

Alan Ward

Never Trust a Wolff?

Philosophical Anarchism and the Problem of Promising

Abstract

The central thesis of philosophical anarchism is that there exists no prima facie obligation to obey the law. Robert Paul Wolff, a seminal philosophical anarchist, has been attacked on the ground that his theory entails that simple promising would also be impermissible. This paper defends Wolff from this charge. I argue first, that Wolff's Kantian roots preclude him from treating promises as merely instrumental and thus violable, and second, that promises to obey the law differs from conventional interpersonal promises. The former are understood to be open-ended while the latter's scope are carefully explicitly or implicitly circumscribed.

Keywords: philosophical anarchism, legitimacy, promising, consent, Wolff, Simmons, Baugh

1. Introduction and preliminaries

The 1970s are widely seen as a period of revival in the field of political philosophy (Ball 1995). John Rawls and Robert Nozick are just two examples of a cohort of philosophers who broke from the behaviorist and utilitarian assumptions of past years, philosophers who tried to articulate the conditions of state legitimacy on the grounds of justice. Robert Paul Wolff is a member of this cohort, but unlike those who sought to define the boundaries of state legitimacy, Wolff unremittingly attacked the concept of the legitimate state itself in his 1970 essay, *In Defense of Anarchism*. Wolff argued that states could not be legitimate because they were necessarily incompatible with moral autonomy, and that the notion of the legitimate state therefore 'must be consigned the category of the round square, the married bachelor, and the unsensed sense-datum' (Wolff 1976, 71). Since the state does not have authority over us, Wolff continued, so the only defensible position for us hold would be what he called 'philosophical anarchism'.

Commentators reacted swiftly and mostly negatively to Wolff's work. Indeed, Wolff has been assailed on multiple fronts. He has been accused of, among other things: not being a 'real' anarchist (Perkins 1972; Frankfurt 1973; Sterba 1977; Dahl 1989; Horton 1992); offering an inadequate defense of the unanimous direct democracy that he concedes may address his concerns (Reiman 1972; Dahl 1989; Sanders 1996); being inconsistent in his claim that the just state is 'logically impossible' (Frankfurt 1974; Martin 1974; Pateman 1985; Simmons 1996); being overly obsessed with moral autonomy at the expense of other virtues (Beauchamp and Witkowski 1973; Pritchard 1973; Dahl 1989; Horton 1992);

confusing legal and moral obligations (Pritchard 1973); and even irresponsibly advancing a dangerous thesis (Laszlo 1973).

After a flurry of critiques and responses, interest in Wolff's philosophical anarchism – what A.J. Simmons later called *a priori* philosophical anarchism – waned as a new version of philosophical anarchism – dubbed *a posteriori* – gained prominence. This new brand of philosophical anarchism, championed by the likes of Simmons (1979), Joseph Raz (1979), and M.B.E. Smith (1973), holds that while the legitimate state is not a logical impossibility, it is either rare or its laws are not morally binding in many cases.

I believe that Wolff's theory can resist many of these critiques and that would-be philosophical anarchists should pause before jumping ship to Simmon et al's weaker version. This paper contributes to the rehabilitation of Wolff's project by addressing and attempting to disarm one the central objections levied against him: that Wolff's position would preclude him from being able to endorse simple acts of promising.¹ The critics who levy this charge are using a *reductio ad absurdum* argument here. Surely Wolff does not want to say that simple promises are impossible or a violation of moral autonomy, but these critics argue, that this is precisely what his framework entails. If this is correct, Wolff's project – and *a priori* philosophical anarchism writ large – is truly dead in the water.

The crux of the charge is as follows. First, Wolff argues that accepting political authority involves an abdication of some or all of our autonomy (Wolff 1976, 16-17). Wolff also believes that our primary moral duty is to act according to our autonomous dictates

whenever possible. Our commitment to acting autonomously implies a duty to safeguard this autonomy. Acknowledging an obedience to obey the state implies an agreement to act contrary to one's autonomy. Thus, Wolff argues, we cannot morally accept the existence of legitimate political authority. But Wolff's critics argue that his radical position vis-à-vis state authority should lead him to reject the institution of promising as well. After all, promising to obey someone – even in a limited capacity – is ostensibly an abdication of one's autonomy. When we promise, we commit to performing an action even if, when it comes time to execute that action, we think it would be right (or otherwise preferable) to do otherwise. We thus bind ourselves to the performance of an action or – worse yet – to the dictates of another. Why should this differ from a promise to obey the state?

This is not the first time that a philosophical anarchist has been accused of being incapable of endorsing the institution of promising. Critics such as Carole Pateman have argued that Wolff should look to William Godwin, an 18th century philosophical anarchist who *embraced* the position that we should ultimately reject promising (Pateman 1985, 136-141). Yet this suggestion does not do justice to the breadth of philosophical anarchism. Philosophical anarchism is a large tent comprised of among others, Christian anarchists such as Tolstoy, proto-Nietzscheans like Max Stirner, and modern-day consent theorists such as Simmons. Philosophical anarchist positions are not interchangeable – Wolff is not Godwin. Godwin's argument against promising amounts to an act-utilitarian objection to the binding nature of promises.² He believes that we must be able to change our plans at any given time; to bind myself now means to possibly forgo selecting the best option in the future (Godwin 1842, Book II, chapter ii). Wolff, on the other hand, considers himself a (not

wholehearted) follower of Kant.³ If Wolff ultimately rejects promising, it will have something to do with autonomy, not with Godwin's epistemic concerns or with appeals to utility.

This paper addresses two versions of the 'Wolff cannot accept promises' challenge. I have called these challenges the 'Alienation Critique' and the 'Promises are Promises Critique'. The *Alienation Critique*, put forth by Graham Baugh, levies the following charges against Wolff. First, it holds that Wolff's conception of autonomy precludes us from developing social conventions such as promising. What's more, even if we could develop some social conventions, Baugh believes that promising would not be one of them because he believes that all promises in a world of Wolffian agents would amount to cheap talk. This is because, on Baugh's interpretation of Wolff, all Wolffians would preserve their autonomy in order to pursue their own moral goals. Promises would be broken the moment they came into conflict with these goals. I argue that this interpretation is simply inconsistent with crucial elements of Wolff's moral theory. Although Wolff explicitly distances himself from Kant, the *Alienation Critique* unfairly exaggerates this distance.

The *Promises are Promises Critique*, expressed by (among others) Simmons, states that the very reasons that lead Wolff to argue against the validity of promises to obey the state also apply to promises writ large. What's good for the goose must be good for the gander – if my promise to obey the state violates my autonomy, so must my promise to meet my friend this afternoon. In other words, those who argue this position believe that Wolff's critique applies to a class of actions which contains all instances of promising. I will argue that this

critique conflates different types of promises. Although we may say 'I promise' in two cases, these promises may amount to very different types of symbolic action. Even though Wolff does not provide an explicit response to the Promises are Promises Critique, his work contains resources that would allow him to differentiate between types of promises. Some but not all promises will entail a violation of autonomy as he understands the term.

2. The Alienation Critique

Summary and description of Wolff

Baugh puts forth what I call the Alienation Critique (AC) in his 1985 essay, *The Poverty of Autonomy*, where he argues that Wolff cannot provide a coherent account of promising for two interrelated reasons. First, Baugh believes that Wolff's understanding of rational agency is so individualistic that it cannot account for the existence of intersubjective meanings and conventions. Promising is one such convention, so, if Baugh is correct, Wolff's autonomous individual either lives in a world without promising or cannot fully understand and participate in any convention of promising that does exist.

Second, Baugh believes that Wolff's autonomous person is characterized by a relentlessly instrumental rationality, one that pushes the individual to pursue her chosen goals as efficiently as possible. If this is right, then any promise will become void when the promisor either comes to have a different goal or when she decides that she can best achieve her goal by reneging on the promise. So even if we reject the first claim – that Wolff's conception of rationality is inherently too solipsistic to account for intersubjective meanings – Baugh will say that Wolff's ethical commitments still lead him to the same conclusion, namely that

promises are simply not possible in Wolff's framework.

Baugh's objections fundamentally turn on his understanding of Wolff's critique and partial appropriation of Kant. In the preface to *In Defense of Anarchism*, Wolff says that he has 'simply taken for granted an entire ethical theory' (Wolff 1976, viii). This statement, taken alongside his uncompromising insistence on the primacy of moral autonomy and his explicit reference to Kant, leads us to believe that he is something very close to a Kantian deontologist.

Wolff fleshes out his position in his subsequent work, *The Autonomy of Reason*, where he offers an extended and ambivalent assessment of Kant's moral philosophy (Wolff 1973).⁴ Kantian freedom as Wolff understands it means not having one's behavior determined by heteronomous, contingent forces. Rather than acting according to these forces, the free person's actions are determined by moral law.⁵ This moral law is *rational* law. There is for Wolff no real 'freedom' in the colloquial sense of the term – either we are obeying moral law or we are effectively getting pushed around by other forces. (Wolff 1973, 216) This is not necessarily depressing, Wolff agrees with Kant that we cannot know that we are not free (in some sense) in the noumenal realm (Wolff 1973, 222). He further agrees that we can presuppose that we are free in that sense, even if we can't say what exactly that kind of freedom amounts to (Wolff 1973, 217).

Wolff and Kant agree that while we cannot intuitively determine whether or not our aims correspond to rational law, we have recourse to the categorical imperative (CI) test as a

mechanism to verify whether or not a potential action (or instance of inaction) is rational. But here Wolff parts ways with Kant. While Wolff agrees that the CI test rules out actions that are impermissible, he disagrees with Kant's assertion that the CI test can generate *positive* moral law.⁶ He wants to retain Kant's demand for consistency while rejecting the position that rationality requires adhering to more substantive positions (Wolff 1973, 216-217). He is unyielding on this latter point: 'The real truth of the matter, I think, is that Kant fails to find a plausible argument for the validity of substantive moral principles because there simply are no such principles!' (Wolff 1973, 219). Yet he immediately follows this statement by declaring himself to be both a 'follower of Kant' as well as a consent theorist:

I am persuaded that moral obligations, strictly so-called, arise from freely chosen contractual commitments between or among rational agents who have entered into some continuing and organized interaction with one another. Where such contractual commitments do not exist, cannot plausibly be construed as having been tacitly entered into, and cannot even be supposed to be the sort that would be entered into if the persons were to attempt some collective agreement, then no moral obligations bind one person to another. (Wolff 1973, 219).⁷

Wolff's deontology is thus much thinner than Kant's. Wolff believes that all social obligations are necessarily contractual, and this is why he objects to Kant's positive duties, e.g., Kant's imperfect duty of beneficence. Wolff believes that these positive duties are not required by reason; they are the kind of duties that one may choose to

adopt, not duties that we *must* accept.⁸ But rather than try to provide substance where Kant supposedly fails to do so, Wolff simply says that positive duties to others only emerge as a result of contracting and thus cannot be specified in advance. The upshot is that while Wolff has much to say about what we ought to avoid, he leaves us without instruction as to what we ought to do other than to keep our word. Baugh seizes on this and argues that it leads to an overly individualistic understanding of morality. Kant's positive duties, the story goes, bind us to a common substantive moral framework in a way that Wolff's moral theory does not.

Baugh's critique

In this section I focus specifically on reconstructing the portions of Baugh's argument that support his position that Wolff cannot make promises. To begin, Baugh worries that Wolff understands rationality in purely instrumental terms. Baugh believes that Wolff thinks that autonomy is the *only* intrinsically morally relevant goal an agent may have and that 'Any agent who acts according to self-chosen policies, which are internally consistent and rationally connected to her goals, is autonomous in a Wolffian sense' (Baugh 2002, 115). In other words, Baugh interprets Wolff as saying that it doesn't matter *what* I do as long as my actions help me achieve whatever goal I happen to have. The fact that my goal may harm others is ostensibly irrelevant from this standpoint of Wolffian autonomy – a sensible statement if it's true that Wolff is only concerned with non-contradiction. But does this mean that discharging this duty to oneself will lead a Wolffian to engage in asocial or antisocial behavior? Baugh thinks so, and that this cannot be helped if the duty to preserve one's autonomy has

lexical priority over other duties and if the duties that each members of a community hold– duties which follow from idiosyncratic goals – end up in conflict.

Baugh's worries about Wolff's conception of autonomy go even further. Baugh believes that Wolffian autonomy ironically divests agents of all responsibility. The idea is as follows. We can presumably agree that we should not be held responsible for nonrational behaviors – our responsibility for an action is eliminated or at least severely attenuated if the action is the result of, say, a tic. But, Baugh says, since Wolff believes that we do not choose our goals or values – in other words, our conception of the good – we cannot (under a Wolffian framework) be held responsible for them (Baugh 2002, 116). Baugh understands Wolff to be saying that we can and must choose to pursue those goals as effectively as possible. We may be responsible for selecting subordinate goals that we believe will help us actualize the more fundamental goals, but if the fundamental goals and values are unchosen, we remain fundamentally unfree. Worse yet, these unchosen values and priorities may or may not take others into consideration.

If Baugh is right that Wolff believes that our primary duty to is act autonomously, it seems that we cannot blame someone who acts in an instrumentally rational fashion in the service of her fundamental goals or values. We may *oppose* her when her fundamental goals conflict with ours, but we cannot *blame* her. In fact, we should in some important sense praise her for acting autonomously and not kowtowing to *our* conception of the good.

Still, we may be disturbed by the thought that the very best we can hope for is either a miraculous convergence of 'given' goals or a tense *modus vivendi*. Baugh seems to believe that a promise in such a world would have little practical value and that I would renege on any promise that I made once it is to my advantage to do so. If my goals are unchosen, they can presumably change without any input on my part. Even if my fundamental goals do not change, perhaps changing circumstances (or even further reflection on my part) will lead me to believe that I should undertake a new course of action that would render my promise void.

Not that we would even get that far. Baugh believes that Wolff's 'extreme' individualist conception of rational agency prevents agents from even communicating intelligibly with each other. As he says,

Wolff cannot provide a coherent account of obligation either, for obligation arises from various social practices, such as promising, and Wolff's purely individualistic and abstract analysis of rational agency simply cannot provide a coherent account of any social practice, for social practices are constituted by intersubjective meanings and impersonal rules which transcend pure, subjective rational agency and arbitrary choice. (Baugh 2002, 117)

This is Baugh's most radical critique. If he is right, then a Wolffian would not even ge

the chance to break her promises willy-nilly, since she would not even be able to understand what a promise *is*. In other words, Baugh believes that Wolff's conception of autonomy presupposes that we each have a mental model that is 'given' and unable to adapt to the mental models of others, except insofar as it can take external stimuli and process them from its own perspective. Baugh seems to be saying that a Wolffian would say that our mental models differ so much that what might seem like conventions would lack sufficiently shared meaning to make them worthy of the name. It is not clear just how far Baugh takes this line of thinking – does he believe that Wolff's position entails that there is no possible *agreement*? Or does he believe that Wolff's position amounts to something close to Wittgenstein's famous statement that 'If a lion could speak, we could not understand him' (Wittgenstein 1958, 223)? Whichever interpretation Baugh favors, we come to the same conclusion: if Baugh is correct, Wolff cannot endorse promising.

While this does not exhaust Baugh's full critique of Wolff, it shows why Baugh believes that Wolff's position precludes promising. To re-iterate and summarize: Baugh believes that moral discussion is impossible in a world of Wolffian agents. Wolff's apparently eccentric commitment to an odd form of autonomy – the efficient pursuit of unchosen goals – leads to an overly individualistic understanding of morality. Any promise becomes cheap talk since one's fundamental goals or the best strategies to achieve one's goals could change without notice. Worse yet, because Wolff's agents (as Baugh understands them) may have radically different goals – both moral and practical – they may end up being unable to communicate intelligibly with

each other.

Fortunately for Wolff, however, both of these interpretations fail to adequately describe his position. In the following subsection, I hope to show that Baugh is unfairly imputing to Wolff an overly bizarre theory of mind. Wolff's autonomy is far more banal than Baugh makes it out to be, and this saves him from Baugh's critique.

Response to Baugh

In this subsection I respond to the foregoing objections in an expansive manner. I hope to not only respond to Baugh specifically, but to use my response to him as a platform to deal with other potential objections.

Baugh's critique is ultimately unpersuasive. Most obviously, Baugh is mistaken when he says that Wolff's autonomous person would break a promise if doing so were instrumentally required in order to pursue another goal. This mistake stems from his unfair critique of Wolff's use of the CI test. Recall that for Wolff, adherence to the CI test does not generate universal positive duties. Baugh believes that this would lead people to hold incompatible goals, with each person morally obligated to try to actualize her own goals.

Yet Baugh does not seem to appreciate that the test nonetheless does effectively stop a follower of Wolff from breaking promises. Even if the autonomous person's goals are, so far as we can tell, completely arbitrary, the autonomous person judges the

acceptability of her own goals by the CI procedure and may do so because the CI test ensures that the person acts according to the principle of non-contradiction.

According to Wolff, any goal that satisfies the CI procedure prior to any promises being made is acceptable, but the pursuit of a goal – would be inadmissible if doing so would lead to the violation of a promise. To use a banal example: it's fine for me to spend next Monday napping or meeting with students in order to give them extra preparation for an upcoming exam, but once I agree to meet with students, I cannot unilaterally renege and decide to nap instead. My decision to meet the students is not an indictment of napping in general; it just means that I can't nap on that particular day. Perhaps my unilateral decision to nap would be defensible if I realized (after making the promise to my students) that my napping was necessary to satisfy some prior commitment that I made, for instance that I be well rested for a late night helping my daughter with her science project. But this is not the arbitrary renegeing that worries Baugh. He is worried that I will renege simply because my attitude changed on Monday morning or that I decided that napping would be the right thing to do all things being equal. Yet this worry is unfounded – running the “napping maxim” through the categorical imperative test after making the promise would yield a contradiction that would not have existed had I ran an analogous “napping maxim” through the test prior to making the promise. Wolff's position is that my attitudes and goals inform the promises I make, but once a promise is made, my status as an autonomous agent forces me to abide by the terms of the promise.

Baugh is mistaken when he says that Wolffians cannot make agreements. What he should be saying is that *some* agreements are not possible because they would predictably bind us to positions that would force us to violate our other moral obligations. Many of these obligations come about due to prior agreements. The only way for Baugh's critique to hold is if Wolff held the position that the values and goals of a given individual could change⁹ in such a way that the individual would find the CI-test silly or irrelevant. Such a person could break promises with impunity, since she would not longer be bound by the non-contradiction clause of the CI test. Unfortunately for Baugh, Wolff is quite clear that this would not happen. Or, more precisely: if I experienced such a conversion, my moral outlook would simply be wrong. The CI-test determines whether or not I my actions conform with freedom; I cannot 'freely' choose to discard it. Doing so would mean to act heteronomously, and thus fail to uphold my autonomy in a Wolffian sense.

The objection that Wolff's autonomous person can switch goals arbitrarily is also odd.¹⁰ Why is it a problem if a follower of Wolff changes her mind on morally pertinent issues? In cases where the change is self-regarding, it is unproblematic. But what about cases that involve a change in how I believe I should treat others? Again, Wolff seems clear on this point: if someone decides that her goals were ill-advised – perhaps she has come to revise her theological convictions – she may pursue other goals so long as these goals pass the CI test and do not violate existing agreements.¹¹

In the extreme, perhaps this does involve renegeing on some agreements. Yet even this

is not necessarily a problem. We should be happy when former racists renege on their sincerely made commitments to try to foment a race war. What's more – and I will address this point in the following section of this paper – some moral commitments cannot be kept or cannot be made. While there may indeed be tragic situations where someone is bound by prior commitments that she now finds unethical, part of being a conscientious moral agent is being careful when making commitments. To constantly face these dilemmas is to fail to take one's moral duties seriously when one is making agreements in the first place.

What's more, Baugh's worry that an individual might suddenly shift her goals or values is overly theoretical. First, most people have remarkably stable – or at least slowly evolving – ways of understanding the world. Worrying about sudden value shifts in those around me is analogous to worrying that one of those people might suddenly snap and attack me. People *do* snap, of course, but fortunately those cases are rare enough for us to generally live our lives without seriously attending to that possibility.¹²

Second, even if an individual (1) suddenly shifted goals and values, and (2) were strictly instrumentally rational in the way that Baugh describes, she would nonetheless recognize the social value in maintaining consistent behavior. Indeed, if Baugh is right and Wolff's autonomous person is instrumentally rational, then this very fact demands that she act consistently in most cases. In other words, her formulation and implementation of an instrumentally rational plan must take into

account the very large value in being seen as someone who acts according to consistent principles. We need others to accomplish our goals – especially our moral goals – and those who constantly renege on agreements will fail to get others to cooperate. Even the caricatured Wolffian that Baugh describes would, by the very fact that she is such a calculating machine, act consistently in order to get the support she needs to make her *real* dreams come true.

Finally, I admit that I do not fully understand Baugh's contention that Wolffian agents could not enter into agreements because they would literally be unintelligible to one another. Pluralism – both value and linguistic – does not prevent the development of shared language. Even if two people do not share *umwelts* and aren't 'really' communicating in a deep way, they can still communicate in such a way that their behaviors are mutually predictable. In fact, Baugh seems inconsistent on this point. On the one hand he seems to think that Wolff's position precludes social conventions of any kind, and on the other he says that

Any moral discourse between such agents will consist of competing but incommensurable and mutually unintelligible claims. In Wolff's anarchist community there could be no meaningful debate or discussion of collective choice of ends. There could be a limited rational discussion of policies but only in an instrumental sense [...] (Baugh 2002, 113-114).

Clearly this type of discussion would not be possible if Wolff were as radical as Baugh says. But it's just not the case. Wolff is committed to a principle of consistency that prevents wanton promise-breaking. In fact, Wolff's concern with non-contradiction would make him even more sensitive than most to the evils of promise-breaking.

3. The Promises are Promises Critique

Are all promises alike? Or, how to fail at promising.

The second critique is what I call the Promises are Promises Critique (PaPC). In formulating it, I draw heavily on Simmons' relatively brief remarks on the matter in his essay, *Philosophical Anarchism* (Simmons 1996, 36-37). Simmons begins by pointing out that Wolff says that not only can a state not unilaterally declare itself to be legitimate, it does not gain legitimacy even if we promise to obey its commands (Wolff 1976, 14-15). According to Wolff, promises in this case do not bind because they would violate our duty to remain autonomous by binding our actions to the will of others. A promise to obey authority is possible; it's just not morally permissible (Wolff 1976, 14-15).

Simmons complains that Wolff does not say why promises to the *state* are especially egregious to autonomy. After all, if I were to promise my friend to help her do something, my friend would have authority over me, at least to some extent. To promise is to *authorize* – to give authority to – someone or some group to demand action on our part. To expand on Simmons a bit, I want to add that open-ended promises are especially problematic. 'I promise to let you decide what we do tomorrow evening' is both a common and formally terrifying promise; after all, how do I know you won't decide to forgo watching a movie in

favor of robbing a bank or, more legal but no less repugnant, have us write angry letters to your ex spouse? For someone like Simmons, Wolff can only be consistent if he rejects the legitimacy of *any* promises for the same reasons that he rejects state legitimacy. Simply put: if a promise to obey the state is invalid because it would violate our autonomy, what gives us the right to make any promises at all?

Simmons' challenge is in one sense quite odd. More precisely, it's odd that this type of challenge comes from Simmons. He is after all a consent theorist per excellence – the central theme running throughout his body of work is that consent is the only defensible basis of state legitimacy. Consent theorists have long acknowledged that acts or utterances that are understood to signal consent may fail to do so, and Simmons is obviously aware of these restrictions. It is thus surprising that he would assume that a supporter of Wolff would think that all promises are created equal. They clearly are not. We can see this by considering promises that are, at the moment of their conception, stillborn.

Theorists differ on the particulars, but Keith Hyams has recently cataloged a number of ways that an act of consent – including a promise – may fail to 'come off'. It may fail if an individual is generally too incompetent to give consent in any case, if the person does not understand that her words or actions actually signaled consent in a given case, if an individual's alternative options are sufficiently unpalatable, or because an individual is not morally authorized to give consent given the substance or scope of the attempted agreement.¹³ (Hyams 2011, 112-119) Before accusing Wolff inconsistency, we should investigate the possibility that his position amounts to saying that a promise to obey the

state fails for one of Hyams' reasons.¹⁴

If Wolff is right, then the promise to obey the state must fail because it violates Hyams' final criterion, as the other reasons clearly don't apply. Going down the list, (1) We can assume that most people are competent enough to give consent in a general sense. (2) An individual who says "I consent to obey the laws of this government" surely knows what those words mean.¹⁵ (3) A Humean objection that we have no *real* (practical) choice but to obey the law misses the point here. Wolff does not base his rejection of promising on the grounds that we have no real choice but to obey; he believes that such a promise would fail even if a person *wanted* to make it (Wolff 1976, 16-17).¹⁶ If promising in this case is objectionable to Wolff, it must have something to do with either the scope or the content of the promise.

We can clarify this further by taking up the argument of Wolff's most persistent critic, Jeffrey Reiman. Shortly after Wolff published *In Defense of Anarchism*, Reiman responded with an extended rebuttal entitled *In Defense of Political Philosophy*. Reiman argues (among other things) that Wolff objects to what Reiman calls a "wild-card morality", which Reiman represents with the symbol *. According to Reiman, Wolff objects to a promise to do * if someone else may determine the content of * in the future. Reiman agrees with Wolff that any such promise must fail and is inconsistent with the exercise of moral autonomy:

Since specification of the source [of the content] does not specify the moral nature of the action to be substituted for the 'wild card,' *such a morality entails a moral obligation to do an action with indifference to*

the moral nature of that action, which is absurd (Reiman 1972, 3).

But, Reiman continues, political legitimacy comes not from a promise (and thus he agrees with Wolff on this point) but on a belief that we are better off with a system of laws than without it (Reiman 1972, 3-9).¹⁷ Reiman thus agrees with Wolff that we cannot give broad consent to obey the laws while also saying that Wolff's whole approach to political legitimacy is doomed from the start. And while Reiman's objection can and should be taken up elsewhere, his idea of "wild-card morality" helps Wolff here.

Reiman's belief that it is absurd to promise to * lines up with Hyams's final case of failed consent. In Hyams' account, a promise may fail due to its substance or scope in a variety of ways. For instance, a promise is invalid if the promise's jurisdiction requires prior authority that the promisor does not have. I can promise my students that they will pass my class if they turn in all their assignments and write sensibly, but I cannot promise that this minimal level of work will be sufficient to pass my colleague's class. To take another example, one injunction against suicide states that we do not have the authority to take our lives because doing so would be a violation against God (e.g., Locke 1980, 9). To those who hold those metaphysical beliefs, my promise to kill myself if I don't win a medal in the next Olympics would be void.

What's more, a promise may fail because we believe the content of the promise to be *evil*. By some – hopefully many – accounts, a promise to kill my dog if I lose a bet would be void because the substance of the promise requires me to violate a value that antecedes the

promise.¹⁸ This would hold even if I lived in an area which legally allowed dog owners to kill their pets at will. Analogously, the problem with a promise to obey the state is that it is a wild-card promise that may require us to either violate the promise in the name of morality or to perform an unjustifiable action.

Attenuated conception political promises

A critic might agree that Wolff is in effect arguing that a promise to * is invalid because (1) its scope is too large, (2) it may commit someone to perform immoral acts, and (3) it directly violates the duty to maintain and act upon one's status as an autonomous being. Yet, this critic might continue, both Wolff and Reiman are mistaken when they say that a promise to obey the state means the government can fill in * with whatever it wants. As numerous commentators have pointed out, few defenders of state legitimacy seriously believe that this is what a promise to obey amounts to (Beauchamp & Witkowski 1973, 536-538). Consent to obey the government is usually understood as a commitment to obey *limited* government, one that does not have an unlimited scope when determining *. So it seems that if Wolff objects to promises to * on the basis of *'s scope, the objection can be overcome if the promisor is the one who can set the limits of of *'s scope. But a supporter of Wolff might reply that this does not obtain when promising to obey the state. A promise to obey a state is a promise to * or a promise to perform certain actions with the understanding that the state can change those actions without rendering the promise null and void. We don't just promise to obey the state as it is; we also implicitly promise the state as it will be.

The upshot is that a promise made to a state is at least in this respect a fundamentally different undertaking than a promise to a friend. When I promise my friend to perform an action, I determine the boundaries of that action. When I promise the state to obey the law, I do so in a context where the state claims the right to modify the scope of the law.¹⁹ The state may in some cases be *justified* in making this modification, but this does not imply that I have a duty to agree to take part in such an arrangement, or even that consent to enter such a lopsided agreement should count as consent at all. More importantly, any agreement that takes place is no agreement between equals, or even between non-equals that have some kind of similar bargaining power. It is reasonable to refuse to consider any agreement legitimate, however substantively benign, when one participant is an all-powerful institution that has the de facto ability to defect or modify the terms of the deal at any time.

Promises and the social meaning of agreements

Those who defend the legitimacy of the state often use a particular dodge that tries to preserve the ghost of legitimacy while conceding nearly everything to those philosophical anarchists – those who superseded Wolff – who say that political obligation is possible but rare. These defenders extend the argument that we only consent to *limited* government by saying that we also consent to obeying as long as there are no clear countervailing reasons that require disobedience. The fact that something is illegal, they go on to say, means that there must be at least a *prima facie* to obey the law. The weight of this obligation is difficult if not impossible to specify, but it would be greater than nil and far below a trump (Smith 1973). These defenders of legitimacy want philosophical anarchists to admit that the law carries at least a vanishingly small amount of weight. They want an admission that the law

is supreme even if this admission has little or no practical bearing on actual moral decisions.

This represents the ultimate rapprochement between the a posteriori philosophical anarchist and “archist” positions. Philosophical anarchists like Simmons say that adopting philosophical anarchism does not entail a change in one's behavior (Simmons 1996, 20). Defenders of political legitimacy who concede that morality may require one to frequently disregard of the law are essentially saying the same thing but from the other side: the fact that something is *legal* might not have any bearing on any of our practical decisions. Something's being the law would have as much force as the Queen of England's opinion on a crisis in Canadian politics.

It is important to note that this is *not* what Wolff is proposing. Wolff's position is incompatible with such a rapprochement. There is a difference between breaking a promise that has weak obligatory force and deciding that a promise is void from the start. Wolff is saying that that the promise to obey is void, *not* that the promise is obligation-generating but overcome by weightier considerations. Wolff's position is the only one that gives philosophical anarchism any actual substance.

Finally, Wolff's critics may retort that the fact that most people understand 'She promised to obey the law' as being both sensical and duty-generating shows that the promise does is not inherently void. To make such a promise, in other words, is not to utter an absurdity. On the contrary, these critics may continue, such a promise gives others legitimate

expectations that the promisor takes her political duties seriously. These expectations, the argument goes, demonstrate the validity of the promise.

Wolff's critics are right that the scope of what counts as a promise, including what others can legitimately expect as a result of the promise's being made, is socially determined. When my friend tells me 'If anything goes down in the bar, I've got your back', he may mean something very different than when Joe says the same thing to Steve. My friend and I implicitly take for granted that the promise is void if I get drunk and begin to pick fights with the other patrons. My friend *meant* that he would help me (to a certain extent) if someone became aggressive toward me or if I wanted to impress someone romantically.

This might not at all be what Joe meant when he made his promise to Steve. In fact, he may have meant precisely that he would be happy to help if Steve became belligerent. Our respective histories – both individual and shared – establish the implicit boundaries of the promise. Similarly, every competent member of a community understands the social significance of some actions. These community members may excuse an outsider for making a mistake (say, not realizing that her actions created legitimate expectations on the part of the community members), but to the extent that they do so they will view this outsider as being socially incompetent. Wolff's agent, the critic might conclude, can only reject the social significance of the promise to obey the law at the cost of announcing their own social incompetence.

This critique fails because it puts too much weight on the proposition that people

understand what I say when I make a general promise to obey the law. We should read Wolff as saying that a promise to obey the state is null from the start because the vast majority of people are making a conceptual – not necessarily a moral – mistake when they consider what it means to promise to obey the law.²⁰ People are not speaking gibberish when they make such a promise. Most of us have a feeling that we know what someone means when they make such a promise. Yet if we were to ask these people if a promise to * (as described by Reiman) would be valid, I imagine that most would say no. So what is going on here?

Wolff's statement about our emotive attachment to the laws of our country provide a clue to resolving this tension:

When I take a vacation in Great Britain, I obey its laws, both because of prudential self-interest and because of the obvious moral considerations concerning the value of order, the general good consequences of preserving a system of property, and so forth. On my return to the United States, I have a sense of reentering my country, and if I think about the matter at all, I imagine myself to stand in a different and more intimate relation to American laws. They have been promulgated by my government, and I therefore have a special obligation to obey them. But the anarchist tells me that my feeling is purely sentimental and has no objective moral basis (Wolff 1976, 18-19).

I believe that Wolff would say that a promise to obey the law is actually a demonstrated commitment to obey the law on some *other* ground, usually pre-reflective emotion.²¹ Most people likely believe that we have an obligation to obey the law regardless of the promise, even if the promise strengthens the associative ties of the members of the community. But – and this is the crucial point – this means that those who say that Wolff cannot make promises have missed their mark. Wolff does not object to all promising. He objects to Reiman's 'wild-card' promises, and he (I believe) would say that promises to obey the state are a different *kind* of performative action than are promises to meet a friend for dinner. Political promises amount to a demonstration of one's commitment to an already existing duty.

Wolff does not believe that this duty exists. Many others do. If it does exist, however, it has nothing to do with promising and thus a different critique of Wolff altogether. If I say that the state is legitimate and I promise to obey it, a Wolffian would say that the promise is void in any case but that I am compounding my error by renouncing my status as an autonomous being by virtue of announcing (not promising) my subservient status.

4. Conclusion

In this paper I sought to address and dispel the argument that Wolff's radical defense of autonomy precludes him from endorsing promising as an institution. His conception of the person is sufficiently robust to allow for convention and cooperation. What's more, although his conception of morality is thinner than Kant's, he nonetheless would not allow people to break their promises willy-nilly. Finally, he picks up on the very plausible

intuition that a promise to obey the state is not 'just another promise'. There are promises and there are *promises*. As we have seen, some promises simply do not 'come off', and others aren't really promises as we normally understand them. For Wolff to say that promises to the state never succeed does not commit him to the position that *all* promises must fail.

If my arguments are sound, supporters of Wolff and *a priori* philosophical anarchism more generally will find their burden lessened. Those supporters still have work to do; the charges against Wolff are varied and scattered. My hope is that this paper leads to the dismissal of one of the most prominent and disturbing of these charges. There may be hope yet for philosophical anarchists who feel that the new breed of philosophical anarchism is too heavy on the philosophy and too light on the anarchism.

References

- Ball, Terrence. 1995. *Reappraising Political Theory: Revisionist Studies in the History of Political Thought*. Oxford: Oxford University Press.
- Baugh, Graham. [1985] 2002. In *The Anarchist Papers*, edited by Dimitrios Roussopoulos, 107-122. Montreal: Black Rose Books.
- Beauchamp, Tom L., and Ken Witkowski. 1973. "A Critique of Pure Anarchism." *Canadian Journal of Philosophy* 2 (4): 533-539.
- Brown, Penelope, and Stephen C. Levinson. [1978] 1987. *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press
- Dahl, Robert A. 1989. *Democracy and its Critics*. New Haven: Yale University Press.
- Frankfurt, Harry G. 1973. "The Anarchism of Robert Paul Wolff." *Political Theory* 1 (4): 405-414.
- Godwin, William. [1793] 1842. *An Enquiry Concerning Political Justice, and its Influence on General Virtue and Happiness*, 4th Ed. London: J. Watson.
- Horton, John. 1992. *Political Obligation*. Atlantic Highlands: Humanities Press.
- Hyams, Keith. 2011. "When Consent Doesn't Work: A Rights-Based Case for Limits to Consent's Capacity to Legitimise." *Journal of Moral Philosophy* 8 (1): 110-138.
- Laszlo, Ervin. 1973. "A Moralizing Note to Professor Wolff's Reply." *The Journal of Value Inquiry* 7 (4): 307-308.
- Locke, John. [1689] 1980. *Second Treatise of Government*, edited by C.B. MacPherson. Indianapolis: Hackett Publishing Company.
- Martin, Rex. 1974. "Wolff's Defense of Philosophical Anarchism." *The Philosophical Quarterly* 24 (95): 140-149.

- Pateman, Carole. 1985. *The Problem of Political Obligation*. Berkeley: University of California Press.
- Perkins, Lisa S. 1972. "On Reconciling Autonomy and Authority." *Ethics* 82 (2): 114-123.
- Pitkin, Hanna. 1966. "Obligation and Consent II." *American Political Science Review* 60 (1): 39-52.
- Pritchard, Michael S. 1973. "Wolff's Anarchism." *The Journal of Value Inquiry* 7 (4): 296-302.
- Raz, Joseph. 1979. *The Authority of Law: Essay on Law and Morality*. Oxford: Oxford University Press.
- Reiman, Jeffrey. 1972. *In Defense of Political Philosophy*. New York: Harper & Row.
- Reiman, Jeffrey. 1978. "Anarchism and Nominalism: Wolff's Latest Obituary for Political Philosophy." *Ethics* vol.89 (1): 95-110.
- Sanders, John T. 1996. "The State of Statelessness." in *For and Against the State*, edited by Sanders, John T. and Jan Narveson, 255-288. Lanham: Rowman & Littlefield.
- Simmons, A.J. 1979. *Moral Principles and Political Obligations*. Princeton: Princeton University Press.
- Simmons, A.J. 1996. "Philosophical Anarchism." in *For and Against the State*, edited by Sanders, John T. and Jan Narveson, 19-40. Lanham: Rowman & Littlefield.
- Smith, M.B.E. "Is There a Prima Facie Obligation to Obey the Law?" *The Yale Law Journal* 82 (5): 950-976.
- Sterba, James P. "The Decline of Wolff's Anarchism." *The Journal of Value Inquiry*, 11 (3): 213-217.
- Wittgenstein, Ludwig. [1953] 1958. *Philosophical Investigations*, translated by G.E.M.

Anscombe. Oxford: Blackwell Publishers.

Wolff, Robert Paul. 1973. *The Autonomy of Reason: A Commentary on Kant's Groundwork of the Metaphysics of Morals*. New York: Harper Torchbooks.

Wolff, Robert Paul. [1970] 1976. *In Defense of Anarchism: With a Reply to Jeffrey Reiman's In Defense of Political Philosophy*. New York: Harper & Row.

Wolff, Robert Paul. "One-Dimensional Man – A Mini-Tutorial Part One." The Philosopher's Stone. <http://robertpaulwolff.blogspot.com/2011/10/one-dimensional-man-mini-tutorial-part.html> (accessed September 15, 2015).

This rehabilitation has occurred in fits and spurts.

2 When we promise, he believes, “[...] we abridge, and that is the most essential point, the time of gaining information, if we bind ourselves today to the conduct we will observe two months hence. He who thus anticipates upon the stores of knowledge, is certainly not less improvident than he who lives by anticipating the stores of fortune” (Godwin 1842, Book III, chapter iii). I should note that Godwin's position amounts to more than a crude utilitarianism based solely on rational calculation. He also subscribes to a perfectionist view similar to that of J.S. Mill: when it comes to our capacity to think and make decisions, we need to use it or lose it.

3 I address his divergence from Kant in the next section of this paper.

4 Incidentally, many of Wolff's critics make no reference to this work, even though it provides crucial detail and in some cases revisions to Wolff's position in *In Defense of Anarchism*.

5 In spite of his sympathy with the struggle of marginalized groups, Wolff uses the term “free man” throughout his text. I replace “man” with “person” throughout this paper.

6 The critique that Kant's ethical theory cannot provide adequate substantive principles is of course well established, and has been since at least Hegel's *Philosophy of Right*. Wolff makes no mention of this. This is unsurprising when one considers Wolff's longstanding anti-Hegel stance. From his “mini-tutorial” of Marcuse's *One Dimensional Man*: “Faithful readers will know that I have an allergic reaction to Hegel, so I consider it an evidence of my admirable broadmindedness that I am willing to take Herbert's works seriously, as I do” (Wolff, *The Philosopher's Stone*).

7 Readers interested in consent theory should note that Wolff's last sentence blurs the line between different types of consent. He begins by saying that there are no binding obligations if people have not *actually* entered into an agreement with each other. By the end of the sentence he is referring to what Pitkin coined hypothetical consent – a pseudo-consent purportedly justified by a belief that the agent *would* have consented had she been fully informed, had access to her intellectual faculties, etc (Pitkin 1966).

8 Take for instance Wolff's commentary on the purported duty to cultivate one's talents: “Nothing much need be said about this argument for cultivating one's talents. Considering the oppressive odor of relentless moralism which clings to the passage, it is just as well that the argument is so obviously bad” (Wolff 1973, 169).

9 I am using the passive voice here quite deliberately. Hopefully I have shown by now that Baugh doesn't think that a Wolffian agent could deliberately revise her fundamental goals and values.

10 Indeed, it would be more accurate to say that the goals would switch on the person, seeing as how Baugh believes that Wolff that we have no choice in the matter.

11 Again, this provides support for Wolff's implicit position that some agreements are invalid from the outset.

12 Those who habitually snap are, generally speaking, already understood to be “the type of person who snaps”. Thus while they might go from calm to aggressive quite quickly, we are less and less surprised each time they shift their demeanor.

13 For instance, I cannot normally consent that you break into my neighbor's house and make off with his television set.

14 There is of course disagreement about the quality, scope and limit of each of these restrictions. For instance, should it matter whether or not my options are limited as a result of coercion or mere bad luck? My point is not that these are settled issues, but rather that the debate is about the scope and not the existence of constraints on consent.

15 There are two points to mention here. First, the situation may not be as clear-cut as it might appear – I discuss the meaning of this type of consent later in this section. Second, I leave out cases of tacit consent, which are far more complex. In such cases, not only do we need to establish that the tacit action (or non-action) could constitute consent, but also that individuals are aware that performing this action (or non-action) constitutes consent.

16 Like other failed promises, the point isn't that the promise is unintelligible, but rather that it is morally indefensible. That said, it is possible that new moral responsibilities when one's (indefensible) promise causes others to modify their behaviors with the expectation that the promise will be fulfilled.

17 As Reiman unfolds his argument, he begins to look very much like a rule utilitarian. Indeed, Wolff simply assumes in his *Response to Reiman* that this is Reiman's position. Reiman however rejects this label (Reiman 1978, 105).

18 As with Hyams' restrictions, I do not want to imply that I think that the question of what counts as evil has been settled. I am pointing out that there exist some actions which do not violate appropriate jurisdictions but are nonetheless void because of the nature of the promised act.

19 Assuming, as Godwin noted, that it even makes sense to say that I've consented to obey the giant and ever-changing

corpus of law, the vast majority of which I will never get a chance to read. It might be more accurate to say that I consent to waive my effective right to consent to obey the law.

²⁰ It is possible to make both types of errors simultaneously. I may be led to a conceptual error by evil motives. That said, Wolff does not give us enough to speculate what proportion of people who believe in state legitimacy are being immoral – abdicating the moral responsibilities – and how many are being intellectually sloppy.

²¹ That said, it is possible that some of these people's belief that we ought to obey the law comes from deep reflection. These people, who may believe that we have a duty to obey the law for reasons of fair play, natural duty, utility, gratitude, etc., could still hold that emotion or promises may further reinforce the commitment to obey.