: 78) 'rest on deliberation about what would be best, all things considered, for myself as a whole' (1995: 215) – they are 'formed by deliberation about instrumental means to the final good' (1977: 170). Non-rational desires, by contrast, do not arise out of deliberation about the final, overall good. Brickhouse and Smith (2013b: 191) distinguish 'non-rational' from 'good-independent' desires as follows: Non-rational desires are 'desires that seek their objects in a way that is independent of our reasoning or deliberation about what is good for us'. Good-independent desires are 'desires that seek their objects in a way that is independent of our universal desire for what is really good for us'; for a discussion of different usages of the term 'irrational', see Carone (2005: 377, n.37).
24. The term 'independent' should here be understood as 'not caused by' or 'not dependent upon' reasoning or deliberation; it should not be understood as 'totally unresponsive to' reasoning or deliberation (Brickhouse and Smith 2013b: 196).
25. Note that Brickhouse and Smith think of our basic attractions and aversions as 'non-rational' but 'good-dependent' (Brickhouse and Smith 2013b: 191). Attractions and aversions are good-dependent because they present their objects as good.
26. Brickhouse and Smith (2010, 2015, 2018) and Martinez and Smith (2018: 70–73) have argued most prominently and extensively for the idea that emotions and appetites can influence our actions in the Socratic account of motivation. But Penner, Rowe and Reshotko have also acknowledged this idea. See e.g., Penner: '"Does desire for drink never generate an action? How can that be? Is it being denied that we have these desires?" No, the desire for drink does occur, but the way it gets us to act is to present itself to our desire for happiness, which turns to the belief-system to produce an estimate of the possible gains from various choices for fulfilling this desire' (2011: 263). See also Reshotko: 'I hold that urges and drives *do influence* our rational assessment of different courses of action. . . . My craving for chocolate makes my calculation of the good, and my consequent actions based on my desire for the good, come out differently than they would, had I not been craving

chocolate' (2006: 87); 'Intellectualism need only claim that these non-intellectualized factors never cause behavior in an unmediated fashion: They cause it by affecting our beliefs. These changed beliefs influence our deliberation concerning which action is the best means to the best end available to us in our situation, so we come to different conclusions about which action is most beneficial' (2006: 84); 'No purposeful action can be the result of a non-rational element (like emotion) except insofar as the non-rational element has influenced the agent's beliefs' (2006: 16). See also Penner and Rowe: 'For most such "feelings" *are* intimately connected with beliefs and actions. Indeed, it is hard to imagine a feeling that does not somehow influence some belief the subject has' (2005: 230).

27. Reshotko attributes the idea that, in the Socratic account, emotions and appetites drive or urge us to Penner: 'For many years, the dominant interpretation of the Socratic denial of *akrasia* . . . depicted Socrates as denying that people feel the urge to do things other than what they believe to be good for them. Penner challenges that tradition . . ., finding that Socrates acknowledges the experience of urges and desires that are not yet integrated with the desire for the good and, so, can be said to conflict with it' (2013: 170).
28. See Reshotko: 'Neither Penner nor I deny that an agent continues to feel an urge even while acting against it in an effort to do what is best. What we deny is that this urge can be a *pull towards a specific instance of a thing*' (2013: 171). See also Brickhouse and Smith's criticism of Penner's and Reshotko's account: 'Missing from this account, we contend, is what is peculiar to the appetites and passions, namely, that they are "drives" and "urges," that is, that they are psychic events that actually do *drive* or *urge* us towards and away from things' (2010: 52, n.6).
29. See Reshotko: 'Devereux believes that my appetite for chocolate can actively work to distort my beliefs concerning a candy, *causing* me to *see the candy as a means to my happiness*. . . . This is not how I have presented the role of unintellectualized drives and urges. In my view, an appetite never plays a role that is more instrumental than any other piece of information that the intellect has used in order to determine what it is best to do as motivated by the desire for the good. I hold that appetites are like sense impressions: they are phenomena that help us form our judgments, but they do not interact with judgments that have already been formed' (2006: 85–86). See also Rowe (2012a: 314) and Penner (2011: 263–264).
30. See, for example, Reshotko: 'While other drives and urges might exist, Socratic intellectualism dictates that they cannot trump or triumph over the desire for the good' (2006: 85, 76–77).
31. Brickhouse and Smith: 'non-rational desires influence [. . .] what we believe by representing their targets as goods or benefits to the agent, so that the agent would come to believe that pursuing or obtaining those targets would serve the universally shared desire for benefit, unless some other process interfered with this' (2015: 11); 'Our very natural attractions and aversions [. . .] present to the soul representations of what is best for us, inclining the agent to come to believe that doing whatever the attraction or aversion indicates actually is the best thing for the agent to do' (2015: 19). Brickhouse and Smith sometimes also describe this process in terms of causation (appetites can cause beliefs about what is best, see Brickhouse and Smith 2010: 53–62, 104; 2015: 11). Both 'causation' and 'inclination' are supposed to capture the same idea: Appetites lead to beliefs about what is best to do by default, unless some other belief-forming process intervenes (Brickhouse and Smith 2015: 19).
32. Jones (2012).

33. In order to be able to generate actions, these beliefs will have to present very specific actions as the best actions to do. Brickhouse and Smith do not give examples of such beliefs. However, in the case of a hoplite's fear, for example, they might suggest a belief such as 'throwing away my shield and running north up the mountain path is the best thing for me to do'.
34. Ravi Sharma has noted that the quoted passage (*Grg.* 480c–d) is dependent upon an antecedent of the form 'if oratory is to be used to expose wrongdoing'. He points out that the 'should' in 'wrongdoing should not be kept hidden' is thus a function of the way one must use oratory to accomplish that goal. Socrates is, then, not making a moral demand, and it is not clear that he is expressing his own view.
35. Fay Edwards (2016) has argued in support of Penner and Rowe's view that Socrates understands teaching as a form of punishment. Her argument focuses mainly on a puzzle in the *Euthyphro*: Socrates claims that if he were to gain knowledge of piety, Meletus should drop the charges of impiety because he will have been punished. But why would learning about piety secure Socrates' acquittal? Edwards argues that the answer lies in understanding instruction as punishment. If Socrates learns about piety, he will have been appropriately punished and there is no need to bring him to court. Edwards suggests that this result 'might encourage us to approach Socrates's statements about punishment in other dialogues, such as the *Gorgias* and *Protagoras*, with caution, as Penner and Rowe themselves suggest, as what Socrates means when he talks about punishment may not be what it, at first, seems' (2016: 18). Brickhouse and Smith have responded to Edwards, arguing that Socrates distinguishes between instruction and punishment: 'Socrates seeks education from Euthyphro that would make further pursuit of the prosecution against him otiose, but not because that education counts as a form of punishment, but because any errors Socrates may have made in the past are the sorts *that do not merit punishment*' (2017: 61).
36. Another possible explanation of Socrates' endorsement of punishment in the *Gorgias* is that this dialogue (either partially or completely) is not in fact 'Socratic' but instead is 'transitional', serving as a point of transition between Plato's early and middle dialogues (Irwin 1977: 291 n.33; 1995: 114). In his transitional dialogues, Plato uses Socrates more and more to express his own views. Thus, one could argue that it may be Plato, not Socrates, who endorses painful punishment.
37. How exactly does wrongdoing harm the soul? Brickhouse and Smith (2007a, 2013b: 204–207) discuss two possible explanations: Wrongdoing harms the soul because it leads the wrongdoer to acquire new false beliefs, or wrongdoing harms the soul because it leads the wrongdoer's appetites and passions to become less disciplined. They argue for the second explanation.