

Forthcoming in *Phronesis*.

Intrinsic Valuing and the Limits of Justice: Why the Ring of Gyges Matters

Tyler Paytas, Australian Catholic University

Nicholas R. Baima, Florida Atlantic University

Abstract: Commentators such as Terence Irwin (1999) and Christopher Shields (2006) claim that the Ring of Gyges argument in *Republic* II cannot demonstrate that justice is chosen only for its consequences. This is because valuing justice for its own sake is compatible with judging its value to be overridable. Through examination of the rational commitments involved in valuing normative ideals such as justice, we aim to show that this analysis is mistaken. If Glaucon is right that everyone would endorse Gyges' behavior, it follows that nobody values justice intrinsically. Hence, the Gyges story constitutes a more serious challenge than critics maintain.

Key Words: Ring of Gyges, Plato, Justice, Glaucon, *Republic* II

In *Republic* II, Glaucon recounts the story of Gyges' ring in order to bolster the case against justice. According to the story, a shepherd named Gyges utilized an invisibility ring to seduce the king's wife, kill the king, and take over the kingdom (360b).¹ Glaucon argues that anyone who came into possession of such a ring would perform similarly unjust acts, and that anyone who would refrain from injustice in such circumstances "would be thought wretched and stupid by everyone aware of the situation" (360d). Glaucon presents this as strong evidence that justice is thought to be valuable only to the extent that one lacks the power to get away with injustice (359b).

Glaucon's argument appears successful in showing that we do not value justice as much as Socrates believes we should. Yet we might still wonder whether it supports the stronger claim that we do not value justice for its own sake *at all*. Commentators such as Terence Irwin (1999) and Christopher Shields (2006) claim that Glaucon's argument could not show that justice is valued only

¹ Translations follow Reeve 2004, with some slight modification. The Greek is from Burnet 1900-1907.

for its consequences because believing that power and wealth outweigh justice in some circumstances is compatible with intrinsically valuing justice to some overridable degree. Let us call this claim the *compatibility thesis*.

In this article, we argue that the compatibility thesis is false and hence that the soundness of the Gyges argument would indeed demonstrate that nobody values justice for its own sake. The upshot is that Glaucon's argument constitutes a more serious challenge to the purported value of justice than is often thought.

1. The Compatibility Thesis

In explicating the compatibility thesis, Irwin writes: "Gyges might suppose that justice has some slight intrinsic value in addition to the value that depends on its consequences, but he might take the good consequences of injustice to outweigh this intrinsic value of justice" (1999, 173). In a similar vein, Shields explains that someone's choosing to follow an unjust path cannot show that they don't value justice intrinsically because it may just be a matter of degree: "In making such a choice, we would no doubt consider the relevant degrees and thresholds and make our decisions accordingly" (2006, 79). These claims certainly conform to ordinary experience; people regularly choose Y over X despite viewing X as intrinsically valuable. With this in mind, Irwin concludes that "Glaucon should ask not only whether we care about justice for its own sake, but also how much we care about it" (1999, 173).

If Glaucon's argument cannot show that justice is valued only for its consequences, this raises questions about the point of its inclusion in Book II. Here one might follow Shields in viewing the Gyges story as of little utility for Glaucon's purposes but rightly included by Plato as a mechanism for causing readers to reflect on their level of commitment to justice (2006, 76-77). But Glaucon seemingly takes the thought experiment to have probative force in regards to his dialectical

aims, and so we should adopt this reading only after we've exhausted alternatives that would give the argument more significance.

Irwin proposes such an alternative. On his reading, what Glaucon aims to show is not that we value justice only instrumentally, but rather that we do not think justice is more valuable than injustice *in all circumstances*. Glaucon asks Socrates, "Do you want to seem to have persuaded us that it is better *in every way* to be just than unjust, or do you want to truly convince us of this?" (357a4-b2, emphasis added). Irwin suggests that we should understand "in every way" to mean "better in every possible circumstance" (1999, 175).

The appeal to the "in every way" clause provides a reasonable explanation for the role of the Gyges argument that meshes with Irwin's analysis. However, the context in which the argument is presented makes this interpretation rather surprising. Note that just before recounting the story, Glaucon provides a threefold classification of goods and claims that most everyone takes justice to be a type-3 good (i.e. burdensome in itself and pursued for the sake of its consequences) (358a).² And shortly thereafter, Glaucon prefaces the Gyges story by claiming that it will allow us to see most clearly that those who practice justice do so only out of necessity (359b6-d1). This suggests that Glaucon's aim is to show that nobody values justice for its own sake. Hence, we have good reason to subject the compatibility thesis to further scrutiny.

² Although the division of goods is a source of interpretive controversy, we shall focus primarily on exploring the rational commitments involved in valuing justice for its own sake. Investigation of this issue does not require taking a stance on whether intrinsic valuing means 'independently of consequences' (Kirwan 1965; Mabbott 1978; Reeve 1988) or 'because of the attached (rather than ancillary) consequences' (Foster 1937; Annas 1981; White 1984; Pappas 1995). For recent discussion, see Heinaman 2002 and Payne 2011.

2. Normative Objects and the Limiting Requirement

At first blush, the claim that valuing justice for its own sake is compatible with believing that considerations of justice can be offset by the allure of wealth and power seems plausible. Viewing something as choiceworthy in and of itself does not commit one to valuing that thing above all else. Nonetheless, if we look closely at the ring of Gyges and what it asks us to consider, this analysis begins to appear less plausible. Consider first that the sincerity of our evaluative judgments is contingent upon our behaviors, dispositions, and the related judgements that we make. This is seen most readily when the object of our valuing is itself partly constituted by ethical norms or standards. Let us call the various ethical principles, moral ideals, and virtues that one might value *normative objects*. If an agent genuinely values a normative object for its own sake (as opposed to merely feigning to do so), then she will not reflectively endorse behaviors and attitudes that directly violate the principles and standards of which it is composed. This is because intrinsically valuing a normative object involves endorsement of the constitutive principles and standards.

To illustrate, consider the case of a hiring manager who claims to value racial equality for its own sake. Suppose this manager consistently and intentionally favors job applications from members of her own race even when they are less qualified than other applicants. When confronted about her hiring practices the manager says, “I do care about racial equality for its own sake, but I also believe that whenever someone has a chance to help a member of their own race, that’s the right thing to do.” Valuing the ideal of racial equality for its own sake is incompatible with endorsing overtly racist attitudes and behaviors. Hence, we can conclude that the hiring manager either does not know what racial equality is, or else her claim to value it intrinsically is mere lip service.

Be that as it may, exceptional circumstances can make it possible for a person to value a normative object intrinsically while still acting in ways that might undermine it. For example, a hiring manager can intrinsically value racial equality while knowingly participating in a racist hiring practice.

It might be the case that a racist CEO threatens to fire the hiring manager if she does not hire persons of a certain race. The hiring manager may sincerely want to refrain from this practice while also knowing that if she does not comply she will lose her job and health insurance, thus having no means to care for her sickly daughter. If the hiring manager decides to follow the CEO's orders, it can still be true that she values racial equality for its own sake; it may just be that she values the life and health of her daughter even more.

The crucial point is that the circumstances that render the flouting of a normative object compatible with valuing it for its own sake cannot be constituted by the very considerations that the normative object precludes from having justificatory weight. The reason it is certain that the first hiring manager does not value racial equality for its own sake is that the ideal of racial equality precludes the mere fact that another person is a member of one's race from justifying the preferential treatment of that person—this preclusion is an essential element of the normative object. In contrast, it might still be true that the hiring manager in the second scenario values racial equality for its own sake because the ideal of racial equality does not preclude a concern for the health and safety of one's child from justifying the preferential treatment of members of a certain race—such a preclusion is not an essential element of the normative object.

Let us turn then to the case of justice. As a normative object, justice is similar to racial equality in that genuinely valuing it for its own sake involves endorsing its essential principles and standards. Hence, valuing justice for its own sake is incompatible with rational endorsement of undermining behaviors when the ostensible justification for the undermining involves considerations that justice itself precludes from having justificatory weight. To see what the relevant considerations are, we need to look closely at what justice encompasses. Returning to the context of the *Republic*, we must remember that the discussants in Book II have yet to complete their investigation into the nature and value of justice. However, any such investigation must begin with

some basic assumptions. One reasonable assumption about justice is what we call the *Limiting Requirement*:

Limiting Requirement: Being a just person requires viewing the interests of others as placing limits on what one can permissibly do in pursuit of effective goods (e.g. wealth and political power).

Not only does the limiting requirement seem central to any plausible conception of justice, the context of the dialogue suggests that it is part of the conception of justice with which Glaucon is working. First, the fact that justice places limits on the pursuit of wealth and power explains why Glaucon has doubts about the value of justice in the first place. It is precisely because justice appears to conflict with self-interest in some circumstances that he implores Socrates to provide a rousing defense. Second, note that the idea expressed by the limiting requirement falls under the broad notion of fairness. A person who does not view the interests of others as placing limits on what can defensibly be done in pursuit of wealth and power is someone who places no weight on considerations of fairness. This is important because Glaucon invokes the notion of fairness when prefacing the Gyges argument in Book II. In explaining why the ostensibly just person will act exactly as the unjust person would in circumstances of immunity, Glaucon says, “The reason for this is the desire to outdo others and get more and more [*pleonxian*]. This is what every nature naturally pursues as good, but by law and force, it is made to deviate from this path and honor fairness [*paragetai epi tēn tou isou timēn*]” (359c). Here Glaucon uses “fairness” (*isos*) as interchangeable with “justice,” and the limiting requirement is clearly involved in any plausible notion of fairness.

Given the centrality of the limiting requirement to the very notion of justice, the analysis we are challenging would attribute the following set of judgments to the agent in question:

A. One ought to view the interests of others as placing limits on what one can permissibly do in pursuit of power and wealth.

B. Whenever the potential gain in power and wealth is sizeable enough, one is justified in acting with utter disregard for the interests of others.

The tension between these judgments undermines the compatibility thesis. It is difficult to see how one could coherently value justice for its own sake while simultaneously judging that it is overridable by the very sort of considerations that it constitutively takes precedence over.

3. Justice as Part of One's Own Good

When considering evaluative attitudes there is a natural tendency to think in terms of the comparative value assigned to possible states of affairs, without particular regard for how one's own agency is connected to them. And if valuing justice for its own sake (in the relevant sense) were merely a matter of preferring states of affairs in which justice is generally realized over those in which it is not (when all else is equal), then the compatibility thesis would be true. To see why, it will help to return to the example of racial equality. Strange as it would be, it is possible for an agent to assign intrinsic value to states of affairs in which racial equality is widely practiced, while also believing that she has decisive reason to prioritize the interests of members of her own race. Perhaps the agent attributes *no* value to states of affairs in which individuals favor members of races other than their own, *some* value to states of affairs in which racial equality prevails, and *most* value to states of affairs in which everyone favors members of their own race. Likewise, it is possible for an agent to assign *some* intrinsic value to states of affairs in which justice is widely practiced while still preferring unjust states of affairs in which she becomes rich and powerful. Hence, if valuing justice for its own sake were only a matter of having certain preferences concerning states of affairs, it would be true that assigning intrinsic value to justice is compatible with endorsing Gyges' behavior.

However, while we grant that agents with the evaluative attitudes described above are conceivable, they would not value the respective normative objects (racial equality and justice)

intrinsically—at least not in the relevant sense. Assigning some intrinsic value to states of affairs in which racial equality obtains is not the same as valuing racial equality *itself*—i.e. racial equality *qua* normative object. A person who intrinsically values a normative object itself is someone who accepts and embraces the central principles and requirements for herself and believes that others ought to do the same. This explains why it would be inapt to describe the agent from the previous example, whose strongest preference is for a world in which everyone prioritizes the interests of their own race, as someone who values racial equality for its own sake.

That the relevant assessment concerns valuing normative objects themselves rather than merely having preferences for certain states of affairs is clear from the context of the debate in Book II. The central question of the discussion concerns the ends and ideals that a person must pursue and adopt in order to live well. This is evident from the text. When Socrates declares that justice is among the finest goods, he says that “anyone who is going to be *blessed with happiness* must love [it] both because of itself and because of its consequences” (*egô mén oimai, ên d’ egô, en tô kallistô, ho kai di’ auto kai dia ta gignomena ap’ autou agapêteon tô mellonti makariô esesthai*) (358a1-3, emphasis added). The reference to the agent’s own happiness is evidence that the topic of concern is whether justice is intrinsically good *for the agent* who pursues it. Consider also Glaucon’s introduction of the third category of goods: “We’d say that these are burdensome but *beneficial to us*, and we wouldn’t *choose* them for their own sakes, but for the sake of the rewards and other things that come from them” (*tauta gar emipona phaimen an, ôphelein de hêmas, kai auta men heautôn heneka ouk an dexaimetha exein, tôn de misthôn te charin kai tôn allôn bosa gignetai ap’ autôn*) (357c7-d2, emphasis added).

Clearly the relevant notion of intrinsic valuing at issue involves choosing the object as a non-instrumental part of one’s own good. And a person who chooses to pursue justice as a non-instrumental part of her good is someone who accepts and embraces the central norms of justice for herself. Such an agent would not think that it is enough to perform the acts that a just person would

on only those occasions in which she would not profit from injustice. This is because the non-instrumental value of justice could not be realized by the mere performance of a handful of ostensibly just acts; the value could only be realized by *being* just (or at least striving to become just). And one cannot make a sincere effort to become just while simultaneously endorsing the complete disregard of others whenever the financial rewards are sufficiently great.³ In light of these points, we reconstruct Glaucon's argument as follows:

1. Any agent who chooses to pursue justice as a non-instrumental part of their good is someone who strives to become a just person.
2. Any agent who strives to become a just person is someone with a reasonably stable commitment to justice; their reflective endorsement of giving consideration to the interests of others does not depend on the amount of riches obtainable by disregarding those interests.
3. Any agent who chooses to pursue justice as a non-instrumental part of their good is someone with a reasonably stable commitment to justice. (1-2)
4. Anyone who would reflectively endorse Gyges' acts does not have a reasonably stable commitment to justice because they reject norms that are central to it.
5. Everyone would reflectively endorse Gyges' acts (at least in private).
6. Nobody has a reasonably stable commitment to justice. (4-5)
7. Nobody chooses justice as a non-instrumental part of their good. (3 and 6)

³ In order for the non-instrumental value of justice to be realized in one's life, the commitment needn't be of such a degree that the agent could never succumb to the temptation to act unjustly. Rather, the agent who values justice in the relevant sense is someone who would not calmly and reflectively endorse clear violations of the basic norms of justice. Glaucon's challenge is made formidable by his claim that everyone would give a rational endorsement of Gyges' behavior, were they to answer honestly.

On this characterization, the goal of the Gyges argument is not just to refute the claim that most people think justice is *always* more choiceworthy than injustice. Rather, the goal is to show that none of us value justice for its own sake *at all*.

If nobody chooses to pursue justice for its own sake, this is a *prima facie* reason to doubt that it *ought* to be chosen for its own sake; we would need strong grounds for believing that so many people are making such a significant error. Hence, Plato includes the Gyges story in the dialogue not merely to force us to think about the extent of our commitment to justice but also because it represents a significant challenge to the supposed superiority of the just life.⁴

⁴ Earlier versions of this article were presented at an American Philosophical Association Central Division Meeting, Australian Catholic University, and Florida Atlantic University. We thank attendees of those presentations for helpful questions and comments. We are especially grateful to Robert Audi, Hugh Benson, Eric Brown, Stephanie Collins, Julia Driver, Fay Edwards, Jason Gardner, Daniel Hagen, Ashley Kennedy, David Killoren, Sarah Malanowski, Richard Rowland, Travis Timmerman, and an anonymous reviewer for useful suggestions.

References

- Annas, J. 1981. *An Introduction to Plato's Republic*. Oxford University Press.
- Burnet, J. ed. 1900-1907. *Platonis Opera*. Oxford University Press.
- Foster, M. B. 1937. "A Mistake of Plato's in the *Republic*." *Mind* 46: 386-93.
- Heinaman, R. 2002. "Plato's Division of Goods in the *Republic*." *Phronesis* 47: 309-335.
- Irwin, T. H. 1999. "Republic 2: Questions about Justice." *Plato 2 Ethics, Politics, Religion and the Soul*, ed. G. Fine. Oxford University Press.
- Kirwan, C. A. 1965. "Glaucon's Challenge." *Phronesis* 10: 162-73.
- Mabbott, J. D. 1978. "Is Plato's *Republic* Utilitarian?" *Plato II*, ed. G. Vlastos. Notre Dame University Press.
- Pappas, N. 1995. *Plato and the Republic*. Routledge
- Payne, A. 2011. "The Division of Goods and Praising Justice for Itself in *Republic II*." *Phronesis* 56: 58-78.
- Reeve, C. D. C. 1988. *Philosopher kings*. Princeton University Press.
- . 2004. *Republic: Translated from New Standard Greek Text, with Introduction*. Hackett.
- Shields, C. 2006. "Plato's Challenge: The Case Against Justice in *Republic II*." *Blackwell Guide to Plato's Republic*. Blackwell.
- White, N. P. 1984. "The Classification of Goods in Plato's *Republic*." *Journal of the History of Philosophy* 22: 392-421.