

Three aspects of Kantian Autonomy: Independence, Self-Determination and Citizenship

Lucas Thorpe

Boğaziçi University, Philosophy Department ✉

Sun Demirli

Boğaziçi University, Philosophy Department ✉

<https://dx.doi.org/10.5209/kant.95646>

Received: 22-04-2024 • Accepted: 18-09-2024

ENG Abstract: In the *Groundwork*, we find three distinct conceptions of freedom: (i) A negative conception of freedom, understood as a capacity for spontaneous action independent of alien causes; (ii) a positive conception of freedom, understood as the capacity of giving law to oneself; and (iii) a second positive conception, understood as the capacity to give laws that bind others as well as oneself. The dominant interpretation of Kant ignores this third conception of freedom and interprets the second conception as a capacity for self-determination or self-governance, which is seen as the most basic and important Kantian conception of autonomy. In this paper, we will focus mainly on explaining the distinction between the second and third conceptions of autonomy and will defend the claim that the third conception is morally the most basic. We will also explain briefly why Kant thinks that having the capacity for citizenship is necessary for having the capacity for self-governance.

Keywords: freedom, autonomy, Rousseau, self-governance, lawgiving.

Summary: 1. Autonomy as the Capacity for Self-Governance. 2. Autonomy as the capacity for citizenship. 3. Why does the capacity for self-governance presupposed the capacity for citizenship?

How to cite: Thorpe, L.; Demirli, S. (2024). Three aspects of Kantian Autonomy: Independence, Self-Determination and Citizenship. *Con-Textos Kantianos* 20, 41-49. <https://dx.doi.org/10.5209/kant.95646>

At the beginning of *Groundwork* III, Kant distinguishes between freedom understood in the negative sense as a “property of such a causality [of the will] that it can be efficient independently of alien causes determining it” (4:446) and freedom understood in the positive sense as “the will’s property of being a law to itself” (4:447)¹. The dominant interpretation of Kant interprets this second conception of freedom as a capacity for self-determination or self-governance, and argues that this is the most basic and important Kantian conception of autonomy. We believe, however, that in *Groundwork* II, Kant implicitly distinguished between two distinct notions of autonomy: the conception of autonomy as self-governance and a richer notion of autonomy as a capacity for giving laws that bind both ourselves and others. This third conception of autonomy is the concept of a capacity that is required if one is to have the potential to be member of an ideal republic, or a citizen in a realm of ends. And we take this conception of autonomy, which we will call the capacity for citizenship, to be the most basic and morally significant conception of autonomy in Kant’s ethics. Thus, in the *Groundwork* we find three distinct conceptions of freedom:

- (1) A negative conception of freedom understood as a capacity for spontaneous action independent of alien causes. We call this conception of freedom the capacity for independence.
- (2) A positive conception of freedom understood as the capacity of giving law to oneself. We call this conception of freedom the capacity for self-governance (or self-determination).

¹ The German here is: “die Eigenschaft des Willens, sich selbst ein Gesetz zu sein”. This could also be translated as: “the property of the will, to be itself a law” – the “to” itself in the English translation is not mandated by the German. Although the “sich” here can be read in reflexive terms, it need not be. On this second reading what is important is being exemplary, not being self-determining and being exemplary is one way of thinking about what one does if one gives laws that bind oneself and others. So we take it that a case can be made for the claim that the positive conception of freedom Kant discusses at the start of *Groundwork* III should be thought of in terms of what we will call the third conception of autonomy (as the capacity for citizenship) rather than the second (the capacity for self-determination). This paper develops ideas from a number of earlier papers and for those interested in seeing these points developed in more detail can look at especially Thorpe (2011) and Thorpe (2024), but also Thorpe (2010a, 2010b and 2013).

(3) A positive conception of freedom as the capacity to give laws that bind others and oneself. We call this conception of freedom the capacity for citizenship².

Although these are three distinct concepts of freedom, this does not imply that Kant thinks that there are three different kinds of freedom³. For Kant thinks that the only way we can be free in the first sense is if we are free in the second sense, and the only way we can be free in the second sense is if we are free in the third sense. We can think of these three conceptions of freedom as having different senses, but (if they refer) the same reference⁴. Kant's thesis is that if the human will falls under any of these concepts it falls under all of them. Therefore, if we are free, there is a single capacity that we must possess that can be described in either of these three ways. If we are right here then the claim that negative freedom entails positive freedom is best understood as a synthetic (a priori) rather than analytic claim. The move from the concepts of autonomy as citizenship to self-determination to independence may perhaps be explained in terms of conceptual analysis as we might think that self-governance is built into the concept of citizenship and the concept of independence is built into the concept of self-governance. But it is not plausible to think that the argument in the other direction can also be understood in terms of conceptual analysis⁵. In this paper we will focus mainly on explaining the distinction between the second and third conceptions of autonomy and will defend the claim that the third conception is morally the most basic. We will also briefly discuss why Kant thinks that having the capacity for citizenship is necessary for having the capacity for self-governance.

1. Autonomy as the Capacity for Self-Governance

There are two sets of passages in *Groundwork II* where Kant develops his account of autonomy of the will. The first set of passages begin at 4:431. In these passages Kant introduces the moral principle of autonomy, which can be understood as the command to “be autonomous!” in the sense of acting as if you were citizen in a realm of ends – being both sovereign and subject to the laws one has given. Because Kant is committed to the claim that ought-implies-can, this moral principle implicitly presupposes that we possess the capacity to be autonomous in the sense of possessing the capacity to be a citizen in a realm of ends.

The second set of passages begins at 4:440 and contrasts autonomy of the will with heteronomy and suggests a more individualistic conception of autonomy as self-governance or self-determination. What is important in this conception of autonomy is the idea of (somehow) giving law to oneself. Most recent anglophone commentators seem to take the second set of passages as central and assume that Kant's most basic conception of autonomy is in terms of self-governance. We believe, in contrast, that the first set of passages contains what Kant takes to be the most basic notion of autonomy, and that we need to interpret the second set of passages in light of the first.

Most commentators take Kant's remarks starting at 4:440 to be something like Kant's canonical account of the nature of autonomy and on this basis identify the most basic Kantian notion of autonomy with something like self-determination or self-governance. According to this conception of autonomy, to be autonomous is to be a law unto oneself in the sense of ruling oneself. What is essential for autonomy understood in these terms is ruling oneself where the law one gives oneself is not imposed by something or someone external to one's will. Unlike the purely negative conception of freedom as independence, this conception of freedom is positive in that involves the idea of being ruled by reason or law. But this conception of autonomy seems to be: (1) essentially individualistic in the sense that one could be autonomous in these terms independent of one's relations with other human beings, and (2) at least partially negative in the sense that to be autonomous in this sense seems to be defined as being non-heteronomous. And so, it is defined negatively as the opposite of its contrasting property.

This understanding of autonomy seems to be the dominant interpretation today. Thus Schneewind (2013) claims that “Autonomy means self-governance” (p. 146); Schaab (2022) defines Kantian autonomy as “our

² Thorpe (2024) names this “the capacity for sovereignty”.

³ Here we agree with Guyer (2012, p. 71) and O'Neill (2000, p. 29).

⁴ As we cannot know whether we are free, we cannot know that any of these concepts refer. So, Kant's claim is a hypothetical one: If any of these three concepts refers, then they must be co-referential, referring to the same capacity.

⁵ There are a number of passages where it seems that Kant suggests that the claim is analytic, but these passages do not need to be interpreted in this way. For example, in the *Critique of Practical Reason*, Kant argues that, “Consciousness of this fundamental law may be called a fact of reason because one cannot reason it out from antecedent data of reason, for example, from consciousness of freedom (since this is not antecedently given to us) and because it instead forces itself upon us of itself as a synthetic a priori proposition that is not based on any intuition, either pure or empirical, **although it would be analytic if the freedom of the will were presupposed** [ob er gleich analytisch sein würde, wenn man die Freiheit des Willens voraussetzte]; but for this, as a positive concept, an intellectual intuition would be required, which certainly cannot be assumed here.” (5:31 – emphasis added) In the clause emphasized it looks like Kant is claiming that the move from freedom to being bound by the moral law (from negative freedom to positive freedom) is an analytic claim. But the next clause suggests only that this move *would* be analytic for an intuitive intellect, and we take it that it is important that this claim about analyticity is in the subjunctive. The fact that a certain claim would be analytic for an intuitive intellect tells us nothing about whether the claim is analytic for us as our concepts are quite different from the concepts an intuitive intellect would have. Similarly, in *Groundwork III*, Kant claims that “[i]f, therefore, freedom of the will is presupposed [vorausgesetzt wird], morality together with its principle follows from it by mere analysis [Zergliederung] of its concept.” (4.447). Now, we suggest that the claim that is trying to make here is that if freedom of the will is presupposed as given, then the claim *would* be analytic. The claim should be in the subjunctive, but is not. But Kant corrects this error in the *Critique of Practical Reason*. And in the following sentence Kant claims that the principle of morality is “nevertheless always a synthetic proposition” (4.447) and he explains that in synthetic propositions “the two cognitions are bound together by their connection with a third in which they are both to be found” (4.447). And he argues that “the positive concept of freedom provides this third cognition” (4.447). This should make it clear that Kant does not think that the relationship between the various sorts of freedom should be understood as analytic.

will's being a law to itself—its ability to be efficacious independently of empirical causes.” (p. 48); Sensen (2022) claims that “Kant and contemporary clusters of autonomy only share a very general definition of “autonomy” as self-determination.” (p. 56); And Andrews Reath (2013) argues that,

Kant's thesis of autonomy of the will should be interpreted somewhat narrowly as the sovereignty of the will over itself, which we understand to have both a negative and a positive dimension... Negatively the rational will is not bound by any externally given principles or authority. Positively it is a law to itself in the sense that (I argue) the nature of rational volition is the source of its own fundamental norm, a principle that Kant identifies as the Categorical Imperative. (pp. 32-3)

Similar claims are very easy to find⁶. And Kant himself does seem to define autonomy in these terms when, in the *Groundwork*, he introduces the distinction between autonomy and heteronomy of the will, arguing that, “[a]utonomy of the will is the property of the will by which it is a law to itself (independently of any property of the objects of volition).” (4:440)⁷. And this definition is in a sense negative as Kant seems to here equate being autonomous with being non-heteronomous, with heteronomous will being defined on the following page as a will that “does not give itself the law; instead the object, by means of its relation to the will, gives the law to it.” (4:441) Let us make three remarks about this conception of autonomy as self-governance.

Firstly, the idea of autonomy as self-governance can be cashed out in either psychological or normative terms. On the one hand it is very common to interpret self-determination in psychological or motivational terms as a capacity to be motivated by pure practical reason rather than by empirical ends or inclinations. And we take it that this is probably the most common interpretation of Kant's most basic conception of autonomy. This seems to be the conception of autonomy proposed by Hill (2013) when he argues that:

To have a rational will with autonomy (as described here) is, primarily, to have certain general capacities and dispositions to act and cause events in accord with our ideas of rational principles without being determined by non-rational ('alien') factors. (p. 20)

On the other hand, there are commentators such as Reath (2013) who argue that autonomy (as self-determination) is best understood primarily in normative rather than in psychological or motivational terms. Understood in such terms, to be autonomous is to have a certain type of normative authority. For a will to be autonomous is for it to be “normatively self-governing” (p. 33) which is for it to be “the source of its own fundamental norm” (p. 33). And Reath argues that other notions of autonomy, such as the motivational account, need to be somehow grounded in this normative notion of autonomy⁸. Two aspects of Reath's account of autonomy are worth noting. Firstly, this normative authority is primarily self-directed; the essential normative authority we have is over ourselves, not towards others. Secondly, the content of the legislation is not particular laws, but our most fundamental and abstract norm. We will return to these two points later in the paper and argue that (1) our most basic type of normative authority is addressed to others and (2) the content of such (authoritative) legislation is not our most basic norm, but particular substantive laws⁹.

Secondly, whether we interpret autonomy in motivational or normative terms both versions of autonomy as self-determination seem to involve a conception of autonomy that makes no essential reference to any human community. One could, one might think, be autonomous in this sense if one were alone in the universe. Autonomy of the will, understood in terms of self-governance, would seem to be an intrinsic (non-relational) property of the will. Understanding autonomy essentially in terms of self-determination or self-governance seems¹⁰ to make autonomy compatible with what David Lewis (1983) calls loneliness: the existence of the capacity of self-governance seems to be compatible with the individual will being the only thing that exists; one might think that God, alone in the universe, could be autonomous in the sense of being self-determining or self-governing. And Langton and Lewis (1998) argue that we can define the concept of an intrinsic property as a property that something can have no matter what else exists and no matter what the laws (of nature) are; “a property is intrinsic just in case it is compatible with loneliness and lawlessness” (Langton, 1998, p. 119). It would seem that those who take autonomy to be essentially understood in terms of self-determination or self-governance are thinking of autonomy as an intrinsic property of the will in this sense. If we are right, however, the capacity of autonomy is not an intrinsic property in Lewis and Langton's sense as it is not compatible with loneliness, because it essentially has to do with our relationship with other agents¹¹. Being autonomous involves striving to interact with other agents on the basis of mutual respect, and so the capacity of autonomy

⁶ An anonymous referee argues that we are reading these authors uncharitably and that they all ultimately have more sophisticated readings. But our aim here is just to show how prevalent this sort of language is in describing Kant's understanding of autonomy.

⁷ “Autonomie des Willens ist die Beschaffenheit des Willens, dadurch derselbe **ihm** selbst (unabhängig von aller Beschaffenheit der Gegenstände des Willens) ein Gesetz ist.” (G 4:440 – emphasis added.) Unlike the passages in which Kant uses “sich selbst” which is ambiguous, the dative “ihm” here clearly warrants the translation “to itself”.

⁸ “[I]n order to understand Kant's conception of autonomy, we do best to begin from this confined notion and build out to related ideas, such as the motivational capacity to act from rational principle independently of sensibly given interests, the freedom of will, the moral standing of individual rational agents as ends in themselves, and so on.” (Reath 2013, pp. 33-34)

⁹ On the second point we agree with Kleingeld & Willaschek (2019).

¹⁰ We use the ‘seems’ language here because, if Kant is right, the capacity for self-governance or self-determination presupposes the capacity for citizenship and so one could not have the capacity for self-determination if one were alone in the universe.

¹¹ O'Neill (2015b) argues that “If individuals could prescribe laws or principles only for themselves, the metaphor of *self-legislation* would inflate what is only a process of choice or decision.” (p. 122). This would suggest that she also thinks that autonomy is incompatible with loneliness and so is not an intrinsic property of the will in Langton and Lewis' sense.

presupposes the existence of other agents with whom one could chose to interact with in this way, and so the capacity for autonomy is not an intrinsic property of the will in Langton's sense¹².

Finally, there are questions about the nature and status of the self that is to be thought of as the giver of the law(s). Is this self to be understood as a particular concrete individual or is the giver of the law impersonal reason somehow speaking through the particular concrete individual? Proponents of a view like Reath's, according to which autonomy involves the authority to legislate the most basic norm to oneself take the self that is the giver of the law in these impersonal terms. And a similar approach is taken by O'Neill (2015a, p. 118) who argues that,

Kant takes himself to be giving an account of the sort of lawgiving that is reason's own, and not an account of lawgivings that are an agent's own. Kantian autonomy is reason's lawgiving rather than the lawgiving of individual agents (whatever that might mean).

And she argues that advocating a conception of autonomy in terms of a lawgiving by particular concrete individual agents is likely non-sensical and at best highly implausible¹³. However, as we will argue in the following section, Kantian autonomy is best thought of in terms of legislation by particular concrete agents. And we hope to show, at the very least, that such a conception of autonomy is not obviously non-sensical.

One might worry here, as an anonymous referee for an earlier version of this paper did, that particular concrete individuals cannot themselves legislate as this is only something that a collective united will can do, and that individuals can at the most participate in this legislation; only collectively as a united will do we have a capacity to legislate; thinking that an (arbitrary and unilateral) individual will has a capacity to legislate is to endorse autocracy or despotism, not Kantian autonomy.

Now, one option here, would be to bite the bullet and argue that Kant (at least in the 1780s) thought that legitimate lawmaking could be done by a despot, as long as the laws that were given were universalizable in principle. Kleingeld (2017) suggests such a reading, arguing that at "no point in his discussion of the Principle of Autonomy [in the *Groundwork*] does Kant claim that others' actual attitudes towards our maxims must be taken into consideration" (p. 65). And she argues that this understanding of autonomy is modeled on his understanding of political legislation (in the 1780s) according to which legislation does not require the united will of all. Thus, Kleingeld appeals to Kant's Lectures on Natural Law from 1784, where she argues that he is committed to the position that "a political legislator can give laws that are fully just even if he does not have the actual consent of any subjects. The possibility of general consent suffices" (p. 70). According to Kleingeld's interpretation Kant only came to embrace the idea that legislation required the united will of all only in the 1790s, and because of this he no longer appealed to his earlier notion of autonomy.

Although Kleingeld's reading has some historical plausibility, we think that Kant's understanding of autonomy in the 1780s is compatible with his claim that only a united collective will can legislate because we think that claiming that each citizen must be thought of as having the capacity to give law, is compatible with the claim that law can only be made by the united will of all, so the capacity to give law is both an individual and collective capacity. Think about the capacity to make a contract with another person. Is this capacity one that both individuals possess, or a joint capacity? Surely, it can be both. One party can make an offer, and this offer only takes on the normative force of a contract once the other party has agreed. But the content is not changed by the offer being accepted by the other party. Similarly, I may adopt a maxim as my subjective principle of action and invite others to endorse it too. For it to become a law would be for it to be taken up by all others, for only then would a subjective principle of action become objective. But the content of the maxim need not change in such a process. An objective maxim in such a case would be fully my maxim and our maxim.

One way an individual can give law to others who are essentially free is by being exemplary. And this is one of the aspects of autonomy that Kant stresses in the *Critique of the Power of Judgment*. Taking classical authors as an inspiration, as long as this is not a matter of blind imitation, does not undermine the autonomy of the new generation.¹⁴ So we can think of the individual who succeeds in giving law (who successfully puts

¹² If our reading is right, then this creates serious difficulties for Langton's (1998) influential account of the distinction between phenomena and things-in-themselves. According to Langton the distinction is to be understood as a distinction between the extrinsic and the intrinsic properties of things, with the claim that we cannot know things-in-themselves being equivalent to the claim that we cannot know the intrinsic properties of things. Now, it is clear that for Kant autonomy of the will, if it exists, would have to be a property of things as they are in themselves and not a phenomenal property. But, if we are right, then the property of autonomy is not compatible with loneliness, as one can only be autonomous if other rational agents with whom one can interact exist, and so would have to be an extrinsic property in Langton's sense. And hence this would imply that we cannot identify things as they are in themselves with intrinsic properties.

¹³ "If Kantian autonomy is pictured merely as *legislation by individual selves*, the convergence of individual wills on anything that should count as morality remains a mystery and the moral importance of autonomy remains obscure. Unsurprisingly, many passages in Kant's writing are reduced to nonsense, or at the very least implausibility, if we equate his conception of autonomy with 'legislation' by individuals." (Oneill, 2015a, p. 114)

¹⁴ Thus, in the *Critique of the Power of Judgment*, Kant claims that "the works of the ancients are rightly praised as models, and their authors called classical, like a sort of nobility among writers, who give laws to the people through their precedence, seems to indicate a *posteriori* sources of taste and to contradict the autonomy of taste in every subject. But one could just as well say that the ancient mathematicians, who have been regarded until now as nearly indispensable models of the greatest thoroughness and elegance of the synthetic method, also demonstrate an imitative reason on our part and its incapacity to produce from its own resources strict proofs, with the greatest intuitive evidence, by means of the construction of concepts. There is no use of our powers at all, however free it might be, and even of reason (which must draw all its judgments from the common source a priori), which, if every subject always had to begin entirely from the raw predisposition of his own nature, would not fall into mistaken attempts if others had not preceded him with their own, not in order to make their successors into mere imitators, but rather by means of

forwards a universalisable subjective maxim that is taken up by all others) as exemplary. Given the possibility of such exemplary commitments, giving law does not have to be despotic.

In addition, one might think that although possession the capacity to give laws is something we all may possess individually, the capacity can only be (fully) actualized collectively. In this the capacity for autonomy may be like our linguistic capacities. We individually possess the capacity to express ourselves linguistically, and there is a sense in which this capacity is innate. But this capacity can only be actualized as part of a linguistic community; in order to express ourselves linguistically we have to learn particular, concrete languages by becoming members of particular linguistic communities, so actualizing our individual linguistic capacities is something we can only do by being part of a collective. Similarly, we each have the capacity to give laws in the sense of acting on subjective maxims that have the potential to be objective maxims (and such potentially objective maxims must be universalisable), but such maxims can only actually become objective maxims in so far as they are the expression of a collective will. This is why Kant argues in the *Metaphysics of Morals* that we ought to leave the state of nature and enter the civil condition (6:312)¹⁵.

2. Autonomy as the capacity for citizenship

Instead of identifying autonomy with the capacity for self-governance, Kant's initial introduction of the idea of autonomy ties being autonomous with the idea of being a citizen of a realm of ends, or (what he seems to treat as equivalent) being a member of an intelligible world. And, as Kleingeld and Willaschek (2019, p. 2) point out Kant is committed to the claim that the highest moral principle commands autonomy.

As many commentators have noted Kant uses the term "autonomy" to refer both to a property of the will and to a moral principle¹⁶. On the reading we offer these two senses are intimately related as autonomy as a property of the will can be understood as a capacity or potentiality, and the principle demands that we actualize this potential; the third formulation of the categorical imperative basically tells us: "be autonomous!" and so what it demands of us is that we actualize our potentiality for autonomy. This capacity to be autonomous is more like a non-verbal baby's linguistic capacity than an adult's capacity to speak, say, English. The first type of capacity is a potentiality that may (or may not) be actualized, whereas the second type of capacity is an ability than can be exercised. So, our capacity for autonomy is more like our capacity to learn a language than our capacity to speak a particular language. In so far as autonomy as a property of the will is understood as a capacity for citizenship in an ideal moral republic, then the moral principle of autonomy demands that we actualize this capacity, which involves developing a character that would make one suitable to be a citizen in such a republic and would involve treating other human beings with respect as co-citizens. This is why the property of autonomy understood as a capacity for citizenship is morally the most basic notion, as this is the aspect of autonomy that is targeted by our most basic moral principle.

To be autonomous, in this sense is a necessary membership condition for being a citizen in a realm of ends. Thus, Kant explains that "[a] rational being belongs as a member to the realm of ends when he gives universal laws in it but is also himself subject to these laws" (4:433)¹⁷. And Kant also thinks autonomy, understood as the capacity to give laws for a realm of ends, is what makes such a community itself possible. Thus, he argues that,

[I]n this way a world of rational beings (*mundus intelligibilis*) as a realm of ends is possible, through the giving of their own laws by all persons as members (4:438)

Note that here Kant identifies the notion of realm of ends with the idea of an intelligible world. And so, it is helpful to read this claim in light of Kant's claims about the idea of an intelligible world in his metaphysics lectures¹⁸. Kant believes that the idea of an intelligible world is the idea of a community of individuals in interaction and thinks that it is possible to conceive of a community of individuals in interaction only if we think of the members of the community as governed by laws, and we can think of the members of a community as governed by laws only if we think of each individual member of the community as the source or giver of these laws. Kant believes that a world is essentially unified, for it is this unity that distinguishes the idea of a world from that of a mere multitude. In addition, he believes that a multitude of individual substances can only be unified or "held together" by laws. So, the idea of a world is the idea of a multitude of individuals unified by laws. Now, if the unity of a world is to be "intrinsic" to the world, rather than merely existing in the mind of some ideal observer observing the world – that is, if there is to be real interaction between the members of

their method to put others on the right path for seeking out the principles in themselves and thus for following their own, often better, course... [G]eneral precepts, which one may either have acquired from priests or philosophers or drawn from oneself, never accomplish as much as an example of virtue or holiness, which, established in history, does not make the autonomy of virtue out of one's own original idea of morality (*a priori*) dispensable or transform this into a mechanism of imitation. **Succession**, related to a precedent, not imitation, is the correct expression for any influence that the products of an exemplary author can have on others, which means no more than to create from the same sources from which the latter created, and to learn from one's predecessor only the manner of conducting oneself in so doing." (5:282-5)

¹⁵ As Kant explains in his Feyerabend Lectures on Natural Right, "I do wrong to others if I wanted to make my will into their law, hence I am obligated to subject myself to an extremal law that is valid for everyone." (27:1338). For an interesting discussion of this passage, and its relationship to our duty to enter the civil condition see Kleingeld (forthcoming).

¹⁶ See, for example, Allison (1990, p. 94).

¹⁷ And he adds that a morally good disposition "makes him fit to be a member of a possible kingdom of ends" (4:435). And Kant argues that it is this fitness to be a citizen in a realm of ends in which the dignity of human beings consists.

¹⁸ For a more detailed account of the argument sketched in this paragraph see Thorpe (2010b) and Thorpe (2011). In stressing the roots of the idea of autonomy in Kant's metaphysics and the idea of being a member of an intelligible world we depart from Kleingeld (to whose interpretation we are generally sympathetic) who roots this conception primarily in his political philosophy.

the world rather than a mere constant conjunction between the state of one substance and that of another (à la Hume or Leibniz) – then the members of the world must be responsible for the unity of the world, and Kant believes that this is possible only if each individual member of the world is the source of, or “the giver of,” the laws that provide the world with its unity. That is, in the language of his mature ethics, we can think of a community of individuals in interaction only if we think of each individual member of the community as autonomous. And to be autonomous is not merely, or even primarily, to rule oneself, but rather to be the source of the laws of a possible community, to be a co-legislator in an ideal republic.

Thus, drawing on ideas developed in his metaphysics lectures, in the *Groundwork* Kant identifies the idea of a realm of ends with the idea of an intelligible world and explains that “by a realm we understand a systematic union of various rational beings through common laws ... what these laws have as their purpose is just the relation of these beings to one another as ends and means” (4:434). Kant makes it clear that the purpose of laws in a realm of ends is to provide the “glue” that gives a community of individuals some sort of unity that allows them to interact intelligibly¹⁹. And he continues by explaining that the only way to be a member of such a community is by being autonomous, that is by being (individually) the source of the laws that provide the community with its unity. And he explains that “a rational being belongs as a member to the realm of ends when he gives universal laws in it but is also himself subject to these laws” (4:434). Here Kant seems to be quite explicit about the fact that being autonomous is the membership condition for belonging to a realm of ends²⁰.

On this reading, Kant’s third positive conception of autonomy, is a descendent of Rousseau’s conception of citizenship introduced in the *Social Contract*. Rousseau argues that in an ideal republic each individual member is both sovereign (as citizen) and subject to the law. Thus, in an ideal republic:

[I]n place of the particular individuality of each contracting party, this act of association produces a moral and collective body, composed of as many members as the assembly has voices, and which receives from this same act its unity, its common self (moi), its life, and its will. This public person, which is thus formed by the union of all the individual members, used to be called a city, and now is called republic or body politic. When it is passive, it is called by its members State, and sovereign when it is active, power when it is compared to similar bodies. With regard to the associates, they take collectively the name of people, and are called individually citizens, inasmuch as they participate in the sovereign power, and subjects, inasmuch as they are subjected to the laws of the State. (2002, p. 164)

Citizens are autonomous in the sense of being both sovereign and subject to the law of their republic. And in a republic citizens renounce their freedom to do whatever they want but in return acquire something more valuable, what Rousseau calls moral freedom. Thus Rousseau (2002) claims that,

We might also add to the advantages of the civil state moral freedom, which alone enables man to be truly master of himself; for the impulse of mere appetite is slavery, while obedience to a self-prescribed law is freedom. (p. 167)

Interestingly, here Rousseau suggests that citizenship is required for self-mastery; only citizens in an ideal republic can truly be masters over themselves. And given Kant’s seamless move from the passages in which autonomy is introduced in terms of the capacity to be a member (citizen) of a realm of ends (starting at 4:431) to the passages which identify autonomy with self-governance as opposed to heteronomy (starting 4:440) it is not implausible to assume that Kant is following Rousseau in assuming that the capacity for self-determination (Rousseau’s “moral freedom”) is somehow tied to citizenship.

Now, of course, there are significant differences between Kant and Rousseau’s understanding of the relationship between self-mastery and citizenship. For Rousseau, the ideal republic is a political ideal, for a particular people and not the whole of humanity; and his claim about moral freedom is that one can be morally free only if one is an actual citizen in an actual republic. Kant in contrast takes Rousseau’s political ideal and repurposes it as a moral ideal. For Kant citizenship in this ideal moral republic is not limited by state borders but encompasses the whole of humanity. And most importantly, whereas Rousseau claims that one can only really be morally free (in the sense of mastering oneself) if one is actually a citizen of a republic, Kant seems to implicitly assume that self-governance (Rousseau’s “moral freedom”) is only really possible if one is governed by principles that would allow you to become a member of such a republic²¹; to be (morally) autonomous for Kant is to act as if one were a citizen-legislator in an ideal republic, and self-mastery is only possible if one acts in this way.

¹⁹ Kant thinks that treating others with respect involves interacting with them in an intelligible way; interacting on the basis of force, violence or threats of violence is not intelligible in this sense.

²⁰ Although we are generally sympathetic to Kleingeld and Willaschek’s (2019) interpretation, we think that the argument in this paragraph suggests we should reject, or perhaps modify, their claim that “the Moral Law is a fundamental a priori principle of pure practical reason that is not grounded in anything more fundamental.” (p. 2). We suggest that the Moral Law is, in a sense, (metaphysically) grounded in something more fundamental, which is our (theoretical) idea of an intelligible world or a realm of ends. The possibility of the Moral Law is a necessary condition for the possibility of such a world. And in normative terms it is also grounded in something more basic, because the reason we care about the Moral Law is because we care about interacting on the basis of mutual respect with other human beings, and such interaction is only possible in so far as we are all self-constrained by the Moral Law. For a fuller defense of this last claim see Thorpe (2024).

²¹ We are assuming here that Kant, unlike Rousseau, thinks that being autonomous, understood in terms of self-determination or self-mastery, is possible in a dictatorship. It is not, however, obvious that Kant is ultimately committed to this and perhaps there is a way of reading Kant that makes his position here closer to Rousseau’s – but we will not develop this thought in this paper.

For Kant morality demands that we act as if we were citizens in such an ideal republic, regarding all our fellow human beings as fellow citizens (and hence potential co-legislators) in such a community. This is what constitutes Kantian respect for humanity – we ought to treat others as if they were fellow citizens in an ideal moral republic and according to principles that they can also potentially endorse. Exercising our capacity of autonomy is, in effect, the prerequisite for becoming a true member of a realm of ends. To be subject to laws without being sovereign is to be a slave; to be sovereign without being subject is to be a dictator. A citizen must be both sovereign and subject, both endorsing and subjecting themselves to the laws and these laws must be co-legislable by all other human beings. On this reading, universalisability is to be cashed out in terms of the possibility of co-legislation; a law is universalisable only in so far as it could potentially be willed by all fellow human beings, considered as fellow citizens in an ideal moral republic encompassing all human beings.

And we suggest (like Reath) that this capacity is best understood in normative rather than motivational or psychological terms. To be a sovereign citizen is to have a certain normative authority to give (universal) laws that bind others (and ourselves), on condition that the laws given can also (potentially) be rationally endorsed by all other members of the community. But in talking about laws in this context, Kant does not mean our fundamental moral norm (as Reath would have it) but specific substantive laws, such as the principle that one does not make lying promises. In saying this we agree with Kleingeld (2019) who argues convincingly that Kant's conception of autonomy presented in the *Groundwork* is modelled on his (fundamentally Rousseauian) conception of just legislation developed in his lectures on natural right. In these lectures Kant argues that positive laws are just insofar as they could have been the result of an original contract, that is that they could have been agreed to by all of those subject to the laws. Thus, Kant argues that,

One must represent all laws in a civil society as given through the consent of all. The *contractus originarius* [original contract] is an idea of the agreement of all who are subject to the law. One must test whether the law could have arisen from the agreement of all, if so then the law is right [*richtig*]. (27:1382)

This idea of an original contract as involving the potential agreement to all who are subject to the law is to be regarded at the a priori constitutional law of the state, and is not itself to be thought of as an object of legislation, but instead as a condition for the justness of particular acts of legislation. And so, insofar as we can make an analogy between moral autonomy and just legislation, then the categorical imperative (Reath's "fundamental norm") corresponds to the a priori constitutional law whereas the "universal law[s]" given by an autonomous will correspond to particular, substantive positive laws²².

We also disagree with Reath and O'Neill that the giver of the laws can only be impersonal reason rather than particular concrete individuals. One of O'Neill's worries is that if concrete particular individuals are thought of as the givers of the law, then this would seem to imply that other agents would be subject to the law of another and hence could not themselves be autonomous. Thus, she argues that "[i]f individuals could promulgate or prescribe laws for others, those others would be subject to that law, so could not it seems be self-legislators." (O'Neill, 2015b, p. 122)²³. But, if we understand autonomy in terms of the capacity of co-legislation, then this worry disappears. Now, of course, O'Neill may reformulate her worry in terms of the incomprehensibility or incoherence of the notion of co-legislation, but we think the onus should be on her to explain why and how such a notion is incoherent. The notion of co-legislation seems to presuppose the coherence of some conception of what contemporary authors call we-intentionality, and is, we think, incompatible with a radical form of individualism²⁴. But we have no reason to think that Kant is, or needs to be, committed to such a form of individualism.

For Kant we ultimately morally value autonomy not because we ultimately value ruling ourselves, but because we value being a member of such a community and the idea of interacting with others on the basis of mutual respect. Kant thinks that in so far as we value the ideal of being a potential citizen in an ideal republic with our interactions with others being governed by mutual respect, then we have to presuppose, as something like a postulate of practical reason, that we have such a capacity to give laws that can constrain others and ourselves; as such we have to regard our principles as potentially authoritative and our willing as exemplary²⁵.

Why might one think that self-mastery is only possible if one acts as if one were a citizen in an ideal republic. It seems possible to imagine a finicky eccentric whose life is very rule governed and ordered, but who

²² "We argue that the relation between the Categorical Imperative and the "universal law" mentioned in the Principle of Autonomy (as described in the *Groundwork*) parallels the relation between the a priori constitutional law of the state and positive state laws (as described in the Feyerabend lectures)" Kleingeld, 2019. p. 164).

²³ This worry is analogous to the worry that led Leibniz to the conclusion that monads must be "windowless" and cannot engage in real interaction. Leibniz's worry is based on the thought that individual monads are essentially active with the determinations of an individual dependent upon the activity of that particular individual. Real interaction between individuals would require that monads are passive and possess some of their determinations as the result of the action of an external substance. But Leibniz cannot understand how, in such cases, the determinations would really be properties of the passive monad. For a longer discussion of this see Thorpe (2011).

²⁴ We believe that we-intentionality cannot be reduced to individual intentionality and see no reason why Kant has to be read as implicitly committed to a reductive individualism about intentionality. For a strong version of a non-reductive account of we-intentionality see Zahavi 2021.

²⁵ Thorpe (2018) has argued that for Kant what it is to be human, in the morally relevant sense, is to possess the capacity to interact with others on the basis of mutual respect, with this capacity being best thought of in terms of a social affordance. And we argued that we must presuppose, as a postulate of practical reason, a capacity to recognize which bits of the phenomenal world are human in this sense. We are here suggesting that this recognitional capacity should be thought of as a component of a broader capacity for autonomy.

has no respect for other people. Couldn't such an individual be capable of ruling themselves without being autonomous in the third sense? We take it Kant would have to say that such a finicky eccentric would not really be giving a law to themselves. Let me finish this paper by trying to sketch why we think Kant thinks this.

3. Why does the capacity for self-governance presupposed the capacity for citizenship?

We can think about the capacity of autonomy in the positive sense in terms of a constrained capacity to coerce the will of others (and oneself) and this helps us to explain why autonomy as the capacity for citizenship is a precondition for autonomy as the capacity for self-governance²⁶.

To begin with it is important to understand some things about Kant's use of the notion of coercion. Kant claims that juridical laws are necessarily coercive, and given how we understand the notion of coercion today, this is taken to mean that such laws are essentially enforceable through the use of force – either actual force or the threat of force (punishment). But this is not what Kant means by coercion. For Kant juridical laws are coercive in the sense that they allow one to necessitate or obligate. Thus, in talking about 'powers of coercion' in his ethics lectures, Kant explains perfect obligation as 'an obligation where the agent can be necessitated to an act of duty by another's choice' (Vigilantius, 27: 289). The capacity of coercion is then the capacity to determine the will of another through one's choice, and an obligation is the effect of legitimate coercion. So, we suggest that autonomy in this third sense can be thought of as involving the capacity for (legitimate) coercion.

Now this capacity to coerce is initially other-directed and unconstrained and unreasonable. Our baby cries whenever he needs attention and food. And these cries have a certain power; they are able to determine my will as his father²⁷. Now we suggest that the best way to think of the categorical imperative is as a constraint on our capacity to coerce or bind others; coercing the will of another can only result in an obligation if the coercion is legitimate, and coercion is only legitimate if it is governed by laws that could be endorsed by all; the demands we make on others can only create obligations in so far as they are subject to laws that all parties involved in can endorse. So, the idea of universal law, is a constraint on the legitimacy of coercion. But we must think of this capacity to bind others as prior to and independent of the constraint. So, the capacity of autonomy is the capacity to bind others in a (lawlike) way that is governed by mutual respect.

Now, on this account, the capacity of autonomy is initially based on the constraint of an essentially other-directed capacity. But as we know, constraints can be creative. Coercion is only reasonable and on the basis of mutual respect if it is based on principles that are universalizable. And such principles, in so far as they are universal must pertain to oneself as well as to others. So, coercing others on the basis of our own choice is only legitimate if we ourselves are governed by the same laws. This, we take it, is why autonomy as a capacity for self-governance presupposes autonomy as a capacity for citizenship.

This account implies, we think, that autonomy understood as self-governance also needs to be understood in normative terms; our authority over others can only be legitimate if it is also an authority over ourselves. In this we agree with Reath. But in contrast to Reath's account, the laws that we have the authority to give ourselves (via giving to all) are particular substantive laws, not the fundamental moral norm. The fundamental moral norm (the categorical imperative) is instead a constraint on what laws can be willed legitimately, and is not itself a law that is (or needs to be) willed.

Let's suppose I have borrowed a book from a friend who says I can keep it as long as she doesn't need it, and my friend asks for it back. As a result of her request, I have an obligation to return the book to her. Now, it might seem that the source of the obligation is the request of my friend, the choice of another. And there is a sense in which the coming into existence of the particular obligation to return the book does depend on the decision of my friend to demand it back. But, of course, such a demand only issues in an obligation in the context of the legitimacy of the principle that one should return things one has borrowed if asked²⁸. In order for there to be any possibility of one person obligating another, there need to be some principles that determine if and when demands result in obligations, and such principles, Kant thinks need to be lawlike and universal. So, my friend's choice can only obligate me insofar as we are (at least implicitly) committed to interaction on the basis of mutual respect, and hence in a way that is governed by principles that all parties involved in the interaction can endorse. And when I interact with other (adult) human beings who I recognize (or should recognize²⁹) as individuals who are not mere means, I am at least implicitly committed to the principle of interacting with them on the basis of mutual respect. Now these principles that govern my relations with others,

²⁶ In what follows our interpretation differs from Kleingeld. We take it that an autonomous agent has the capacity to *actually* give laws that bind others and themselves, whereas Kleingeld thinks that this notion of legislation should merely be used as a criterion to evaluate the moral status of their maxims. Thus Kleingeld (2019) explains: "The moral agent is not required literally to give laws but rather to determine the permissibility of his maxims by using the model for evaluating the justice of state laws." (p. 168)

²⁷ Now, and here I am perhaps speaking in my own voice and not Kant's, I think that when my baby cries this crying has normative power. He is able to determine my will, and it is not just that I am motivated to comfort him but that I ought to comfort him. This ought, however, is more basic than the "ought" of obligation; but is a necessary condition for the possibility of legitimate obligation. Thorpe (2024) suggests that this normative force that makes obligation possible is to be identified with the dignity of humanity.

²⁸ And this principle needs to be legitimate in the sense of being an actual law that governs our relationship, not merely a principle that could govern our relationships and as such must not merely be legitimately willable (as Kleingeld suggests) but must be (at least implicitly) willed as a law.

²⁹ The suggestion here that we *should* recognize the humanity of others presupposes a moral realist reading of Kantian ethics, and a notion of a moral demand that is deeper and more basic than the notion of a moral obligation developed in the following paragraphs that is somehow dependent upon the will. It is because we recognize (or *should* recognize) the humanity of other that we should only interact with them in ways that are universalizable. The normative demand that the humanity of others makes on us is a demand the recognition of which grounds the moral requirement of universality; we only morally care about universality because we care about humanity. So the value we place on universality presupposes the dignity of humanity.

Kant thinks, need to be universal, and, as so, must govern my relationship to myself as well as my relations to others. If I am obligated by a promise I have made to another, then another must be obligated by a promise they have made to me and I must be obligated by a promise I have made to myself. So, my capacity to govern myself is parasitic on my capacity to give laws to others; my (normative) capacity for self-governance is parasitic on my (normative) capacity for (moral) citizenship. This is why, we think, the third concept of autonomy is morally the most basic for Kant; the capacity for self-determination presupposes a capacity to give universal laws that bind others as well as oneself.

An earlier version of this paper was presented by Lucas Thorpe at the Third Chinese Kant Congress, hosted by SEU in Nanjing in June 2023. And we would like to thank the organizers and participants (particularly Professor Bian Shaobin and Zhao Quing) at this event for their useful feedback.

Bibliography

- Allison, Henry E. (1990), *Kant's Theory of Freedom*, Cambridge University Press.
- Guyer, P. (2012). "Progress toward autonomy." In O. Sensen (Ed.), *Kant on Moral Autonomy* (pp. 71-86). Cambridge: Cambridge University Press.
- Kleingeld, Pauline (2017). "The Principle of Autonomy in Kant's Moral Theory: Its Rise and Fall." In Eric Watkins (ed.), *Kant on Persons and Agency*. New York: Cambridge University Press. pp. 61-79.
- Kleingeld, Pauline (2019). "Moral Autonomy as Political Analogy: Self-Legislation in Kant's 'Groundwork' and the 'Feyerabend Lectures on Natural Law'." In Stefano Bacin & Oliver Sensen (eds.), *The Emergence of Autonomy in Kant's Moral Philosophy*. Cambridge University Press. pp. 158-175.
- Kleingeld, Pauline & Willaschek, Marcus (2019). "Autonomy Without Paradox: Kant, Self-Legislation and the Moral Law." *Philosophers' Imprint* 19 (6):1-18.
- Kleingeld, Pauline (forthcoming). "Independence and Kant's Positive Conception of Freedom" in Martin Brecher and Philipp-Alexander Hirsch (eds.), *Law and Morality in Kant*, Cambridge University Press.
- Langton, Rae (1998). *Kantian humility: our ignorance of things in themselves*. New York: Oxford University Press.
- Langton, Rae & Lewis, David (1998). "Defining 'intrinsic'." *Philosophy and Phenomenological Research* 58 (2):333-345.
- Lewis, David (1983). "Extrinsic properties." *Philosophical Studies* 44 (2):197-200.
- O'Neill, Onora (2000). "Agency and Autonomy" in her *Bounds of Justice*. Cambridge University Press
- O'Neill, O. (2015a). "Autonomy: The Emperor's New Clothes." In *Constructing Authorities: Reason, Politics and Interpretation in Kant's Philosophy* (pp. 103-120). Cambridge: Cambridge University Press.
- Reath, Andrews (2012). "Kant's Conception of Autonomy of the Will." In Oliver Sensen (ed.), *Kant on Moral Autonomy*. Cambridge University Press. pp. 32-52.
- Ripstein, Arthur (2004). "Authority and coercion", *Philosophy and Public Affairs*, 32/1 (2004),
- Rousseau, Jean-Jacques (2002). *The Social Contract and the First and Second Discourses*, edited and with an Introduction by Susan Dunn, Yale University Press
- Schaab, Janis David. (2023). "Kant on Autonomy of the Will" in *The Routledge Handbook of Autonomy*, edited by Ben Colburn, Routledge.
- Schneewind, J.B (2012). "Autonomy after Kant." in Oliver Sensen (ed.), *Kant on Moral Autonomy*. Cambridge University Press.
- Sensen, Oliver. (2023). "Autonomy after Kant" in *The Routledge Handbook of Autonomy*, edited by Ben Colburn, Routledge.
- Thorpe, Lucas. (2024). "Guyer, Sellars and Kant on the Dignity and Value of Freedom". *Knowledge, Freedom, and Taste: Internationaler Kant-Preis 2024: Paul Guyer*, edited by Konstantin Pollok, De Gruyter, pp. 21-38.
- Thorpe, Lucas. (2018). "Kant, Guyer and Tomasello on the Capacity to Recognize the Humanity of Others" in Kate Moran (ed.) *Kant on Freedom and Spontaneity*, Cambridge University Press.
- Thorpe, Lucas. (2013). "One Community or Many? Community and Interaction in Kant: From Logic to Politics via Metaphysics and Ethics." In *Politics and Metaphysics in Kant* edited by Howard Williams, Sorin Baiasa and Sami Pihlström, University of Wales Press.
- Thorpe, Lucas. (2011). "Autonomy and Community" in Thorpe, Lucas and Payne, Charlton (eds.) *Kant and the Concept of Community, A North American Kant Society Special Volume*, University of Rochester Press.
- Thorpe, Lucas. (2010b). "Is Kant's Realm of Ends a Unum per Se? Aquinas, Suárez, Leibniz and Kant on Composition." *British Journal of the History of Philosophy*, 18:3, 461-485.
- Thorpe, Lucas. (2010a). "The Realm of Ends as a Community of Spirits: Kant and Swedenborg on the Kingdom of Heaven and the Cleansing of the Doors of Perception." *The Heythrop Journal*, XLVIII, pp. 1-24 AHCI
- Zahavi, D. (2021). "We in Me or Me in We? Collective Intentionality and Selfhood." *Journal of Social Ontology*, 7(1), 1-20.

