CHAPTER 4

*Socrates and Coherent Desire (*Gorgias *466a–468e)*

Eric Brown and J. Clerk Shaw

4.1 Introduction

In Plato's *Gorgias*, Polus says that he and others esteem orators for their tyrannical power (466a9–c2). Socrates grants that orators can act tyrannically, by doing whatever seems best to them (*dokein beltiston*). However, he also insists that orators and tyrants do almost nothing they want (*boulesthai*), so that they have the least power in their cities (466c9–e1). This perplexes Polus, and most readers are in the same boat. What is wanting, and how does it differ from thinking best? And why does Socrates suppose that orators and tyrants fail to satisfy their wants (*boulêseis*)?

Socrates argues for his distinction between wanting and thinking best (467c5–468e5), but the argument has proven difficult to grasp. One key claim in the argument is that wanting is for one's real good (468c2–d7). If wanting is for one's real good, and orators and tyrants fail to achieve their real good, then they fail to do what they want. However, merely *stipulating* that wants are for the real good is unlikely to convince anyone that such desires are real, let alone that orators and tyrants fail to satisfy them.[1] This approach presents textual problems too: Socrates motivates his

For comments on earlier drafts, we thank Scott Berman, Thomas Bonn, Agnes Callard, Jake Daly, Rachana Kamtekar, Casey Perin, and Rachel Singpurwalla. We also thank audiences at the Arizona Colloquium for Ancient Philosophy, Duke University, the University of Tennessee, and the University of Utah for their questions and objections – especially Hugh Benson, E.J. Coffman, Michael Ferejohn, Eric Hutton, Rachana Kamtekar, Mathew Lu, Elijah Millgram, Tim O'Keefe, Nick Smith, and Josh Watson.

[1] For references to and discussion of earlier readings that focus on this claim that wanting is for the real good, see McTighe (1992). More recent and developed versions of such readings include Penner (1991) (cf. Penner 1988 and Penner and Rowe 1994), and Segvik (2000). We relate our reading to these and others as we go, but we do not respond systematically to any. For an overview of the literature, see the opening pages of Kamtekar (2006).

notion of wanting subjectively, from the agent's perspective (467c5–e1). But this introduces a new difficulty: how does Socrates combine a subjectively motivated notion of wanting with the objective constraint of being for one's real good? Many readers hold that Socrates moves from the former to the latter fallaciously.[2] In light of these puzzles, some go on to infer that he exaggerates his view or only argues *ad hominem*,[3] despite his later approving reference back to this argument (509e5–7).

The main problem in grasping Socrates' argument about wanting lies in accommodating subjective and objective constraints on this sort of desire. We offer a reading that does just that, focusing especially on how Socrates motivates his claims from any agent's perspective. First, we examine the lead-up to his argument and especially his initial refutation of Polus (4.2). We then reconstruct the argument about wanting (4.3). Next, we directly confront difficulties in this argument, including the deeper nature of wanting and why Socrates assumes that wanting is for one's real good (4.4). Finally, we return to orators and tyrants. Socrates' argument about wanting does not establish that orators and tyrants fail to satisfy their wants, despite occasional allusions to this claim. By looking at the end of our passage and at its larger context, we identify the specific wants that Socrates thinks they fail to satisfy, and why he supposes these specific wants are for real goods (4.5). Overall, we provide a reading of *Gorgias* 466a–468e on which Socrates argues sincerely and plausibly, and on which this passage is integral to the dialogue as a whole.

4.2 The Setup

The main themes of our passage emerge from the first part of the dialogue. Polus expresses admiration for orators like Gorgias early on (448c4–9), and Gorgias agrees that oratory is admirable (451d6–8, 452d5–8). However, Socrates exposes an ambiguity in Gorgias' conception of oratory: either it is a general ability to persuade on all topics without knowledge of any, or it is (or at least requires) expert knowledge of justice. Gorgias is keen on the power that oratory, understood in the first way, offers (452e1–8, 456a1–c7, 459c3–5). However, he is ashamed to deny that orators know justice, and Socrates' refutation of him starts from the notion of oratory as requiring such knowledge (458e3–461b2). When he concludes this refutation and Polus demands that Socrates give his own account of oratory,

[2] See Irwin (1979); Santas (1979); Vlastos (1991); McTighe (1992); and Waterfield (1994).
[3] See Vlastos (1991); McTighe (1992); Weiss (1992); and Segvik (2000).

Socrates returns to and develops the other conception, arguing that oratory only pretends to know justice and flatters its audience by guesswork (462b3–466a8).

This dismal picture of oratory does little to discourage Polus; he only cares that orators are respected and powerful in their cities (466a9–c2; cf. 462c8–9). These are the same features Gorgias stressed earlier, and to which Socrates said they would return as needed (459c6–8). When Polus claims that orators have power, Socrates quickly contradicts him: orators have least power in their cities, assuming (*toinun*) that power is something good for its possessor. Polus agrees that power is good for its possessor, but insists that orators have it: "[D]o not orators, as if they were tyrants, execute whomever they want [*boulôntai*], and confiscate property from and exile whomever they see fit [*dokêi autois*]?" (466b11–466c2) Socrates claims these are two separate questions, and he draws a distinction: "I say, Polus, that both orators and tyrants have the least power in their cities, as I was saying just now, because they do just about nothing that they want [*boulontai*] although they do whatever they think is best [*autois doxêi beltiston*]" (466d6–e2). This flabbergasts Polus, who assumes that the ability to do what one thinks best is the same as the ability to do what one wants (466e3). But Socrates rejoins that Polus does *not* think these abilities are the same (466e4): he supposes that power is good for its possessor (466e6–8), but also grants that doing what one thinks best without intelligence [*noun mê echôn*] is bad (466e9–11, 467a4–5).

At this point, Socrates has elicited from Polus four mutually inconsistent claims about power, which largely abstract away from the original question about orators and tyrants:

(a) Power is the ability to do what one wants.
(b) Doing what one wants is the same as doing what one thinks best.
(c) Doing what one thinks best without intelligence is bad for the agent.
(d) Power is good for its possessor.

Given (a) and (b), Polus must grant that power is the ability to do what one thinks best. But then, given (c), he must concede that power is bad for unintelligent agents, which contradicts (d). At least one of these claims must go. Socrates offers to show Polus why he should reject (b) (467b11–c4), and this is the burden of the immediately ensuing argument.

First, though, we might ask about the strength of claims (c) and (d). Must power be *always* good for its possessor, or could it be good in some weaker way? And if it is good in a weaker way, could Polus admit that unintelligent action sometimes harms its possessor without falling into an

inconsistent tetrad? At this stage, we note two things. First, the stronger (d) is as a claim about power's goodness, the weaker (c) can be and still produce a conflict. Second, Polus is committed to a quite strong claim about power's value. Both he and Gorgias have in mind power that enables one to live well. They think that since oratory offers such power, it is the finest craft, providing its possessor greater benefits than any other expertise (451d–452d). If power is this sort of good, then (c) need say only that unintelligently doing what one thinks best will not enable one to live well. (One might also wonder whether Socrates accepts a strong version of [d], and if he does not, whether this shows that he proceeds to argue only *ad hominem*. At the end of 4.4, we argue that Socrates accepts a very strong version of [d] indeed.)

Polus might also respond by denying the relevance of this puzzle about power to orators. If orators are intelligent, then (c) is irrelevant to them, and they will not harm themselves by doing what they think best. Socrates invites this response, asking Polus to refute his claim that oratory is not an expertise and to show that orators are intelligent (466e13–467b2). Socrates admits that orators are clever (463a6–8), but only expertise systematically distinguishes benefit from harm. However, Polus keeps admiring orators even after hearing they lack expertise (466a9–c2). He does not grant that orators lack intelligence; like Gorgias, he supposes that the relevant intelligence – knowledge of what is good, fine, and just (459d) – is easily had.[4] Still, rather than grappling with Socrates' argument that orators lack intelligence, he insists that they have power anyway (466a9–c2). In this context, Polus cannot deny the relevance of (c) to orators. Instead, he and Socrates carry on to Socrates' argument against (b).

4.3 Wanting and Thinking Best

Socrates' argument that wanting differs from thinking best falls into three main stages. In the first stage, he articulates a principle about the relationship between actions and ends, and says how wanting figures into this picture (467c5–e4). In the second stage, he introduces an axiological division that develops and explains the initial principle about actions and ends. This stage carries over earlier claims about how wanting relates to the principle about action, and introduces claims about how thinking best relates to it (467e4–468c1). In the final stage, Socrates shows how the first

[4] This is precisely why Gorgias is ashamed to deny that orators know the just fine, and good; cf. Shaw (2015: Ch. 4).

two stages entail that one can do what one thinks best without doing what one wants (468c2–e5).

In the first stage, Socrates introduces a principle about action (467c5–e1). The basic idea is that when agents act, they want some end. His first formulation runs like this: "Do you think that people want [*boulesthai*] to do whatever they are doing on a given occasion or do they want that for the sake of which they do what they are doing?" (467c5–7). He illustrates with examples: nobody wants to take painful medicine or to set out on grueling and risky sea voyages. Instead, they want the health produced by medicine or the wealth produced by sea voyages (467c7–d5). This principle and its examples are well-motivated from the agent's own perspective. Patients and seafarers are not so dim as not to recognize these ends of their actions. They want and act for the sake of health and wealth, rather than treatment and voyages, *by their own lights*. Thus, the start of the argument does not simply posit a want for ends independent of how agents conceive of their actions.

This initial statement, and the examples that illustrate it, might at first seem to imply that nobody ever wants an action itself, but only some distinct product. But Socrates immediately rephrases his point conditionally: "*if* a person does anything for the sake of something, he does not want this thing that he's doing, but the thing for the sake of which he's doing it" (467d6–e1). The rephrased version allows that one might do an action for its own sake, without saying so explicitly. We suppose the initial phrasing and examples (which feature painful actions and distinct ends) are there for Polus' benefit: they make it easy to see the point that agents do what they do for the sake of an end. The rest of the passage confirms that Socrates allows for the possibility of actions done for their own sake, and also that he generally pitches his argument at a level Polus can understand.

Socrates now introduces a threefold distinction among things that are good, bad, and neither good nor bad, or "intermediate" (*metaxu*, 467e1–4). Polus agrees that goods include wisdom, health, and wealth, and bads the opposites of these. Intermediates include objects such as sticks and stones and action-types such as walking or sailing, some of whose tokens are good and others bad. It is no accident that the examples in this stage overlap with those in the first stage. In that stage, sailing was an example of something one does for the sake of an end, while wealth was an example of an end for the sake of which one sails. This axiological distinction helps to explain the first stage's claim. Earlier, Socrates argued inductively that actions aim at ends, and that we want ends, not the actions we do for their sake. The axiological division offers the principle behind this claim: the actions we do for the sake of ends are intermediates,

while the ends for the sake of which we act are goods.[5] Again, this is motivated subjectively, from the agent's perspective. One does not sail for the sake of wealth unless wealth seems good by one's own lights.

This more-developed principle about action raises a new version of our earlier question. Before, we asked whether anybody ever wants an action as an end; here, that question becomes whether there are any actions in the class of goods. When Socrates introduces the axiological distinction, he mentions actions only in the class of intermediates (walking and sailing), and not in the class of goods. This might seem to show that we never want to do actions after all. However, Socrates later says that while orators and tyrants do not want to, say, banish others *simpliciter*, they do "want to do these things if they are beneficial" (468c3–4). In other words, they want to banish beneficially, and this leaves open whether such banishing is beneficial for its results or because of some feature of the action itself.

Ultimately, then, Socrates' principle about action says that when agents act, they want some good end; for all he says, that good end may be internal to the action.[6] At the same time, every action can be described in neutral terms such that its tokens are sometimes good, sometimes bad, and as such intermediate. For example, helping friends might be a good action-type. But any action that counts as helping friends also falls under some intermediate description – for example, "returning weapons". In such cases, we return weapons for the sake of helping friends, even if these are the same token action. As with actions that have a productive end, here too the intermediate action-type and the good action-type can come apart (cf. *R.* 331c–d).[7]

Socrates' axiological division leads to significant complications to which we return, but for now we press on. The next step is to gain agreement that the intermediate actions we do for the sake of good ends are what the agents think best (*oiomenoi beltion*, 468b2; *oiomenoi ameinon*, b6). In contrast, again, what we want are the good ends at which our intermediate actions aim (468b7–c1), not the intermediate actions without qualification (468c2–8; *haplôs houtôs* at c3). Thus, Socrates' fully developed principle about action says:

[5] Contrast McTighe (1992), who finds reasons for Socrates' principle only in the induction, and deems the argument sloppy or defective.

[6] So, too, Vlastos (1991: 150n77); Segvik (2000: 40–43); Wolfsdorf (2008: 122–124); contrast Irwin *ad* 467cd, 467e–468a, and 468bc; and Reshotko (2006). Penner (1988, 1991) also encourages an instrumentalist reading of the principle about action, but Penner and Rowe (1994: 7n12) seem to suggest a different view.

[7] Cf. *Symp.* 180e–181a, 183d. Cf. also recent discussions of "particularism".

> Whenever an agent acts, they *think it best* to do some intermediate action for the sake of some good end, but they *want* the good end for the sake of which they act, not the intermediate action they think best.

From this, it follows that:

> Whenever an agent acts, and they manage to do the intermediate action they think best, but they do not realize the good end for the sake of which they act, they do what they think best but they do not realize what they want.

Socrates and Polus agree that this can in fact happen. Socrates thereby establishes his distinction between doing what one wants and doing what one thinks best. By extension, he also shows that one can do what one thinks best without having great power, understood as the ability to do what one wants (i.e., to realize the good ends of one's actions). Polus sees this and sulks; Socrates must press him to answer twice (468c6–7, d6).

However, this argument leaves two loose ends. The first concerns the relationship between subjective and objective construals of the claim that agents want good ends. By the end of the argument, Socrates seems to say that every agent wants what is really, objectively good for them. This is suggested most of all by his phrasing in talking about good ends (*tunchanei de on kakion*, 468d3–4; *tunchanei tauta kaka onta*, 468d5–6). We could stretch to understand these phrases as implicitly qualified: in fact they are worse *by the agent's own lights*. But most plausibly, by the end of the argument Socrates claims that every agent wants what is really, objectively good for them. Rhetorically, he proceeds from a subjective construal to an objective one over the course of the passage. His first steps clearly have subjective import, and can be understood purely subjectively (467e–468b). Next, he offers an ambiguous claim: orators and tyrants do not want to banish and so on without qualification, but "only if these things are beneficial" (468c2–5). This can be read subjectively or objectively without distortion. Finally, he clarifies that his claims have objective purport (468d1–7).

Part of what explains this sequence is that Socrates pitches his argument at Polus, starting from claims he will readily understand. However, this raises worries that he slides from subjective construals to objective ones, thereby arguing fallaciously. Here we must ask what his considered view is. Socrates' clarification does not underwrite a purely objective reading of our good ends through the entire passage. The subjective versions of his claim are not mere provisional versions to be rejected later.[8] Rather, his considered view seems to be that we want what is, in fact *and* by our own

[8] Cf. Brennan *apud* Kamtekar (2006: 135n14); contrast Wolfsdorf (2008: 126, 128–129).

lights, good for us. But this does not explain how he is entitled to such a conception of wanting. Below, we argue that Socrates can reasonably connect the subjective and objective sides of wanting. For now, we simply claim that Socrates' view has this feature, but we grant freely that this requires further explication and defense (4.4).

The second loose end is that while Socrates' argument shows that agents can in principle do what they think best without doing what they want, he has not explained how this actually happens. One might suppose that the disconnect is not so common. Minimally clever agents often realize good ends they want and act for the sake of, such as health and wealth. So, we need to understand better why Socrates might think such failures are common. (Here we are concerned with the general point; as promised before, we return to orators and tyrants in 4.5.)

Socrates does not make this fully explicit, but the answer lies in his axiological division, and more specifically in his account of intermediates as "sometimes good and sometimes bad". This implies that goods cannot be sometimes good and sometimes bad. In other words, on his taxonomy the only goods are *unconditional* goods. However, we know from elsewhere that Socrates thinks health and wealth are mere conditional goods; when guided by folly, they are harmful (*Euthyd.* 280b–282a; *M.* 87e–88c). He introduces the category of goods using examples that include health and wealth, but again, we think he does this to make his claim accessible to Polus and to common sense (cf. *Euthyd.* 279a–c; *M.* 87e). His principle implies that someone who sails for the sake of wealth does *not* want wealth after all. Wealth is a mere intermediate, sometimes good and sometimes bad. The agent must ultimately seek some further, unconditionally good end.[9]

This clarifies why we often fail to do what we want even while doing what we think best. Agents often achieve proximal, intermediate aims of their actions without achieving anything unconditionally good. But the explanation comes at a significant cost. It seems plausible to say that when an agent acts they aim at something good by their own lights, whether conditionally or unconditionally.[10] It is harder to argue that agents always

[9] More strictly, one should distinguish "pure goods", which never harm but may fail to benefit, from "unconditional goods", which always benefit. We suppose that Socrates thinks wisdom is purely but not unconditionally good in this strict sense, while wise activity is unconditionally good. However, the distinction does not much matter for present purposes, since Socrates clearly thinks that orators and tyrants fail to satisfy both their desires for merely pure goods like wisdom and their desires for strictly unconditional goods, such as wise activity.

[10] The contemporary "guise of the good" literature focuses on this weaker thesis. In favor, see Anscombe (2000); Davidson (1980); and Stampe (1987); against, see Velleman (2000). For a

aim at something *unconditionally* good by their own lights. Further, the more we stretch to attribute such desires to every agent from their own perspective, the harder it becomes to tie up the other loose end: we already needed to explain how Socrates secures the claim that agents aim at ends that are *in fact* good (and not just good by their lights). Now we need to explain how he secures the claim that agents aim at ends that are *in fact* unconditionally good (and not just unconditionally good by their lights). This desire is harder to establish starting from each agent's own perspective, howsoever corrupt.

4.4 Wanting as Reflective Desire

Socrates atributes to every agent, in every action, a want for something unconditionally good by their own lights; he further thinks that the object of such a want is in fact unconditionally good for them. In this section, we aim to show how he can satisfy these subjective and objective constraints on wanting through a notion of wanting as reflective desire. We also give an example of the sort of desire he attributes in this way.

We proceed in part by comparing Socrates' attribution of wants to orators and tyrants with his attribution of beliefs to Polus in the same passage. As we saw earlier, when Socrates first says that orators and tyrants do what they think best but not what they want, Polus asks whether doing what one thinks best is not great power. Socrates responds "Polus says it is not" (466e4), but Polus splutters that he says it is. Why does Socrates attribute this belief to Polus that Polus disavows – and more generally, why does he attribute beliefs to agents who reject his belief-attributions?[11]

Recent scholarship has made headway on this question about belief.[12] Socrates often finds that all who deny a given claim *p* can be shown that *p* follows from beliefs they already hold more deeply than their denial of *p*. In such cases, he expects the pattern to continue. Here, he expects that if Polus reflects on everything he thinks and reasons towards coherence, he will deny that orators and tyrants do what they want. That is, he attributes beliefs not based on how things seem to Polus unreflectively, but based on how things would seem to him after reflection on everything he already thinks. Attributing beliefs this way is not unique to Socrates within the

collection of essays, see Tenenbaum (2010). See also n. 17 on whether Socrates thinks all desires satisfy the stronger thesis, or only the special category of wants.

[11] See also 495e1–2, and cf. *Ap.* 26e–27a, *Ion* 539e, and *Symp.* 202c.

[12] See Vlastos (1983); even if his particular account is flawed, we think something broadly like it must be right.

Gorgias; Polus and Callicles do much the same to Socrates (461b3–4, 471e1, 474b; 481b6–7, 495a8–b1, d6–7, 521c3–4). Nor is this practice strange more generally. If a friend heaps scorn on a long-cherished colleague, we might say "You do not really think that!" In such cases, we say that our friend believes not what they say unreflectively, but what they would say on fuller reflection. This belief is not merely attributable to them after they reflect; they already had the attitudes within them that would lead them, upon reflection, to reject how things recently seemed to them in the moment. In that moment, our friend thought they believed something they did not. It might be odd always to attribute beliefs this way, but Socrates does not do that (472d1–4, 473a2–b6, 476a3–6). Rather he does so on particular occasions, for particular reasons. In our passage, among other things, he provokes Polus to examine himself.

More generally, while we have beliefs of many kinds – some true, some false; some close to awareness, some not – the key distinction for Socrates is between unreflective, unexamined beliefs and intelligent, examined beliefs. You might think you believe that *p* simply because it seems to you that *p*. But if *p* conflicts with other things you believe more deeply, you may come to see that *p* is not something you believe reflectively. Socrates constantly urges people to examine themselves, and so to discard some unreflective beliefs and transform others into intelligent, examined beliefs. Further, he consistently finds that certain beliefs lead people to contradict themselves. They find these beliefs untenable on reflection. Everyone whom Socrates has examined has a fund of beliefs that conflict with the untenable beliefs, and he extends this observed regularity by repeatedly appealing to beliefs from the common fund and bringing his interlocutors to reject the untenable ones. So, he expects that his interlocutors will, on reflection, reject their untenable beliefs. This focuses on untenable beliefs, since Socrates' examinations so often uncover beliefs to be discarded. But he is just as confident that on reflection, everyone will retain the common-fund beliefs, and can be brought to endorse those entailed by the common fund. So, he not only expects that his interlocutors will, on reflection, reject certain beliefs, but also that they will accept certain others.

Socrates does not question the status of the common-fund beliefs he finds in everyone, nor does he ask how they come to share these beliefs. He simply presupposes that any true belief must cohere with beliefs in the common fund. This presupposition guides his examinations and underpins his confident assertions about truth and falsity in the *Gorgias* (473b10–11; 508e6–509a4 with 479e8). There is nothing clearly wrong with this. We might reasonably press further, of course, even if Socrates

insisted that at least some common-fund beliefs are indubitable.[13] But if Socrates could justify his confidence in the common-fund beliefs, and if (as seems plausible) those shared beliefs are sufficiently rich and diverse, then Socrates would be right to suppose that anyone whose every belief coheres with the common-fund beliefs has only true beliefs. It is not too much to think that one's genuine beliefs – those that are tenable on reflection because they cohere with one's other beliefs, including the common-fund beliefs – are *true*.

This account of Socrates' attribution of beliefs to Polus (and others) has three important features. First, he is working with a plausible distinction between how things seem and what one truly believes: the former can be completely unreflective and one-off, while the latter must take account of the full range of how things seem to the subject. Second, and as a result, attributions of such beliefs are not made in isolation from the subject's own subjectively accessible attitudes; they must attend to how things seem from the subject's own perspective. Third, it is not unreasonable for Socrates to suppose that these subjectively motivated belief-attributions also meet an objective constraint – that beliefs aptly attributed this way are true.

We aim to extend this account of Socratic belief-attribution to his desire-attributions (and in particular his want-attributions).[14] Indeed, Socrates, unlike Hume, does not distinguish beliefs from desires as if they were entirely different animals. He sometimes treats desires as affectively charged motivating beliefs about value. This is suggested by his use of phrases like *dokein*, *dokein beltion*, and *dokein beltiston* in our argument. Such language suggests a conative attitude that is just belief (*doxa*) about what is good, better, or best. If this were not a conative attitude, it would be odd for Polus to doubt whether there is a difference between wanting and thinking best.[15] So, we should expect the reflective/nonreflective distinction among beliefs to apply to desires as well.

First, desires come in various stripes. Each of us desires a wide variety of things, but some of our desires are unreflective, while others are intelligent and examined. Intelligent, examined desires are not hastily formed but

[13] Arguably, the theory of recollection fills this gap for Plato (see *Phd.* 73a with Vlastos 1994, 28–29).

[14] For approaches that are similar in this regard, see Brickhouse and Smith (1994) and Kamtekar (2006).

[15] We prescind from the question of how *epithumiai* fit into Socrates' larger taxonomy of conative attitudes (cf. 491d ff.), but we would anyway insist that they satisfy the weak "guise of the good" requirement.

take into account consistency with one's other desires. You may take yourself to want to do something simply because that action (or its projected results) seems best to you unreflectively. But if that action conflicts with other, deeper desires you already have, then you may come to see that you do not reflectively desire to do it. Second, Socrates has found that certain desires consistently lead their possessors to conflict with themselves. Their possessors find these desires untenable on reflection. Everyone that Socrates examines has a fund of beliefs and desires that conflict with the untenable desires, and he extends this observed regularity by repeatedly appealing to beliefs and desires from the common fund and bringing his interlocutors to reject the untenable desires. So, he comes to expect that his interlocutors will, on reflection, reject their untenable desires. Moreover, he comes to expect that they will, on reflection, retain the common-fund desires, and can be brought to endorse those entailed by the common fund.

This explains how Socrates secures both the subjective and objective sides of his notion of wanting. Attributions of wanting are grounded in the subject's own present attitudes. Aptly ascribing a want for X to someone requires that they *already* have desires such that, if they reasoned to psychological coherence, X would seem best to them. In other words, Socrates attributes *actual* wants to someone on the basis of what *counterfactually would* but does not now seem best to them, but where this counterfactual claim is grounded in actual, subjectively accessible attitudes the agent already harbors.[16] The same is true of Socrates' strange attributions of belief and the relationship of these to how things presently seem. In some cases, as when Socrates says that Polus does not believe that orators and tyrants have great power, he already has in hand the actual, immediately accessible attitudes that entail the belief he attributes. In other cases, as when he attributes beliefs and desires to people he has never met, only his inductive confidence grounds his attributions. This inductive confidence could in principle be undermined, of course; Socrates has not examined everyone. But that seems to us a strength of his view, enabling him to rebut worries that his notion of wanting serves merely as "an excuse for dictation".[17]

[16] Gulley (1965) claims that wanting is the desire one would have if one knew the good. This makes wanting a purely counterfactual matter, whereas Socrates ascribes actual wants to orators and tyrants. Here we show why Socrates is entitled to actual and not merely counterfactual want-attributions.

[17] Dodds *ad* 467c5–468e5, note (c). One might also wonder about how to understand thinking best. In particular, does Socrates think of every such desire as constitutively (if inchoately) aimed at some unconditional good for the agent? Or might he rather suppose that desires in general aim at some good for the agent, but not necessarily an unconditional one, and that aiming at unconditional goods is distinctive of wants? We favor the latter view, not least because Socrates says that orators

Once we accept the idea of attributing desires based on how things seem to the agent upon reflection, Socrates is off and running. He finds that on even minimal reflection, everyone avows that they want to live well, to be successful or happy (*eudaimôn*), and that everyone takes themselves to aim at living well in their actions (cp. *Euthyd.* 278e–279a, *Symp.* 204e–205a). This is no mere conditional aim of their actions, sometimes good and sometimes bad. Rather, Socrates finds that upon reflection, everyone avows a desire for happiness *without qualification*, and that they condition their other, narrower desires on this one. This is not to say that every agent is always consciously aware of seeking the unconditional aim of living well. Rather, we have this aim and would acknowledge it on minimal reflection. So, someone might banish for the sake of revenge, without thinking that revenge is unconditionally good for them, but they would on reflection acknowledge that they want success or happiness, both in this action and in general.

More particularly, and as we have seen, Gorgias and Polus so value oratory and the putative power it provides precisely because they think it enables its possessor to live well and successfully. So, Socrates is on fairly firm ground in attributing both to them and to orators and tyrants a strong version of the claim that power benefits its possessor – in particular, that power is unconditionally good for its possessor.[18] If success or happiness is an unconditional aim that everyone has on reflection (i.e., something they want), then it is also desirable for anyone to have the ability to act in ways that satisfy that desire. This is power worth having, and if we were assessing ways of life, we might well consider whether people living that way can realize their goal of success or happiness (*eudaimonia*). Hence, Socrates can reasonably attribute to every agent a want for something

and tyrants manage to do what seems best to them, so that on his view they manage to satisfy certain desires (though not their wants).

This is related to a problem for Penner's defense of Socrates. Briefly, Penner thinks the claim that wanting is for our real good reflects a wider rejection of Fregean psychological states – that is states whose identity is independent of the realities at which they are directed. This may explain why orators and tyrants do not manage to do what they want (achieve their real good). But Penner faces a dilemma about thinking best. He could maintain that thinking best is also non-Fregean, but then it seems that by parity of reasoning, orators and tyrants should be unable to do what they think best. Alternatively, Penner could allow that thinking best is Fregean, but he then owes an explanation for why some psychological states are Fregean after all, while others are not. Penner (1988: 182, 190n33) seems to take the first arm of this dilemma; Penner and Rowe (1994: 9) seem to take the second arm.

[18] Kamtekar (2006: 135) notes this well. However, she takes 461b to suggest that Polus does not think orators must know the good. Polus is more committed to oratory's being fine than to its being an expertise (462c, 466a), but again, we think this is because he takes what is good to be obvious, not the object of specialized knowledge.

unconditionally good, and he can reasonably focus on the ability to satisfy this reflective desire.

This also helps us to understand how Socrates might endorse the same claim about power. It is true that he denigrates power elsewhere (e.g., *Euthyd.* 278e–281e). However, he can surely use 'power' in two senses. If we have a desire (i.e., wanting) whose object is unconditionally good, then ability to do what one wants will be unconditionally good. This does not require erasing the other kind of power (ability to do what seems best and to achieve conditional goods). Both senses of 'power' figure later in the *Gorgias* (esp. 509c–513c), strongly suggesting that Socrates distinguishes two kinds of power precisely because he distinguishes wanting from thinking best.[19] So, Socrates can accept the claim that power is (unconditionally) good for its possessor, and we should not suppose that his argument about power, wanting, and thinking best is narrowly *ad hominem*.

4.5 Orators and Tyrants

Socrates can thus reasonably attribute to each agent, in each action, a want for something unconditionally good for them, both by their own lights and in fact. He can also reasonably insist that doing what one wants (realizing such unconditional goods) is distinct from doing what one thinks best. As stated in 4.3:

> Whenever an agent acts, and they manage to do the intermediate action they think best, but they do not realize the good end for the sake of which they act, they do what they think best but they do not realize what they want.

Because this is a perfectly general claim about agents and actions, it applies to the agents and actions that most concern Socrates and Polus. To wit:

> Whenever orators and tyrants kill, exile, and confiscate, but they do not realize the good end for the sake of which they kill, exile, and confiscate, they do what they think best but they do not realize what they want.

Of course, orators and tyrants do not spend all of their time killing, exiling and confiscating; these actions are salient because Polus points to them as

[19] Contrast McTighe (1992); see also Dodds (1959) *ad* 510a3–5; Penner (1991: 154–155, 169); Wolfsdorf (2008: 116–117). Socrates does similar things elsewhere. For example, sometimes he calls courage unconditionally good (*G.* 507b4–8; *La.* 192c–199e; *Pr.* 359a–360e) and sometimes he calls it merely conditionally good, because it is bad when not guided by wisdom in addition (*Euthyd.* 281b4–e1; *M.* 88b3–6; *R.* 491b ff.).

examples of tyrannical power. Socrates' claim is more general: in these actions and others, orators and tyrants fail to do what they want. Further, his claim is quite strong: in their actions more generally, orators and tyrants do *least* of what they want and so have *least* power in their cities. However, Socrates has not yet given any clear sense of what specific wants orators and tyrants fail to satisfy in their actions, and why they *most* reliably fail to satisfy those wants.

To answer this question, we must look at the immediate aftermath of Socrates' argument about wanting. When he reaches the conditional claim about orators and tyrants,[20] Polus concedes it (468d1–e3), and Socrates relishes his concession (468e3–5). His revelry provokes Polus, who reverts to form (468e6–9): "As if you, Socrates, would not welcome being in a position to do whatever you see fit in the city, rather than not, and as if you would not feel envy whenever you saw someone putting to death some person he saw fit, or confiscating his property or tying him up".[21] In reply, Socrates drops a big hint about his reasons for supposing that orators and tyrants so reliably fail to do what they want. He simply asks (468e10), "Do you mean justly or unjustly?".

Socrates and Polus now clarify their disagreement over three theses (468e–474c). Socrates affirms and Polus denies (i) that it is worse to do than to suffer injustice (469b–c); (ii) that one cannot be unjust and happy (470e, 472d); and (iii) that if one acts unjustly, it is better to be correctly punished than not (472d–e, 473b–e). Polus sees that this conversation connects with what came before: when Socrates says that he would rather suffer than do injustice, he denies wanting to be a tyrant (469c3; cf. 510a–511a).[22] On Socrates' view, killing, exiling, and confiscating are in themselves intermediate, better only when done justly and worse when done unjustly (470c–471a). Nor is this merely his private view. True, Polus denies the theses that Socrates affirms. But Socrates thinks Polus is not expressing his real beliefs. He insists, "I think that I and you and the rest of humanity believe that doing what is unjust is worse than suffering it and that not paying the penalty is worse than paying it" (472b2–4, cf. 475e3–6). So, he examines Polus and shows that, upon reflection,

[20] Stauffer (2006: 52–53) rightly emphasizes the conditional phrasing, but wrongly concludes that Socrates' argument accomplishes relatively little.

[21] Notice that Polus here attributes desires to Socrates based on coherence with what he thinks everyone values, rather than based on how things seem to Socrates in the moment.

[22] This puts Polus a step ahead of McTighe (1992), who finds the shift to justice and injustice erratic. Stauffer (2006: 12, 55–56, and 138–140) rejects a version of the reading offered here, for unpersuasive reasons.

Socrates' theses cohere with Polus' own commitments better than Polus' denials do (474c–476a, 476a–481b). He thus extends his grounds for thinking that everyone wants to kill, exile, and confiscate only if they do so justly.

This clarifies the grounds for Socrates' claim about orators and tyrants – why he thinks they so regularly fail to satisfy their wants. Orators and tyrants, like everyone, want what would seem good for them on reflection – among other things, to act justly. One might wonder whether this characterization of their aims in acting is only available after extended reflection, or if some version is relatively close to the surface. Plato seems to suppose, in fact, that the unjust typically conceive of themselves as acting justly and that they care about this aspect of their actions. For example, Gorgias emphasizes that he teaches oratory so that his pupils can help their friends (456c7–457c3). And despite often being called an "amoralist", Callicles too values oratory in part so that he can help friends (486a2–3, b6–7).[23] But Socrates' deeper claim is that acting justly is no mere intermediate aim of their actions, sometimes good and sometimes bad. Like everyone, orators and tyrants agree on reflection to the theses that Polus unreflectively rejects, including the thesis that living well is impossible for the unjust. Hence, orators and tyrants want to act justly as an unconditional aim of their actions, and Socrates later characterizes the upshot of this discussion as an agreement that "no one does what's unjust because he wants to" (509e). This unconditional desire to act justly is not so close to the surface, but attribution of such a want to orators and tyrants is still grounded in Socrates' inductive confidence.

However, as Polus himself insists, orators and tyrants regularly act unjustly. In fact, they act unjustly and so do what they do not want more than others, because they have extraordinary means at their disposal to escape punishment (479c, 510e, 525d–526a). Most of us, if only for fear of punishment, avoid doing serious injustice. Orators and tyrants lack such constraints on their unreflective beliefs and desires. Hence, they do *almost nothing* they want and have *least* power in the city, even though they are most able to do intermediate actions such as killing, exiling, and confiscating (470b1–5) – indeed, at least in part *because* they are so free to do such actions. Further, because everyone shares the commitments about justice that underwrite Socrates' claims, orators and tyrants are held in no regard in their cities (cf. 466b) – that is, in no *reflective* regard.

[23] See further Shaw (2015: sections 4.2.1, 7.1).

The point that orators and tyrants want to act justly but fail to do so also connects with the claim that orators lack intelligence. In the setup (4.2), Socrates said not only that oratory is not an expertise (*technê*), but more specifically that oratory imitates a particular expertise, judicial expertise (*dikastikê*), which concerns justice (*dikaiosunê*) (462d6–9; 464c5–d4; 465c1–5). So, orators *qua* orators lack intelligence about justice. If orators were intelligent about justice – if they were real experts – they would have consistent views about justice. But Socrates insists that nobody has consistent views about justice unless they deem it unconditionally good for the just person. Hence, expertise about justice entails thinking it best to act justly. Experts about justice are able to satisfy this desire expertly. So, experts about justice always act justly. And finally, if someone who always acts justly is just, then experts about justice are just people. This is what Socrates told Gorgias earlier, in reverse order: an expert in justice will be just, act justly, and want to act justly (460b6–c2). Had Polus earlier insisted the orators have intelligence in this sense, this would simply have left him needing to respond directly to Socrates' refutation of Gorgias.[24]

One loose end remains. We have seen why Socrates thinks that the objects of coherent desire in general are genuinely, unconditionally good for the subject. However, his conversation with Callicles develops this idea further with respect to the specific claim that justice is genuinely and unconditionally good for us. Psychological coherence looms large in this later conversation, right from the start. When Callicles butts in to reject the claims that Socrates has argued for with Polus, Socrates insists that his examined life supports those claims. He stresses the value of psychological coherence and claims that, thanks to his love of philosophy, he is psychologically coherent, while Callicles is not. He thus lays a challenge at Callicles' feet: he should either refute Socrates or remain incoherent and unharmonious for his whole life (482a–c).[25]

Later, Socrates explains the value of psychological coherence or harmony, and puts a new twist on the relationships among coherent desire,

[24] Some scholars (e.g., Cooper 1999) criticize the epagogic argument at 460a–c. As we see it, Socrates' discussion with Polus begins to fill the gap in the argument. The *epagogê* fails to fully justify its conclusion because the doctor need not act healthily, nor the musician musically, as the person who knows justice must act justly. That is because health and musicality are mere conditional goods, whereas justice is unconditionally good (as an expert on justice would know). See *Hippias Major* and *Republic* I for rejected inductions to the claim that the just person can do the greatest injustice, and cf. Penner (1988: App. IV).

[25] See further Woolf (2000) on how Plato portrays Callicles as inconsistent right from the start.

justice, and the good. So far, again, we have seen that coherent desire tracks the real good, and that justice is among the things a coherent soul desires (and so among the things that are genuinely, unconditionally good for us). Socrates' later account draws these still closer together: the good *is* psychological order and organization (i.e., coherence), and psychological order and organization *are* the virtues of justice and temperance.

Socrates argues as follows: first, he draws an analogy between craftsmen generally and the good orator, who improves people's souls. (As the Gorgianic orator is the ersatz judge, so the true orator is what the Gorgianic orator merely pretends to be – viz., a true judge.) Craftsmen in general fit together their materials and impose order and organization on the whole that is their product. The true orator imbues the souls of their audience with order and organization, which are temperance and justice (503d–504e). Socrates then pushes on to a general account of goodness that substantiates his claim that the human good is order and organization of the soul, not pleasure (506c–507a). On the assumption that the elements ordered and organized in the just and temperate soul are psychological attitudes,[26] then, Socrates thinks that psychological coherence simply *is* the human good, and the states and activities of a coherent soul are also good. These goods include the virtues,[27] each of which is some state of a coherent soul, and virtuous activities. And that completes the explanation for why what would seem best to orators on full Socratic reflection – what they want, including justice and just activity – is in fact good for them.[28]

We do not appeal to these developments to give them a full reckoning. That would take far more attention than we can give here. Here, we only want to support our interpretation of Socrates' notion of wanting and his reasons for thinking that orators and tyrants do least of what they want,

[26] Some, drawing on 493a–494a and its similarity to the divided soul of the *Republic*, might suppose instead that order and organization of the soul are proper relations among its parts. We suppose that orderly relations among parts entail orderly relations among attitudes, so that for present purposes the difference would not much matter.

[27] Socrates focuses on justice and temperance because Callicles rejects these virtues, even though he accepts practical wisdom and courage (491d–492c). Of course, Socrates agrees that practical wisdom and courage are parts of the human good (for the former, recall Socrates' emphasis on intelligence and craft; for the latter, see 507b). For his description of virtue as 'lawfulness' and 'law', as against Callicles' rejection of convention, see Brown (2009a: 345).

[28] Some, inspired by Sachs (1963), might question the relationship between activities of an orderly and organized soul, so described, and pre-theoretical notions of temperate and just activity. We see the moderately revisionary aspirations of Socrates' view as acceptable and even admirable, but again we cannot address the issue fully here.

and to indicate the central role of these claims in Socrates' overall project in the *Gorgias*. He maintains that we want all and only what would seem best to us if our desires were perfectly coherent, that what we want is really good for us, and finally, that what we ultimately want is precisely psychological coherence. Properly understood, Socrates' exchange with Polus points forward to his rich and provocative account of the human good.