

**SOCRATES AND COHERENT DESIRE
(GORGIAS 466b-468e)**

ERIC BROWN and J. CLERK SHAW
Department of Philosophy
Washington University in St. Louis
eabrown@wustl.edu and jcshaw@artsci.wustl.edu

1. Introduction (466b9-466e5)

In Plato's Gorgias, Socrates makes some strange claims during his conversation with Polus.¹ The conversation begins with an apparent standoff. Socrates asserts that orators have "the least power in the city" (466b9-10), and Polus jealously defends the orators' capacity to act as tyrants, to "put to death anyone they want and confiscate property and banish from their cities anyone they see fit" (466b11-c2). But Socrates believes that there is no real disagreement. He acknowledges the orators' ability to do tyrannical things (466c9-d3), but he nevertheless maintains that they have the least power (466d6-8). In support of his counter-intuitive position, Socrates offers a strange claim about desire: tyrants and orators do whatever seems best to them but almost nothing that they want (466d8-e2).² When the confused Polus asks whether the ability to do whatever seems best is not great power, Socrates answers with a strange claim about belief: "No, at least Polus says it is not" (466e4). At this point, Polus can only sputter, "I say it isn't? I certainly say it is!" (466e5). Readers are often just as confused. Why is Socrates confident that Polus

We are cognizant of some gaps in our consideration of the secondary literature (not to mention the gaps in our argument!). We have not yet taken the full measure of Wolfsdorf's forthcoming article and Stauffer's recent book, nor have we revisited the full mass of material which shaped earlier stages of our work. We regret these gaps, and hope to correct them soon.

¹ Unless otherwise noted, we cite the Gorgias of Dodds' edition, and our translations follow (with only occasional and slight alterations) Zeyl's rendering, as printed in Cooper's edition of Plato's Complete Works. (Here and throughout the notes, we cite modern sources by the author's name and a (sometimes truncated) title; further details follow the paper.)

² For 'almost nothing', see Dodds ad 450b7 and Burnet ad Ap. 17a4 and b8 on hôs epos eipein. Socrates does not deny that orators and tyrants occasionally do what is in fact good for them, contra Penner, "Desire and Power," nn. 10 and 18. Socrates says that orators and tyrants get nothing good from their ability to do what seems best to them (touto, 467a3), not that they never get anything good by accident, say. Though Penner is surely correct to say that tyrants and orators cannot luck into success in life generally, this does not prevent them from occasionally doing or getting something good. If no one who lacks knowledge ever did anything he wanted, then tyrants and orators would be no less powerful than everybody else who lacks knowledge—that is, so far as Socrates can tell, everybody—and would be least powerful only in the sense that they are isodynamic with most or all people. In fact, however, he has good reason for thinking that tyrants and orators not only lack power but are least powerful, as we show in §4 below.

speaks his mind but not what he really believes? And why is Socrates so sure that the orators and tyrants do what seems best to them but not what they really want?

Recent scholarship has made significant headway on the first of these questions.³ In brief, Socrates' experience examining others sometimes underwrites his confidence that a given claim p follows from beliefs that his interlocutor holds more deeply than he holds his belief that not- p . So, based on his experience examining others, Socrates believes that if Polus were to reflect on all of his beliefs and reason toward the most coherent set that could be constructed of what he takes himself to believe, he would endorse the thesis that tyrants and orators do not do what they want and reject the opposite thesis. With this confidence, Socrates attributes to Polus not the beliefs Polus unreflectively professes but those he would accept on full reflection. This way of attributing beliefs is perfectly intelligible, and not even that unusual in more modest forms. "You do not really believe that," one might say to a friend who is swept up in the moment and saying things "out of character," as one might describe things. It would be odd to attribute beliefs this way all the time, but Socrates does not do that.⁴ Rather, he uses this special way of attributing beliefs on particular occasions, for particular reasons.⁵ Here, for example, he is trying to provoke Polus to some self-examination.

But our task in this essay is not to explore Socrates' approach to belief. Instead, we address the question about desire, for it seems to us that this question has not received a satisfying answer. To substantiate this judgment, we critically survey the existing answers in the next section. Then we proceed to a new account that is animated by the hypothesis that Socrates' curious claim about desire is just the other side of his curious claim about belief.⁶ That is, Socrates attributes only those desires to tyrants and orators that the

³ Thanks especially to Vlastos, "The Problem of the Elenchus." There are reasons to resist Vlastos' particular formulation of the solution, but we see no good reason to refuse something broadly like it.

⁴ For ordinary belief attributions in the immediate vicinity, see 472d1-4, 473a2-b6, 476a3-6.

⁵ See also 495e1-2. Polus (461b3-4, 471e1, 474b) and Callicles (481b6-7, 495a8-b1, d6-7, 521c3-4) also exhibit a tendency to attribute to Socrates those beliefs that they take to be true. Elsewhere, Diotima expresses confidence about what Socrates believes (*Symp.* 202c), as does Socrates about Ion (*Ion* 539e). Compare, too, Socrates' suggestion that Meletus cannot believe what he says and must be jesting (*Ap.* 26e-27a; cp. 481b6-7). See also Segvic, "No One Errs Willingly," 28-29.

⁶ This hypothesis is not original; see Brickhouse and Smith, *Plato's Socrates*, 73-102, esp. 88. But we go beyond Brickhouse and Smith in three main respects: first, we show why alternative construals of Socrates' argument at 466-468 are inadequate (see §2 below); second, we show how Socrates' argument at 466-468 is valid (see §3 below—Brickhouse and Smith do not much address the claims about action that are fundamental to Socrates' argument); and third, we do more to connect Socrates' position to other parts of his ethical view in the *Gorgias* (see §§4-5 below).

tyrants and orators themselves would accept upon fully reflecting toward the most coherent set that could be constructed of what seems best to them. We draw our interpretation out of the text of Gorgias 466-468 in two broad stages. First, in section three, we lay bare the structure of Socrates' argument to vindicate its validity. Then, in section four, we explore his grounds for the argument's most controversial premises. The principal upshot of our reading is quite simple: in the Gorgias, coherence is the key to the Socratic understanding of genuine wanting as justified desire.

2. Unsatisfactory Approaches

There have been several attempts to explain Socrates' claim that the orators and tyrants do not generally do what they want, but each comes with a significant cost.

The oldest and simplest approach is to say that Socrates is using 'want' in a special, restricted sense. One might say that Socrates is simply offering something like a "persuasive definition" of 'want' (boulesthai) according to which boulesthai has as its objects only what is in fact good for the wanter. The most immediate problem with this view is that it is unmotivated. Why should Polus (or anyone else) accept this restricted definition of boulesthai? At least since Plotinus, readers have been inclined to answer this question by insisting that boulesthai is restricted to a true self and the true self goes only for what is in fact good for it.⁷ But even if Socrates indulges in this way of talking,⁸ it can hardly solve the problem or persuade Polus. Why should anyone agree that we have a true self that goes only for what is in fact good?

Gulley has proposed to understand boulesthai as the desire that one would have if one knew the good.⁹ This reading removes any need for a subject of boulesthai other than an ordinary agent, and so requires no reference to a mysterious true self. (It does this by, in effect, identifying the true self as the ordinary agent plus knowledge of the good.) We believe that this approach has something right about it, but it fails to explain the nature of the knowledge in question and especially the relationship between wanting (boulesthai) what one knows to be good and going for what seems best to each agent. The latter point echoes the problem with the traditional approach: the restricted sense of boulesthai remains

⁷ For discussion of this view and references to those who hold it, see McTighe, "Socrates on Desire," 195-199 and nn. 8-19.

⁸ McTighe admits that Socrates says something like this in some of Plato's dialogues, e.g., the Phaedo; see also Brickhouse and Smith, Plato's Socrates §3.6. Alcibiades I, whether genuine or not, is another location for this sort of talk.

⁹ Gulley, "Interpretation," 89-90.

unmotivated. Why should Polus, or anyone, care about the desires he would have if he knew the good? These desires are, after all, merely hypothetical.¹⁰

More recently, Santas has offered a subjective motivation for the restriction on boulesthai.¹¹ According to this development of the traditional view, if an action has results that the agent recognizes as bad, then the agent did not want that action. This restriction is easy to motivate and requires no true self (because it appeals to things that the actual agent actually recognizes as good and bad), and it enjoys some textual plausibility.¹² Again, in some ways this is on the right track. Socrates does restrict the sense of boulesthai and he tries to motivate the restriction subjectively, based on what the agent himself takes to be good. But we believe that more needs to be said. Socrates gives an argument for his claim that tyrants and orators do not do what they want, and we suppose that this argument should illuminate and bolster his special sense of wanting.

Unfortunately, most of those who have carefully studied the argument do not have good news to report. The problem, put in technical jargon, is that Socrates either equivocates on de dicto and de re desire, or consistently speaks of desire de dicto but illicitly substitutes equivalents within an opaque context. Socrates argues that a tyrant who banishes his enemies but then finds that banishing his enemies was, in fact, bad for him (perhaps because it destabilized the city and unleashed more threatening opposition) did not in fact want (boulesthai) to banish his enemies. Either we take the tyrant's realization to be that he did not desire banishment under some further description (de dicto), for example, under the description that it destabilizes the city, or we take the tyrant's realization to be that he did not desire this actual banishment (de re). If the former, then Socrates' argument makes an invalid substitution. From the fact that the tyrant did not desire banishment as a destabilizer it does not follow that he did not desire banishment as, say, sweet revenge. Even though the same action might provide sweet revenge and destabilization, one can desire sweet revenge without desiring destabilization. If the latter, though, then Socrates argues fallaciously by changing the topic from desire for banishment

¹⁰ There is also a textual problem with Gulley's reading, as noted by, e.g., Kamtekar, "Conative Attitudes," **10 in ms**. Socrates says that the tyrants and orators actually want (boulesthai) what is good for them, not that they hypothetically do, even though the tyrants and orators do not know what is good for them.

¹¹ Santas, Philosophy of Socrates, 224-225.

¹² See 469d-470b, but as with several moments in this dialectical encounter, we take Socrates' example to be one that is easy for youthful Polus to understand, and not a representative example of Socrates' considered views (see §3.2). After all, the results to which Polus objects, and which Socrates uses to get him to recognize the general point about action, are punishments in return for injustice, and Socrates thinks that Polus would eventually realize that being punished correctly is better than escaping punishment for injustice.

as sweet revenge, which is desire under a particular description (de dicto), to desire for this actual banishment, which is desire de re.¹³

But this last problem arises only if we suppose that the tyrant's desire for banishment as revenge is desire for an object under a particular description (de dicto). If all wanting (boulesthai) were desire de re, for the objects themselves, then Socrates would not be fallaciously changing the topic by appealing to the tyrant's realization that he did not desire this actual banishment. In this way, Penner offers an interpretation of Socrates' argument according to which it is perfectly valid.¹⁴ The crux of his interpretation is that Socrates rejects what Penner calls "Fregean psychological states"—that is, psychological states whose identity conditions do not depend on the reality at which they are directed. Put another way, according to Penner's Socrates, all psychological states can be identified only by reference to the realities at which they are directed. So Penner construes Socrates' assertion that all desire aims at the good as the claim that all desire aims at the real final good, whereas what merely seems best to us are the means, instrumental or ingredient, that we propose to the real good. When one has a desire, then, the content of one's desire is not exhausted by what seems good—there are no Fregean psychological states—but includes a set of means-end relations that culminate in what actually is good. Hence, knowledge of what one desires requires knowledge of what is actually good and of the means-end relations that link the immediate content of the desire (e.g., to walk) to what is actually good. Unfortunately, agents commonly lack knowledge about the correct means to happiness, and so they fail to identify the action they desire to do, namely, the action that actually constitutes the best means to what is actually good. So agents commonly are stuck doing what seems best to them, though what seems best to them does not constitute means to what is actually good and so is not what the agents actually desire to do. Without

¹³ See especially McTighe, "Socrates on Desire," 205-207, and Waterfield, Gorgias, **pages**. Cf. Irwin, commentary, ad 468d; Santas, Socrates 223-225 with 186-189; and Vlastos, "Cheat?" 149-154. Irwin allows that Socrates' argument justifies a weaker conclusion that is sufficient to refute Polus, but he nevertheless thinks that the argument fails to establish the conclusion Socrates actually draws. Santas invokes a distinction between the intended object of (de dicto) desire and the actual object of (de re) desire to explain how Socrates could maintain that the tyrants and orators take revenge, *say*, to be the intended object of (de dicto) desire without taking banishment to be the actual object of (de re) desire. But this explanation does not give Socrates a valid argument for his conclusion that the tyrants and orators do not want the banishments they order; rather, it accentuates the difficulty Socrates is in. This difficulty becomes even clearer when Vlastos highlights the invalidity of substituting equivalents in an opaque context and uses the distinction between actual and intended objects of desire to argue that Socrates should not even accept his famous dictum that no one does wrong voluntarily. Irwin and Vlastos, in different ways, save Socrates only by changing his message, and this underscores the problems they find in his argument.

¹⁴ In "Desire and Power." Relevant material also appears in "Belief-Relative Sciences" and "Desire for Good."

knowledge of the good and the causal relations the good enters into, an agent is unable to do anything he desires; only with such knowledge will the agent have the power to choose the best means to the best end.

The problems with this approach are at least twofold.¹⁵ First, Penner's view requires an externalism about mental content so extreme that it is implausible that Socrates holds it.¹⁶ Second, Penner faces a dilemma. Either Socrates thinks that there are absolutely no Fregean psychological states, or he thinks that only wanting (boulesthai) has identity conditions that depend upon the real good at which it is directed. If the former, then no one can think that something is the best means toward some end unless it is in fact best. But in that case tyrants and orators cannot do what seems best to them by doing things that are in fact bad for them, even though Socrates happily grants them this ability (466e1-2, 467b3-5).¹⁷ Alternatively, if the content of what seems best does not depend upon what really is best, Penner needs to explain why some psychological states (e.g., what seems best) are Fregean while others (e.g., wanting [boulesthai]) are not.¹⁸ Thus, Penner's interpretation either contradicts the text and renders mysterious how Socrates might explain action (without reference to what seems best to people de dicto, which is how

¹⁵ We also reject Penner's representation of Socrates' argument. He thinks that Socrates' first formulation of the argument (466a4-467a10) is adequate so far as it goes, and runs like this: ("Desire and Power," 154): (i) Power is a good thing for its possessor; (ii) Orators and tyrants are unintelligent; (iii) To do what seems best without intelligence is to get what is bad for one; therefore, (iv) Orators and tyrants have no power. In addition to misstating the conclusion (see note 2 above), Penner finds no role for the distinction between what you want and what seems best to you, even though this figures prominently in the text of 466a4-467a10. Penner ("Desire and Power," 168-170) gives a slightly convoluted reason for this: that material is introduced only to give additional structure to the argument by justifying (iii) and especially (i). If this were so, we would expect to find Socrates concluding (i) at some point in 467c-468e, but (i) occurs only as a reiterated premise at 468e1-2.

¹⁶ We think that the considerations Penner raises in favor of the externalism (e.g. the advisory 'you don't want to do that' ["Desire and Power" 191], parents' concern for their children's happiness ["Desire and Power" 191], and the possibility of recognizing error ["Desire for Good," 7]) are not decisive, but this is not the place to reopen the debate about mental content. Our primary aim is to give an account according to which Socrates argues validly from premises that are generally recognized as plausible. We also think that the evidence Penner adduces to attribute to Socrates (and Plato) an antipathy toward "Fregean psychological states" is insufficient, but this is also not the place for a wide-ranging investigation into the Ion, Republic I, the Philebus, and Phaedrus. (Cf. Penner, "Belief-Relative Sciences" and The Ascent from Nominalism, passages indexed s.v. 'logic of psychological contexts'.) But note the second horn of the ensuing dilemma.

¹⁷ Penner appears to take this arm of the dilemma in "Desire and Power," given his phrasing of S1 on 182 and his reply to a referee on 190n33, but see the next note.

¹⁸ See "Desire for Good," 9, where Penner and Rowe write "for these sorts of cases, we shall resort to some variant of a Fregean approach." But this would appear to place Penner on the second arm of the dilemma, in which he admits Fregean psychological states. (We regret that we have not yet been able to take the full measure of Penner and Rowe's work on the Lysis, which might help to clarify Penner's reading of our passage.)

Socrates in fact does explain action [see 468b1-8 with 468d3-4]),¹⁹ or it, like the traditional interpretation, insufficiently explains why wanting (*boulesthai*) is special.²⁰

Thus, the existing accounts attribute to Socrates an invalid argument or premises that are unmotivated or implausible. Given this state of affairs, some readers have unsurprisingly decided that Socrates cannot mean what he says. They say that he is exaggerating his real position,²¹ or that he is merely arguing *ad hominem*.²² We reject these desperate maneuvers, because we feel no despair. We believe that Socrates argues validly from premises that are neither unmotivated nor implausible. Appreciating this argument will point the way to the grounds for Socrates' restricted account of wanting (*boulesthai*), which turns out to be restricted on both subjective and objective grounds.²³

¹⁹ It is not easy for externalists to explain action. For a contemporary attempt to reconcile the denial of narrow content with (scientific) intentional psychology, see Fodor, *Elm and Expert*, Ch. 2; for a reply to Fodor see Wakefield, "Broad vs. narrow content." Fodor's argument should be unacceptable to Penner, since one premise of his argument is that human action is generally successful (either Fodor is just less of a rigorist about living well than Socrates, or he thinks that desire does not aim at the good, or he thinks that there are no properties of good and bad and so no science of those properties, so that extensional truths about the good and the causal relations into which it enters do not form part of the content of desires).

²⁰ For further critical engagement with Penner's interpretation, see Kamtekar, "Conative Attitudes."

²¹ **Weiss.** Cf. Vlastos, "Cheat?" 149-154.

²² See McTighe. McTighe must brush aside a later passage in which Socrates refers back to his argument (509e5-7), and he must countenance a Socrates who cheats. He gives five main reasons for his defeatist reading, but we reject all of them. First, we think that none of the three problems McTighe finds in Socrates' argument applies: (1) there is no inconsistency among 467c5-7, 467d6-e1, and 468c3-4 (see our explanation in §3.2); (2) Socrates does not substitute illicitly into an opaque context (not, that is, in the argument we lay out in §3.2-4); (3) Socrates does not, as Waterfield (**page**), Irwin (ad 468ab), and Penner ("Desire and Power," 198-199) also charge, simply slide from the apparent good to the actual good to The Good: he has a principled reason for presupposing that what people seem to want on reflection is what they should want, as we suggest in §3.4 and justify in §4. Moreover, we think that neither of the oddities McTighe finds in *Gorgias* 466-468 obtain: (4) Socrates does not change the topic erratically either at 467a8 or at 468e10 (see §§3.1 and 4, respectively) and (5) does not contradict anything he elsewhere says about power (see the note in §3.1 below).

²³ Segvic, "No One Errs Willingly," esp. 5-16 and 40-45, also defends Socrates' position without attributing unmotivated or implausible premises to him, and we agree with her interpretation in many significant particulars. Still, we disagree about two important things. First, Segvic says that Socrates' crucial move is warranted only dialectically (44); on her view, Socrates has reasons for the move but does not deploy them in the argument. We think that this undersells what Socrates says in the argument. Second, drawing on Stampe, "The Authority of Desire," Segvic explains Socrates' reasons for the crucial move with an analogy between wanting and perception: just as perception requires not just a seeming but also an actual recognition (i.e., correct grasp) of how things are, so too Socratic wanting requires not just a seeming but an actual recognition (i.e., correct grasp) of what is good (Segvic does not insist that all desire is like this, and so avoids the problem we raise for Penner above). We believe that this misses the relevance of what is in the agent's mind to the distinction between mere desire and Socratic wanting, which is, in fact, exactly the resource in Socrates' argument that Segvic ignores. To reach these different conclusions, we proceed differently from Segvic. She goes after the larger quarry of Socratic intellectualism, and we limit ourselves to a more detailed reckoning of the *Gorgias* argument. We believe that patient exegesis pays philosophical dividends by revealing an even more interesting

3. A New Approach to a Valid Argument

3.1 The Set-Up (466e6-467c4)

We begin where Socrates does, to take a hint from the way in which he introduces his argument. After Socrates has made his curious claims about belief and desire and immediately after Polus has sputtered his outrage, Socrates puts his point in a nutshell (466e6-12):

By the...—you do not [believe that the tyrants and orators have great power], since you say that having great power is good for its possessor.

I do say that.

Then do you think that it is good if someone does whatever seems best to him when he does not have intelligence [nous]? And do you call this great power?

I do not.

Consistent with his earlier claim that oratory is not a real expertise (esp. 462b-466a), Socrates holds that orators do not have great power because they do not have intelligence. He thereby issues Polus a challenge: if Polus wants to show that orators do what they want, he should show that they have intelligence, that oratory is an expertise (466e13-467b2).

Polus does not rise to this challenge. Instead, he continues to sputter in amazement that Socrates would concede that tyrants and orators do what seems best to them but deny that they do what they want (467b1-10). Polus is in the grips of the ordinary assumption that doing what one wants simply is doing what seems best to one. He fails to realize that it might be reasonable to adopt a technical sense of doing what one wants. This is unfortunate, for Polus is in a pickle. He subscribes to four mutually inconsistent claims:

- (a) Power is the ability to do what one wants.
- (b) Doing what one wants is the same as doing what seems best to one.
- (c) Unintelligent agents sometimes harm themselves by doing what seems best to them.
- (d) Power is always good for its possessor.

Socratic intellectualism. Segvic also has a serious problem in that Socrates does two things: he attributes wanting in his special sense to everybody, and he insists that wanting, in this special sense, is not sufficient for correct action. Since Segvic's "Socratic wanting" presupposes a correct grasp of what is really good, it should be sufficient for correct action, but Socrates insists that it is not (509d7-e7). (We adapt this last point from Gulley.)

Given (a) and (b), Polus must agree that power is the ability to do what seems best, but then, given (c), he must concede that this power is sometimes not good for its possessor, which contradicts (d). At least one of these claims must go, but Polus only sputters.²⁴

Socrates therefore intervenes and offers to lead Polus through an argument that will show him why he should reject (b) (467b11-c4). This saves Polus from self-contradiction without directly answering the question whether orators have intelligence and therefore power. But Socrates has hardly changed the topic. In fact, throughout the ensuing argument (and see also 505a-b, 527d), Socrates maintains that doing what one wants just is doing what seems best for one if one is intelligent. Put another way, Socrates believes that what you want is what would seem best to you if you were reflecting properly on everything that seems best to you.

So we claim. We argue for this claim over the next three sections by first establishing the structure of Socrates' argument and then assessing his most controversial premises.

3.2 Stage One: An Analysis of Intentional Action (467c5-468b8)

To begin, Socrates insists that intentional actions are done "for the sake of something," and that agents do not want the actions themselves but that for the sake of which they do their actions. He introduces this principle slowly, and he puts it two different ways. First, he asks Polus, "Do you think that people want 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). This suggests the claim

- (1) If A does x, then A does x for the sake of some y, A does not want x, and A wants the y for the sake of which he does x.

²⁴ Notice that Socrates needs (d) to put pressure on Polus to surrender (b). Some scholars maintain that Socrates cannot mean to endorse (d) because he elsewhere denigrates power (e.g., Euthd 278e-281e). So McTighe (219-220 and 220n64) charges that Socrates is insincere, and Vlastos (Socrates, 149n74) insists that Socrates everywhere accepts only that power is conditionally good. But as Irwin points out (ad 466e), the argumentative context requires the stronger claim. Socrates has little reason to deny tyrants and orators the possession of mere conditional goods (see, e.g., 479c, 510e-511b, 513e-514a, 521b-c), but he still denies that they are powerful in our passage. We think that there is no conflict or insincerity because Socrates uses 'power' in two different senses. In fact, both senses of 'power' are prominent later in the Gorgias (particularly at 509c-513c). Compare the way Socrates treats 'courage'. Sometimes, he claims that courage is good for the one who has it (Gorg 507b4-8, Prot 349d-350c and 359a-360e, La 192c-199e), and sometimes, he claims that courage is actually bad unless one is wise in addition (Euthd 281b4-e1, esp. c6-7, Men 88b3-6, Rep 491b ff., Pol 306a ff.). In the former cases, Socrates assumes that courage is the same as (or at least a part of) wisdom; in the latter, he has in mind something like the natural boldness mentioned in Protagoras 349d-350c. Cf. Penner, "Desire and Power," 154-155 and 169.

In other words, taken at face value, Socrates' question suggests that a person cannot do an action for its own sake. But we should be careful, lest we forget whom Socrates is addressing. Socrates might believe that a person can do \underline{x} (exiling those people, say) for the sake of \underline{y} (restoring the health of the city) where \underline{y} (restoring the health of the city) is just another description of \underline{x} (exiling those people), and he might realize that this subtle point would only confuse the colt-like Polus. Socrates wants not only to offer a clever argument but to persuade Polus, and so he must approach his points in terms that Polus can understand. He must work toward his principle—whatever that principle turns out to be—from a simple starting point. So, to introduce the general idea that agents desire the ends for the sake of which they act, Socrates offers a blunt version of his intended principle, and he invokes examples in which the action and the end are distinct and the action is, by itself, undesirable (467c7-d5). This works. Polus readily understands the point of taking painful medicine for the sake of health, and by means of this example, he comes to see the general idea that intentional action aims at some end. But we are struck by the fact that once Socrates gets this general idea across, he never again insists that every action is distinct from its end, or that every action is, by itself, undesirable. Once they have served their pedagogical function, the blunt version of the principle and the blunt illustrations are dropped.

Socrates now aims for a subtler, more permissive articulation of his principle: "Is it not like this in every case? If someone does something for the sake of something [else], he does not want what he is doing, but that for the sake of which he is doing it?" (467d6-e1). This conditional commits Socrates only to

(2) If \underline{A} does \underline{x} for the sake of \underline{y} , then \underline{A} does not want \underline{x} and \underline{A} wants \underline{y} .

So understood, Socrates is silent about the possibility of doing an action for its own sake.²⁵ This must be intentional, and Socrates must mean to replace the misleading, introductory (1) with the subtler (2), because he clearly believes that one can do an action for its own sake. He later affirms that tyrants and orators do not want to, say, exile others simpliciter, but "want to do these things if they are beneficial" (468c3-4), and this says that a person

²⁵ Another interpretation arises if one narrowly construes the phrase 'like this' in Socrates' question, "Is it not like this in every case?" If Socrates means that his principle captures all of the features of the examples he has just articulated (painful medicine and the like), then one should take the ensuing conditional to be a mere approximation of the more accurate and restrictive version of the principle, namely, (1). We find this reading clearly inferior for two reasons. First, it is easier to take the phrase 'like this' loosely than it is to take the whole conditional loosely. Second, as we go on to say, Socrates later contradicts the restrictive version of the principle. The only thing to be explained, on our reading, is why Socrates starts with a restrictive version of the principle and with examples of that fit the restrictive version. But we take Socrates' pedagogical aims to explain this.

could want to do an action (and not just want that for the sake of which the action is done), which is inconsistent with (1) but consistent with (2).²⁶

Having induced Polus to accept the subtle (2), Socrates proceeds to justify it by locating it in a more general account of intentional action.²⁷ He first appeals to a tripartite classification of things in terms of value (467e1-4): some things are good, some are bad, and some are neither good nor bad (or, as we will say henceforth, intermediate). The intermediate things, Socrates explains, are those which are sometimes good and sometimes bad (467e6-468a4), and they include certain types of action, such as walking or making sea voyages, some of whose tokens are good and others bad. We suppose that Socrates wants us to generalize: every action is susceptible of being described in flat, neutral terms so that tokens of the flat, neutral action-type to which it is assigned are good in some circumstances, bad in others, and intermediate qua actions of that type. The generalization is controversial, perhaps, but not utterly implausible. Those action-descriptions that pick out good things ("helping friends") or bad things ("murdering") are, it might be supposed, not flat and neutral.²⁸

To complete this picture of intentional action, Socrates and Polus next agree that people do these intermediate actions for the sake of some good (468a5-b1). Indeed, they agree that people always act for the sake of what is good (468b1-8).²⁹ This clarifies (2):

²⁶ So not only should Socrates have (2) in mind rather than (1), as Irwin suggests ad 468bc, but he does. See also Vlastos, "Cheat?" 150n77, and Segvic, "No One Errs," 41-43.

²⁷ Contrast McTighe, 267. He finds reasons for (2) only in what precedes it, the induction from examples at 467c7-d6, and he condemns both the examples and the induction. Once we see the induction as Polus' way to the subtle claim (and not Socrates'), we are free to find Socrates' reasons for the claim in the articulation that follows (2). The asymmetry between Polus and Socrates and the conflict between (1) and (2) encourage the reading we take, but the best reason for our reading is that it attributes neither sloppy thinking nor deceptiveness to Socrates.

²⁸ Cf. Pausanias in Plato, Symposium 180e-181a and 183d. Also relevant are Socrates' rejection of "returning what is owed" as an account of what justice is, in Republic I and (perhaps) the contested argument concerning suicide at Phaedo 62a and following.

²⁹ This idea—for which, see also Meno 77b-78b and Anscombe, Intention, Davidson, Essays on Actions and Events, and Stampe, "The Authority of Desire"—can be doubted. Dodds, in note a ad 467c5-468e5, makes three objections. First, he says (following Murphy, Interpretation, 47) that if all actions aim at the good, then we will not be able to explain cases of psychological conflict such as that of Leontius (Rep 439e-440a). But one can be torn between two cases of action both of which seem, in different ways, to be good for one. Second, he argues (following Prichard, "Duty and Interest") that if all actions aim at the good of the agent, then actions motivated by disinterested love are impossible. But these are possible, understanding 'disinterested' in a certain way, since I may take pursuit of someone else's good to be constitutive or expressive of my own. Third, Dodds doubts "whether many of our actions are consciously governed by [a general conception of the agent's good]." But Plato is silent on whether actions are done under an occurrent, explicit judgment that the action is generally good for one; quite possibly, agents do not generally consider consciously what makes their actions desirable until others start asking them to justify their actions. For more searching objections, see Velleman, "The Guise of the

when one does \underline{x} for the sake of \underline{y} , one does not want the intermediate action \underline{x} , but one wants the good \underline{y} .

Unfortunately, there are limits to the clarity. In this stretch of the conversation, it is not decisively clear whether Socrates and Polus have an objective or subjective conception of goods in mind. Do people always act for the sake of what is actually good, or do they act for the sake of what they think is good? But things seem to be subjective, especially when Socrates asks, "And don't we also put a person to death, if we do, or banish him and confiscate his property because we think [*oiomenoi*] that doing that is better for us than not doing it?" (468b4-6, emphasis added; cf. *oiomenoi* at 468b2). We suppose that if Socrates also means that one always acts for the sake of what is actually good for one, he has to earn the point. Polus will be willing to agree, since he has in mind tyrants and orators and he takes them to be experts at living well.³⁰ But it would be premature to give Socrates the stronger claim at this stage.

It is important to notice that Socrates' picture of intentional action supports (2) by offering the resources to explain both sorts of intentional actions that (2) allows, those done for their own sake and those done for the sake of some consequence. On Socrates' picture, one always acts for the sake of something one takes to be good. We take this to require that one's action be done "under a description" that characterizes the action as in some way desirable. That is, intentional actions must allow for a reason to be given in response to the question "Why are you doing that?" and no response will offer a reason unless it identifies something the agent can intelligibly take to be good about the action.³¹ Now, in some cases, the neutrally described action is done for the sake of doing that same token action under a different description that characterizes it as desirable.³² For example, walking can

Good." We believe that these objections can be answered, and that in any case they are not so strong as to make Socrates' picture of intentional action obviously implausible.

³⁰ This point is well made by Kamtekar, **8 in ms**.

³¹ We here borrow some analysis from Anscombe, *Intention*.

³² Hence Dodds' objection in note b) *ad* 467c-468e misses the mark. Socrates recognizes that actions are intermediate only under certain (neutral) descriptions; all actions will admit of some description that renders the action intelligible by providing a desirability characteristic, whether that description mentions an effect of the actions or not. Dodds is also wrong to think that because "every event stands in a causal relationship to later events" there are no final ends; a mere causal result of *phi*-ing need not be an intentional result of *phi*-ing. Much the same can be said of Irwin's objection at 467cd and 467e-468a: it is not true that Socrates is talking purely about instrumental means-end relationships. (Irwin, *ad* 467cd, also takes *Lysis* 220ab to "rule out" "the possibility of wanting something both as a means and as an end." But see Lesses, "Plato's *Lysis*".) While much of Penner's language in "Belief-Relative Sciences" and "Desire and Power" encourages an instrumentalist reading (witness Segvic, "No One Errs" 12n15), it is clear in "Desire for the Good" 7n12 that he does not wish to restrict the means-end relations relevant to the science of happiness to instrumental means.

be done for the sake of getting exercise, and in these cases walking just is getting exercise, though the description 'getting exercise' reveals what is desirable about walking. Socrates recognizes these cases when he allows that one can do an action for its own sake. In other cases, the neutrally described action is done for the sake of some consequence to be brought about, where that consequence is conceived as good. For example, walking is sometimes done for the sake of getting to the festival, where arrival at the festival is the good to be brought about by walking. Socrates recognizes these cases when he first induces Polus to see that one can intentionally do an action without thinking that the action itself is desirable. In both sorts of case, we can say that the intentional action x is done for the sake of some y .³³

3.3 Stage Two: The Key Consequence (468b8-c8)

If we accept Socrates' picture of intentional action, at least for the sake of understanding his argument, then we should be prepared to draw a consequence. Recall that Socrates says that action x is done for the sake of y because x is intermediate and goods, not intermediates, are the objects of desire. From this it follows that

- (3) If A does action x for the sake of y but A 's doing x does not realize y , then A does not thereby realize what A wants.

We here use 'realize' ambiguously, in order to capture the fact that A might be doing x for the sake of doing good action y which is action x under a different description or A might be doing x for the sake of bringing about good consequence y . So on the one hand, if the action x fails to live up to its description as good action y , then doing action x fails to realize good action y for A . And on the other, if the action x fails to bring about the good consequence y , then doing action x fails to realize good consequence y for A . In either of these scenarios, A fails to get what A wants.

3.4 Stage Three: Tyrants and Orators (468b4-7, d1-e5)

Now all of the conceptual machinery is in place for Socrates to make his bold claims about tyrants and orators. From here, the argument is relatively simple. First, Socrates and Polus agree to several examples of what tyrants and orators can do: they can put

³³ Socrates does not consider actions done "for no particular reason," but it is available to him to say that they are done for the sake of satisfying a whim.

people to death, banish them, and confiscate their property. But these actions—call them \underline{x} , \underline{y} , and \underline{z} —in themselves are intermediate, sometimes good and sometimes not (470b). So

- (4) Tyrants and orators are able to do intermediate actions \underline{x} , \underline{y} , and \underline{z} .

Next, we apply to this fact Socrates' explanation of intentional action:

- (5) Therefore, when tyrants and orators do intermediate actions \underline{x} , \underline{y} , or \underline{z} , they do it for the sake of some good \underline{a} , \underline{b} , or \underline{c} .

Here, the premise is supposed to be neutral between the subjective and objective interpretations of the idea that people always act for the sake of the good. We will assume the subjective interpretation until there is warrant for the stronger reading. We can now draw the appropriate inferences:

- (6) Therefore, tyrants and orators do not want intermediate actions \underline{x} , \underline{y} , and \underline{z} .
 (7) Therefore, tyrants and orators want goods \underline{a} , \underline{b} , and \underline{c} .

From here, the conclusion we want is easily reached by means of a bold hypothesis:

- (8) When tyrants and orators do actions \underline{x} , \underline{y} , and \underline{z} , they do not realize goods \underline{a} , \underline{b} , and \underline{c} .
 (9) Therefore, when the tyrants and orators do actions \underline{x} , \underline{y} , and \underline{z} , they do not realize what they want.

The hypothesis and the ensuing conclusion explain the sense in which the tyrants and orators fail to have power: they lack power insofar as they fail to achieve what they want by their actions. Of course, Socrates and Polus focus on a narrow range of the tyrants' and orators' actions, but the argument generalizes easily. In fact, if the hypothesis holds of a larger range of the tyrants' and orators' actions than it does of anyone else's in the city, then the conclusion will be that the tyrants and orators have the least power of achieving what they want by their actions in the whole city. We will show in the next section why Socrates and Polus might see that the hypothesis does hold more broadly for the tyrants and orators than it would for anyone else in the city.

On the reconstruction presented here, Socrates' argument is perfectly valid; there are no substitutions into opaque contexts to be scorned as unwarranted or to be evaded via an implausible assumption about psychological attitudes.³⁴ But our interpretation attributes to Socrates a rather dubious hypothesis, (8). Why should we not suppose that the tyrant or orator exiles his enemy in order to exact revenge, and that in exiling his enemy he does in

³⁴ In fact, if we are right, Socrates must presume that substitutions into certain opaque contexts are not allowed. For instance, he must think it unallowable to infer, from the facts that \underline{A} desires to do \underline{y} unconditionally (where description \underline{y} carries a desirability characterization) and $\underline{y}=\underline{x}$, that \underline{A} desires to do \underline{x} unconditionally (where description \underline{x} offers only a neutral characterization).

fact exact revenge? Why suppose that tyrants and orators always aim at some good or goods that they fail to realize?

The first of these questions can be answered quickly if we complicate the argument a bit. So far, our version of Socrates' argument has supposed that one does an intermediate action for the sake of something one conceives to be good, but things are generally more complicated than that. If I walk for the sake of taking exercise, you might ask me why I am taking exercise. So it would seem that I walk for the sake of taking exercise, and I take exercise for the sake of my health. In this more complicated explanation, is taking exercise an intermediate or something I take to be good? The test is this: is taking exercise sometimes good and sometimes bad? Intermediate it is, we suppose. But then, health might also appear to be an intermediate. If so, you should ask me why I seek health, and you should continue to ask until my answer picks out an aim of my walking that is unconditionally good.

With this, Socrates' account for explaining intentional action begins to show its colors as an account of justifying intentional action. The complication does not entail that when I act, I act for the sake of what is actually good for me, but it does entail that when I act, I act for the sake of what I take to be unconditionally good for me. Of course, many agents might fail entirely to conceive of anything unconditionally good for them. We can still explain their actions in terms of the incomplete purposes they recognize, but we cannot justify their actions in those terms. To justify their actions, we would have to appeal to some unconditional good for them that they do not recognize. They might sometimes realize this good. Although perhaps they cannot intentionally realize it, there is such a thing as getting lucky.

Polus is in no position to reject the complication, as, again, he takes the tyrants and orators to be experts in living well. Surely, then, he thinks, when the tyrants and orators act, they realize what they take to be unconditionally good for them: they realize happiness and living well. Nor is it obvious that we should reject the complication. The idea that reasons for action are supposed both to explain and to justify action is plausible, as is the idea that justifying reasons need to bottom out in what is unconditionally valuable. At any rate, we have no qualms about attributing this to Socrates. It is enough to answer our first question. The tyrants and orators might execute, banish, or confiscate to exact revenge, but unless exacting revenge is unconditionally good, then the good at which they aim must lie deeper. In that case, even if they successfully exact revenge by executing, banishing, or confiscating, they might nevertheless fail to achieve the good for the sake of which they act.

But that leaves our second question. Why, again, does Socrates suppose that the tyrants and orators regularly fail to achieve the good for the sake of which they act?

4. What the Tyrants and Orators Want but Fail to Get

This is the crucial question for the argument, and the answer requires one look forward and another look back. Socrates has very general reasons for thinking that tyrants and orators ultimately want to act justly—this is an unconditional good that is an ultimate aim of their actions³⁵—and it is this want that they so often fail to satisfy by executing, banishing, and confiscating. So we argue.

Let us first look back to Socrates' introduction of the argument. According to the set-up, to understand why the tyrants and orators do not get the goods for the sake of which they act is to understand why they do not do what seems best with intelligence. But what does Socrates suppose would be added by intelligence (nous)?

Here we take a hint from Socrates' confidence about Polus' beliefs.³⁶ For now, two points are important. First, although a person has many beliefs of many different kinds—some true, some false, some close to awareness, some not—the distinction most crucial to Socrates is between unreflective, unexamined beliefs and intelligent, examined beliefs. Intelligent, examined beliefs are not hastily formed but instead take into account their consistency with one's other beliefs. So you may believe that p just because it seems to you that p. But if p turns out to conflict with other propositions that you already believe, you come to see that p is not something that you intelligently believe, on examination. Socrates is constantly asking people to examine themselves, and thereby to discard (or transform) unreflective beliefs in favor of intelligent, examined beliefs.

Second, in his examinations, Socrates finds that certain beliefs consistently lead their possessors to contradict themselves. These beliefs seem to their possessors to be untenable on reflection, and Socrates experiences this with remarkable consistency. Everyone Socrates examines has a fund of beliefs that conflict with the untenable beliefs, and Socrates extends this observation by regularly appealing to beliefs from the common fund in order to bring his interlocutors to reject the untenable beliefs. So Socrates comes to expect that his interlocutors will, on examination, reject these untenable beliefs.

³⁵ We say "an ultimate aim" here and at the end of this section to avoid prejudging the question of whether Socrates takes there to be more than one (and in particular, whether he takes it to be possible for one to value x for the sake of y and y for the sake of z where y is not an intermediate).

³⁶ We are still borrowing from Vlastos, "The Problem of the Elenchus."

The same two points hold for Socrates' examination of desires. Socrates, unlike Hume, does not distinguish between beliefs and desires as if they were two entirely different animals. Instead, Socrates treats desire as a motivating belief about value. So, let us apply the two steps. First, each of us desires an enormous variety of things, but some of our desires are unreflective (or undeliberate) and others are intelligent and examined (or deliberate). Both aim at some perceived good, based on one's subjective preferences and beliefs. But only the reflective desires cohere with one's other subjective preferences and beliefs. So, you may phi because it seems best to phi. But if phi-ing here and now turns out to conflict with other desires that you already have—perhaps you did not consider these other desires, or perhaps you failed to note that phi-ing would lead to consequences which happen to conflict with these other desires—then you will learn, perhaps painfully, that phi-ing was not something that you deliberately, intelligently desired to do.³⁷

Second, in his examinations, Socrates finds that certain desires consistently lead their possessors to be in conflict with themselves. These desires seem to their possessors to be untenable on reflection, and Socrates experiences this with remarkable consistency. Everyone Socrates examines has a fund of beliefs and desires that conflict with the untenable desires, and Socrates extends this observation by regularly appealing to beliefs and desires from the common fund in order to bring his interlocutors to reject the untenable desires. So Socrates comes to expect that his interlocutors will, on examination, reject these untenable desires.

We put these points in terms of untenable beliefs and desires, since Socrates' examinations so often uncover beliefs to be discarded. But his procedures and his search for examined, reflective beliefs and desires also lead him to recognize the common fund of beliefs and desires and those that entailed by the common fund. So he comes to expect not only that interlocutors will, on full examination, reject certain beliefs and desires, but also that they will, on full examination, accept certain others.

Against this background, suggested by Socratic practice and the opening moves in Socrates' conversation with Polus, to say that the tyrants and orators do not realize what

³⁷ So concerns like Wooten's (cited in Dodds' note c ad 467c-468e) that appeal to 'real' desire is really nothing more than "an excuse for dictation," are misplaced in this case, since each person's real desires are given by reflective, intelligent consideration of all the things that seem best to her: for this reason she is committed to caring about those "real desires," and in principle can be brought to do so starting from where she stands. If Socrates' inductive confidence were misplaced, he would be keen to find that out and open himself to the same process (458a), but in practice he regularly manages to produce his interlocutor as a witness (472b-c, 474a-b, 475e-476a) by appealing to the common fund of beliefs and desires. One might think that the methods by which Socrates works on the antecedent preferences of his interlocutors to produce revealed preferences are manipulative or mistaken, but the point would have to be argued.

they want when they put people to death, banish them, and confiscate their property is to say that these actions fail to cohere with certain other desires that the tyrants and orators have in doing them. But with what desires, exactly?

Here we need to look forward. At the end of our argument, when Socrates introduces his bold hypothesis, Polus concedes (468e2-3) that tyrants and orators do not do what they want and thus fail to have great power *if* what they do is actually bad for them (see 468d5-6). Socrates revels in his "I told you so!" moment (468e3-5), but Socrates has not explained why the tyrants' and orators' actions are actually bad for them. His revelry provokes Polus, who reverts to form (468e6-9): "As if you, Socrates, would not welcome being in a position to do what you see fit in the city, rather than not, 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."³⁸ In his reply, Socrates provides our second big clue. He simply asks (468e10), "Do you mean justly or unjustly?"

Socrates and Polus go on to clarify their disagreement over three central theses (468e-474c). Socrates affirms and Polus denies, first, that it is worse to do than to suffer what is unjust (esp. 469bc), second, that it is impossible to be unjust and happy (esp. 470e and 472d), and third, that if one is unjust, it is better to be punished than not (472de, 473b-e). Polus sees the connection between this conversation and what preceded: when Socrates says that he would rather suffer than do an injustice, he denies wanting to be a tyrant (469c3).³⁹ (From the start (461b-c), Polus had insisted that oratory was great because of the tyrannical injustice it was capable of. He thinks that Gorgias, too, should have been shameless enough to stand by this initial advertisement (452e, 456a-c).) But Socrates has a different point of view. He makes it perfectly plain that putting people to death, banishing them, and confiscating their property is better to do when one does them justly and worse to do when one does them unjustly (470c1-3).

Now, this is not merely Socrates' own private view. True, Polus denies the three theses that Socrates affirms. But Socrates thinks that Polus is not expressing his genuine beliefs. Socrates insists, "I think that I and you and the rest of humanity (tous allous anthropous) believe that doing what is unjust is worse than suffering it and that not paying the penalty is worse than paying it" (474b2-4, cf. 475e3-6). So he proceeds to examine Polus and to show that, on reflection, Socrates' affirmations cohere with Polus' commitments better than Polus' own denials do (474c-476a for the first thesis, 476a-481b

³⁸ Here, notice, it is Polus who attributes desires based on coherence with what he thinks everyone takes to be valuable. Cf. n. 5 above.

³⁹ This puts Polus one step ahead of McTighe, who finds the shift to justice and injustice at 468e10 to be erratic.

for the second and third). In doing this, he extends his grounds for holding that human beings want to execute, banish, and confiscate only if they are doing so justly.

This makes clear the grounds for Socrates' bold empirical claims about the tyrants and orators, why he thinks that they regularly fail to realize the goods for the sake of which they act. Tyrants and orators, like all human beings, really want what seems good for them on reflection, which is, among other things, to act justly. Nor is acting justly some intermediate aim of their actions, good sometimes and bad at other times. They, like all human beings, agree on reflection with the three theses that Polus initially rejects, and so they recognize that happiness (which people take to be unconditionally good for them) is impossible for someone who is unjust.⁴⁰ Rather, tyrants and orators really want acting justly as an ultimate end of their actions.⁴¹ But they regularly act unjustly. In fact, they act unjustly and fail to do what they want far more often than others because they have extraordinary means at their disposal to escape punishment and retaliation for their injustice (479c, 510e, 525d-526a). Most people, from fear of punishment if not from understanding what they really want, avoid doing serious injustice. Tyrants and orators do not. That is why they have the least power in the city.

5. Coherent Desire and the Good

There remains a loose end. We have argued that on Socrates' restricted account, one wants (boulesthai) what seems best to one on full (Socratic) examination. This understanding of wanting explains why he attributes wants to people in spite of their professed beliefs and desires. There is an important difference between our unreflective desires and the desires that optimally cohere with our commitments upon examination. Examined desires express what we really want because they agree with all our commitments. Unreflective desires frequently fail to express what we really want because they conflict with some of our other commitments. Hence Socrates moves from what seems best to us to what we want. Moreover, this understanding of wanting is the only

⁴⁰ Socrates explicitly states that all humanity agrees only to the first and third theses (474b2-4 and 475e3-6), and not to the thesis that happiness is impossible for one who is unjust. But the premises by which Socrates induces Polus to accept the second thesis strike us as bland or borrowed from the arguments by which he induces him to accept the first and third. So we cannot imagine that all humanity would accept the first and third theses on reflection but reject the second.

⁴¹ We also find support for this construal of the structure Socrates' argument in his retrospective characterization of it: "Do you think that Polus and I were or were not correct in being compelled to agree in our previous discussion when we agreed that no one does what's unjust because he wants to...?" (509e3-7).

understanding Socrates needs to argue that the tyrants and orators regularly fail to realize what they want. It is because acting justly would seem best to the tyrants and orators on full examination that they actually want to act justly, and this is what they regularly fail to do.

But this account of wanting does not establish that one wants only what is in fact good for one. There are indications that Socrates has this objective restriction in mind, as well.⁴² As we acknowledged earlier, Socrates seems to think that one acts always for the sake of what is really good for one. We insisted that he would need to earn this. Now we want to show how he does so, by extending the lessons of the previous section. Again, we need one look back and one forward.

Taking our cue from Socrates' suggestion that a person wants what seems best if one is intelligent, we have argued that Socrates sometimes attributes beliefs and wants based on a common fund of commitments that he has discovered through his examinations. When he makes attributions in this strange way, he says that one believes and desires what follows from one's common fund of commitments, on full examination. We believe that he has some reason to suppose that what one wants, on this conception of wanting, is also what is actually good for one.

Consider first the case of beliefs. Socrates does not question the status of common-fund beliefs, nor does he ask how the people he examines come to share them.⁴³ Instead, he simply presupposes that any true belief must cohere with beliefs in the common fund. This presupposition guides his examinations, and it underpins his confident assertions about truth and falsity in the Gorgias (473b10-11 and 508e6-509a4 with 479e8). There is nothing obviously wrong with this. We might reasonably press for more justification, of course, even if Socrates were to insist that at least some of the common-fund beliefs are indubitable.⁴⁴ But if Socrates could justify his confidence in common-fund beliefs, then, supposing (as is plausible) that these shared beliefs are sufficiently diverse and rich, Socrates would be right to think that anyone whose every belief cohered with his common-

⁴² See especially 468d3-4, 468d5-6, and 475e5-6, with Penner, "Desire and Power" 198-199; Segvic, "No One Errs," 43-44; Kamtekar, "Conative Attitudes," **4 in ms**. Kamtekar's excellent account of why Plato believes that desire is for what is actually good for one starts from this point that we are trying to reach. We do not think that we have to reject anything she says to insist that the subjective restriction on wanting (as what seems best to one on full examination) is also an important part of the story in the Gorgias.

⁴³ This is not entirely fair: see 481cd and Protagoras 348cd.

⁴⁴ We see no evidence that the Socrates of the Socratic dialogues felt this need, but it seems as though Plato and the Stoics, among other followers, did. Arguably, recollection provides Plato his first response (for this suggestion see Vlastos, "The Problem of the Elenchus" 28-29), and certainly, the kataleptic impression provides the Stoics with theirs.

fund beliefs had nothing but true beliefs. It is not too much to think that one's genuine beliefs—the beliefs that are tenable on reflection because they cohere with one's other beliefs, including one's common-fund beliefs—are true.

We suppose that similar reflections can bolster the wants (*boulêseis*). If desires aim at the good in the way that beliefs aim at the true, and if full Socratic examination of the common fund of commitments fixes the aim of belief on what is in fact true, then the same sort of examination should be able to fix the aim of desire on what is in fact good. To the extent that Socrates' confidence in the common-fund commitments is warranted, he can reasonably take beliefs that cohere with the common fund to track what is really true and desires that cohere to track what is really good. On this sort of reasoning, Socrates might plausibly suppose that what seems best to a person on full examination is in fact best. Thanks to the reliable common-fund commitments, his subjective restriction on wanting flowers into an objective restriction.

It should not be surprising that Socrates would think that the just actions that the tyrants and orators want are in fact good for them. Still, we want to look forward from his argument about tyrants and orators one more time to bolster these reflections, for Socrates' later conversation with Callicles develops the connection between coherent desire and the good.

The theme of psychological coherence dominates this later conversation. Socrates makes it explicit right at the start. Callicles butts in, to reject the theses that Socrates has argued for with Polus, but Socrates insists that his life of philosophical examinations supports what he has said fully. So he lays a challenge at Callicles' feet:

Believe, then, that you are bound to hear me say other such things, too, and do not be surprised that I say them, but stop my beloved, philosophy, from saying them. For she always says what you now hear me say, dear friend... So either refute her and show that doing what is unjust without paying what is due for it is not the ultimate of all evils, as I just now was saying it is, or else, if you leave this unrefuted, then by the Dog, the god of the Egyptians, Callicles will not agree with you, Callicles, but will be dissonant with you all your life long. And yet for my part, my good man, I think it is better to have my lyre or a chorus that I might lead out of tune and dissonant, and have most people disagree with me and contradict me, than to be out of harmony with myself, to contradict myself, though I am only one person. (482a2-5, b2-c3)

Socrates claims at least two things here. First, psychological coherence is tremendously important. Second, thanks to his love of philosophy, he possesses psychological coherence and Callicles does not.⁴⁵

Later, Socrates identifies psychological orderliness as the good for human beings. First, he draws an analogy between craftsmen generally and the man who possesses the genuine craft of rhetoric to illuminate the aim of the rhetorical craft (503d-504e). Craftsmen in general give their products some shape; they place what they do into a certain organization, fitting together its parts until the whole is put together in an orderly way. So the true rhetorician should imbue the soul of his audience with a certain organization and order, and these just are justice and temperance or self-control. Then he pushes still further, to offer a general account of goodness with which to substantiate his claim that the good for human beings is psychological order or harmony and not pleasure (506c-507a). He argues as follows:

- (1) The good for X is that whose presence in X makes X good. (506d1-2)
- (2) X is made good by the presence of excellence in X . (506d2-4)
- (3) Excellence comes to be present in X most finely by the organization, correctness, and craft bestowed on X . (506d5-8)
- (4) Therefore, the excellence of X comes to be something organized and ordered by organization. (506e1-2)
- (5) Therefore, when the order proper to X comes to be in X , X becomes good. (506e2-4)

From the first two premises, it follows that the excellence of a thing is its good. From the first and last claims, it follows that the order proper to a thing is its good. From these two implications, additional support for (4) follows: the excellence of a thing is its order or organization. So, Socrates infers, the good for a human soul, its orderliness, is temperance or self-control (506e-507a, cf. 503d-504d).⁴⁶

On Socrates' view, then, psychological coherence is the good for a human being, and the parts of a coherent soul, its states and activities, are also good. These goods include the virtues, each of which refers to the state of a coherent soul, and virtuous activities. And

⁴⁵ We are largely in agreement with Woolf, "Callicles and Socrates," that Plato portrays Callicles as inconsistent from the start. On Socrates' psychological coherence and on how it falls short of knowledge, see Shaw, "True Political Craft," §3.

⁴⁶ Socrates focuses on justice and temperance or self-control because Callicles rejects these cooperative virtues of restraint, although he acknowledges the value of practical wisdom and courage (491d-492c). We do not suppose that Socrates would deny that practical wisdom and courage are also parts of the human good.

that completes the explanation why that which seems best to the tyrants and orators on full examination—just activity—is also in fact good for them.

We appeal to these developments not to give them a full reckoning—that would require more attention than we give here—but to support our interpretation of Socrates' special sense of wanting in his conversation with Polus and to indicate its central role in Socrates' overall ethical project in the Gorgias. Socrates maintains that the special sense of wanting is reserved for coherent desires, whose satisfaction realizes what is genuinely good for an agent. Properly understood, Socrates' conversation with Polus points the way forward to his rich and provocative account of the good for a human being.

Sources Cited

- Anscombe, G.E.M. Intention. 2nd ed. Cambridge: Harvard University Press, [1963] 2000.
- Brickhouse, Thomas C., and Nicholas D. Smith. Plato's Socrates. Oxford: Oxford University Press, 1994.
- Burnet, John. Plato: Euthyphro, Apology of Socrates, Crito. Oxford: Clarendon Press, 1924.
- Cooper, John M. "Socrates and Plato in Plato's Gorgias." in Reason and Emotion: Essays on Ancient Moral Psychology and Ethical Theory, 29-75. Princeton: Princeton University Press.
- Cooper, John M., ed. Plato: Complete Works. Indianapolis: Hackett Publishing Company, 1997.
- Davidson, Donald. Essays on Actions and Events. Oxford: Clarendon Press, 1980.
- Dodds, E.R. Plato Gorgias. Revised text, with introduction and commentary. Oxford: Clarendon, 1959.
- Fodor, Jerry A. The Elm and the Expert: Mentalese and its Semantics. Cambridge: MIT Press, 1994.
- Gulley, Norman. "The Interpretation of 'No one does wrong willingly' in Plato's Dialogues." Phronesis 10 (1965): 82-96.
- Irwin, Terence. Plato Gorgias. Translated with notes. Oxford: Clarendon, 1979.
- Kamtekar, Rachana. "Plato on the Attribution of Conative Attitudes." Archiv für Geschichte der Philosophie 88 (2006): 127-162.
- Lesses, Glenn. "Plato's Lysis and Irwin's Socrates." International Studies in Philosophy 18 (1986): 33-43.
- Murphy, N. R. The Interpretation of Plato's Republic. Oxford: Clarendon Press, 1951.

- McTighe, Kevin. "Socrates on Desire for the Good and the Involuntariness of Wrongdoing: Gorgias 446a-468e." in Essays on the Philosophy of Socrates, ed. Hugh H. Benson (Oxford: Oxford University Press, 1992), 263-297. reprinted from Phronesis 29 (1984): 193-236.
- Penner, Terry. The Ascent from Nominalism. Dordrecht: Reidel Publishing Company, 1987.
- ". "Desire and Power in Socrates: The Argument of Gorgias 466a-468e that Orators and Tyrants Have No Power in the City." Apeiron 24 (1991): 147-202.
- ". "Socrates on the Impossibility of Belief-Relative Sciences." Proceedings of the Boston Area Colloquium in Ancient Philosophy 3 (1988): 263-325.
- Penner, Terry and Christopher Rowe. "The Desire for Good: Is the Meno Inconsistent with the Gorgias?" Phronesis 39 (1994): 1-25.
- Penner, Terry, and Christopher Rowe. Plato's Lysis. Cambridge: Cambridge University Press, 2005.
- Prichard, H.A. "Duty and Interest." in id., Moral Writings, ed. J. MacAdam (Oxford: Oxford University Press, 2002), 21-49. originally published 1928.
- Santas, G. X. Socrates: Philosophy in Plato's Early Dialogues. London: Routledge & Kegan Paul, 1979.
- Segvic, Heda. "No One Errs Willingly: The Meaning of Socratic Intellectualism." Oxford Studies in Ancient Philosophy 19 (2000): 1-45.
- Shaw, J. Clerk. "Socrates and the True Political Craft." forthcoming.
- Stampe, Dennis. "The Authority of Desire." Philosophical Review 96 (1987): 335-381.
- Stauffer, Devin. The Unity of Plato's Gorgias. Cambridge: Cambridge University Press, 2006.
- Velleman, J. David. "The Guise of the Good." in id., The Possibility of Practical Reason (Oxford: Oxford University Press, 2000), 99-122. reprinted from Nous 26 (1992): 3-26.
- Vlastos, Gregory. "Does Socrates Cheat?" in id., Socrates, Ironist and Moral Philosopher (Ithaca: Cornell University Press, 1991), 132-156.
- ". "The Socratic Elenchus." in id., Socratic Studies, ed. Myles Burnyeat (Cambridge: Cambridge University Press, 1994), 1-37. Revised and reprinted from Oxford Studies in Ancient Philosophy 1 (1983): 27-58.
- Wakefield, Jerome C. "Broad versus narrow content in the explanation of action: Fodor on Frege cases." Philosophical Psychology 15 (2002): 119-133.
- Waterfield. Plato's Gorgias. Translation with notes. Oxford: Oxford University Press, 1994.
- Weiss, Roslyn. "Killing, Confiscating, and Banishing at Gorgias 466-468." Ancient Philosophy 12 (1992): 299-315.
- Wolfsdorf, David. "Gorgias 466a4-468e5: Rhetoric's Inadequate Means." forthcoming.

Woolf, Raphael. "Callicles and Socrates: Psychic (Dis)harmony in the Gorgias." Oxford Studies in Ancient Philosophy 18, 1-40.