Reasons Internalism, Cooperation, and Law

I shall paint a surprising picture. First, I shall present a reasons internalist argument for thinking that many moral norms depend on agreements between agents, as agents all have a fundamental reason to cooperate.[1] Then I shall argue that if one construes some of the moral norms that depend on agreement as laws, slotting them into the theoretical framework natural law theorists usually defend, a natural law theory looks surprisingly attractive. The cooperation-based norms can solve some fundamental theoretical problems for natural law theories. I shall not, however, endeavour to conclusively defend natural law theory, for I am not certain about whether we should prefer a natural law framework to some other one, such as a legal positivist framework. My fundamental aim is, instead, to paint the picture.

To do so, in section 1, I introduce my preferred interpretation of reasons internalism. In the lengthy section 2, I present an argument which suggests that a desire to cooperate with other cooperative agents is partially constitutive of ideal agency. By reasons internalism, this desire can explain moral norms. In section 4, I transpose these norms into a natural law framework and show how the emerging framework is attractive. I conclude in section 5.

(1) Reasons Internalism

Reasons internalists take there to be a necessary relation between a reason for action and a person's psychology, whereas externalists deny that this is always the case. I think internalism is correct, but I do not have the space to defend it, so I shall just assume it and my favourite formulation of the view.[2] This formulation is:

> (*REASONS INTERNALISM*) For all $r(F,A,a,C)$, $r(F,A,a,C)$ is a reason relation holding between a fact F and an agent A's action a in circumstances C iff (and because) $r(F,A,a,C)$ holds in virtue of the desires that feature in P's idealized psychology.[3]

To clarify: when I mention a fact F, I am indifferent between talking about a (true) proposition, an instantiated property, or whatever else one may take a fact to be (cf. Peter, 2019). Furthermore,

[1] This argument is developed in greater depth in Leffler (2019).
[2] See, however, Leffler (2019) for that too.
[3] $r(F,A,a,C)$ may hold between more *relata*, e.g. times, as well. But I stick with this formulation for simplicity.

I shall not make any commitments about how an action *a* works, but I shall discuss *A* and *C* in more depth below.

Moreover, I shall not discuss the 'in virtue of'-relation that holds between idealized desires and reasons in much depth. There are many theoretical options about what this relation might be – e.g. constitution, or even identity – but for the sake of convenience, I shall usually say that agents' reasons have their sources in, or are grounded in, ideal desires (cf. Chang, 2009; 2013). For present purposes, nothing turns on what one thinks here.

But what does 'the desires that feature in *P*'s idealized psychology' mean? I shall explicate this aspect of *REASONS INTERNALISM* at some length, for it will be important in my argument below. First of all, *REASONS INTERNALISM* is a particular type of desire-based reasons internalism. While not all forms of internalism must look like it, let alone be desire-based, this is still probably the most common version of reasons internalism in general (cf. Williams, 1981; Joyce, 2001; Smith, 1994; 1995; 2012*a*). On this view, the agent *A* whose desires explain reasons has desires, beliefs, and is otherwise suitably rational – not least instrumentally rational. I shall also assume that this psychology should be given a functionalistic interpretation; in particular, the function of beliefs is to represent the world accurately, the function of desires is to motivate the agent to act, and (instrumental) rationality functions to make the agent takes the best means (that she believes there are) to her ends (set by her desires). From here and on, I shall follow convention and call fully idealized agents of this kind '*A*+', while non-idealized ones still will be called '*A*'.

It is the desires of *A*+-style agents, suitably idealized, that explain reasons for *A*. As *A*+ has beliefs, desires, and is rational, and desires do the key explanatory work here, the core idea behind *REASONS INTERNALISM* is that if *A*+ desires to φ, then *A* has a reason to φ. And the idealization of *A*'s psychology, i.e. that which turns *A* into *A*+, plays a supporting role in ensuring that *A*+ has the *right* desires to explain *A*'s reasons. Perhaps some of our desires are short-sighted and incompatible with some of our other, deeper, desires. Or perhaps some of our desires are based on faulty information. But in better conditions, our desires might explain our reasons better.

How may we reach better conditions? As *REASONS INTERNALISM* explains reasons by appealing to the desires that feature in an idealized psychology, and that psychology is functionalistic, I shall take an ideal agent to be a fully functional agent, in the sense that all the psychological states and capacities she has *qua* agent function fully, and that she manifests these capacities fully insofar as she acts. I shall also assume that, *qua* ideal agent, she has or is in the right background conditions to be able to exercise those well-functioning psychological states and capacities. Here, a 'background condition' is any fact about the agent or the context she is in which may affect her psychology, decisions, or actions.

To be idealized, then, *A*+ is supposed to feature all the properties of *A* just mentioned, fully functioning or fully manifested when acting. So, for example, the idealized agent does not just have the capacities for believing, desiring, or instrumental rationality, but actually is instrumentally rational insofar as she acts (cf. Smith, 2012*b*). This idealization condition rules out the possibility that the mental states involved in idealization are blocked so the agent is unable to make use of them, e.g. by accidie (cf. Hurtig, 2006).

Furthermore, full idealization requires that *A*+ has or is in the background conditions that allow her to have the states or capacities of her psychology fully functioning or manifested when acting. The point here is that the agent cannot be in background conditions that hinder her psychology from living up to its functions. For example, insofar as having capacities to deliberate requires ways of applying the instrumental principle in deliberation, and deliberation is something that an ideal agent may do, having the ability to apply the instrumental principle in those ways is an idealizing condition of the agent. As Williams (1981) famously noted, instrumental rationality need not just involve the taking of already known means to given ends, but can also involve deliberation about how to do so by finding constitutive solutions in cases where desires conflict, using one's imagination to find new possible solutions, etc.

This kind of background conditions idealization can also explain why *A*+ ought to be epistemically refined. Epistemic refinement does not just mean that *A*+'s beliefs are fully functional or manifested so that she satisfies the aim of belief, viz. represents the world accurately. Just by assuming the full functioning of beliefs, I shall assume that, insofar as the ideal agent has beliefs, these live up to their aim. However, many authors have also tended to assume that the agent must have some particular set of beliefs, e.g. all relevant true beliefs and no false ones (Smith, 1994, ch. 5). I am not sure exactly what set of beliefs matters here, but it is plausible that there is a set like that, for it is plausible that *A*+ needs fairly many true beliefs to manifest her agential capacities fully insofar as she acts – without them, she might take the wrong means to her ends. And if there is such a set, *A*+ has it.

Summing up, then, idealization involves fully functioning or manifested capacities *as well as* relevant background conditions. The latter is also what explains why *A*+ is able to apply the instrumental principle in complex ways and is epistemically refined.

(2) Cooperation

Assume, then, the version of reasons internalism that I have formulated. It will allow us to sketch an *argument from idealization* for a certain conception of morality. I shall try to present it and motivate

its premises, though I do not have the space to respond to all potential objections to the view here. The argument goes like this:

(1) If *A*+'s psychology is able to explain the reasons of an agent *A* in our world, then *A*+ is suitably idealized.
(2) If *A*+ is suitably idealized, then *A*+ has a set of idealized desires (based on *A*'s desires) for what to do in a range of situations or circumstances, many of which feature the circumstances of justice.
(3) If *A*+'s psychology is able to explain the reasons of an agent *A* in our world, then *A*+ has a set of idealized desires (based on *A*'s desires) for what to do in a range of situations or circumstances, many of which feature the circumstances of justice.
(4) If *A*+ has a set of idealized desires (based on *A*'s desires) for what to do in a range of situations or circumstances, many of which feature the circumstances of justice, *A*+ must have a desire to cooperate with other cooperative agents as a matter of being suitably idealized.
(5) If *A*+ must have a desire to cooperate with other cooperative agents as a matter of being suitably idealized, *A*+ has a desire to cooperate with other cooperative agents.
(6) If *A*+ has a desire to cooperate with other cooperative agents, her desire to cooperate and its presuppositions and implications can explain some central moral norms in terms of cooperation.

(C) If *A*+'s psychology is able to explain the reasons of an agent *A* in our world, her desire to cooperate and its presuppositions and implications can explain some central moral norms in terms of cooperation.

How does the argument work? To start off, premise (1) might seem fairly obvious already. I have already presented the work idealization must do to explain reasons; I claimed that it makes sure that the agent has the right desires to explain reasons. This is because it ensures us that the ideal agent is fully functional or fully manifests her capacities insofar as she acts, and that she is in the relevant background conditions for doing so. So premise (1) seems safe.

Nevertheless, suitable idealization has some important implications that I shall introduce here. They, in turn, have surprising normative upshots. I wrote above that I would discuss the circumstances *C* an agent may be in in more depth later, and it is now time for that. By 'circumstances', I mean the natural, social, physiological or psychological background conditions that an agent faces or could face, viz. the conditions of those kinds that may affect her psychology, decisions, or actions. Examples of what I have in mind are what species she belongs to, what planet she inhabits, and what society she lives in.

To introduce some more terminology, I shall call a particular set of circumstances that an agent may be in a *situation*. As an ideal agent can plausibly be in or have desires for what to do in many situations, situations may sometimes be understood as possible worlds, however they should

be interpreted if they are to be compatible with interpretations that do not have any controversial ontological implications about what they or the ideal agent must be like.[4] For the same reason, a situation may sometimes be a subset of the circumstances inside some world – an ideal agent can be in or have desires about what to do in many possibly subsets of circumstances there too. However, if you do not like possible worlds-talk, feel free to reinterpret what a situation is using your preferred interpretation of 'sets of circumstances'. Fundamentally, what matters here is that the ideal agent may inhabit or have desires about what to do in different natural, social, physiological and psychological circumstances. This has important ramifications.

Why? I have presumed that the desires of ideal agents explain the reasons of non-ideal agents. But *REASONS INTERNALISM* does not, by itself, say which circumstances ideal agents must be in or have desires about what to do in, only that their desires explain the reasons *actual* agents have in their circumstances *C*. This seems to make it possible that an *A*+'s situation (or the situations she has desires for what to do in) may differ from an *A*'s.

But too great divergences between *A*+'s situation or desires and *A*'s circumstances could, in turn, give *A*+ different kinds of desires than those *A* plausibly has – and therefore give *A* different reasons than those she plausibly has. This risks generating several explanatory worries for *REASONS INTERNALISM*. For example, how do we know our reasons if *A*+ may be in very different circumstances from us? How can the desires *A*+ has in such cases be related to us and our actions? And might such a view get the extension of our reasons wrong?

Fortunately, these worries can be handled by resources internal to *REASONS INTERNALISM* as I understand it. Idealization is supposed to have two main dimensions: *A*+ should be fully functional (or a fully capacity-manifesting agent when she acts) in background conditions relevant for maintaining her full functionality. Hence, *A*+'s psychology (or circumstances, which might alter her psychology) should not plausibly be altered in more ways than by idealization in these two dimensions, for further changes are irrelevant. So something like the following constrains any *A*+:

> (*CLOSE*) *A*+'s idealized desires explain *A*'s reasons only if *A*+'s desires and other psychological states range over circumstances that are similar enough to those *A* may be in.

[4] Of course, if the reader prefers more ontologically heavy-duty possible worlds, she should feel free to go with them instead.

This means that at least some circumstances that *A*+ inhabits or has desires for what to do in must be similar to those *A* is in.[5] There are hard questions to ask about how we should understand that similarity (e.g. in terms of possible worlds in some technical sense?), but while such questions are interesting, taking positions on them would risk making controversial commitments that do not matter for present purposes.

Instead, here, it is enough to have an intuitive grasp on the limits *CLOSE* sets: *A*+'s desires must range over circumstances that are similar enough to those *A* is in if they are to explain *A*'s reasons. Hence, for example, if *A* is a human, we can safely rule out cases such as when *A*+ is Cthulhu from explaining *A*'s reasons. Cthulhu's desires, let alone background natural, social, physiological or psychological conditions, are plausibly very different from those of any human. Understood like that, a condition like *CLOSE* seems extremely plausible.

There is also another, similar, property of *A*+'s that does very important work when it comes to explaining *A*'s reasons. This property is:

> (*ROBUST*) *A*+ must have psychological dispositions and capacities that remain the same over minor changes in the circumstances she may inhabit or otherwise have desires for what to do in.

ROBUST, too, can be explained by idealization. Idealization involves making an agent fully functional or fully capacity-manifesting insofar as she acts, as well as making sure that she is in the right background conditions. *ROBUST* is such a background condition. This is because *A*+ hardly can manifest her psychological states to act if they were to change capriciously with various more or less randomly occurring events – and this would soon also undermine their functionality. If *A*+'s desire to drink when thirsty, for example, were to turn into a desire to wear a red jumper when thirsty because her neighbours have acquired a cat, she would not be able to act on the desire to drink if her neighbour indeed did acquire a cat. Then she would soon die of thirst, completely undermining the functionality of her psychology. A 'minor' change, then, is a change in *A*+'s circumstances which is such that, if *A*+ had been sensitive to it, it would undermine her being ideal. *A*+'s psychology must be *ROBUST* in the face of such changes.

On to premise (2). I assume that the circumstances of justice include the sort of things that Rawls (1971, pp. 126-130), following Hume (1978, pp. 473-534), took for granted to apply in ordinary human circumstances. The most significant one is that people's desires usually cannot all be easily satisfied given the constraints that their social circumstances put on them. Moreover, in

[5] Is *A*+ not a *hypothetical* agent, and hence unable to 'inhabit' situations where her desires might change? Well, if we can think of hypothetical agents, we may also think of the habitation hypothetically.

the circumstances of justice, there is a moderate scarcity of resources, moderate generosity on part of others (or moderate ideological agreement between agents), and it is within others' power to – either individually or together – thwart any given individual's attempts to satisfy her desires by overpowering her. (To be clear, this use of power need not be moral or nice; the point is that agents are able to use their power to harm each other.) Under these circumstances, living in cooperative societies usually benefits individual agents, but participating in them does not always lead to the best results for any individual agent, given what they desire.

CLOSE ensures us that ideal agents must have desires about what to do in an extensive set of situations that feature these circumstances of justice. To recapitulate, *CLOSE* says that *A*+'s idealized desires explain *A*'s reasons only if *A*+'s desires and other psychological states range over circumstances that are similar enough to those *A* may be in. But our situation contains the circumstances of justice, and situations that do not would be very different from ours just because they would not feature the circumstances of justice. Moreover, there may be all kinds of differences between different versions of the circumstances of justice that we inhabit. There are already many such versions in the actual world, and there may be further ones still. So if *A*+ had lacked desires for what to do in an extensive set of such circumstances that we may inhabit, *A*+'s desires would not range over circumstances that are similar enough to ours to explain our reasons.

True, it is also possible that some agents do not inhabit these circumstances, or if they presently inhabit them, they may come to leave them. Hence, ideal agents should also have desires about what to do in at least some *other* circumstances. But any even remotely humanlike creature is also likely to risk being in the circumstances of justice, so their ideal counterparts should have desires about what to do in situations that feature them, too. This means that premise (2) is in place. And premise (3) follows.

Premise (4) says that if *A*+ has a set of idealized desires (based on *A*'s desires) for what to do in a range of situations or circumstances, many of which feature the circumstances of justice, *A*+ must have a desire to cooperate with other cooperative agents as a matter of being suitably idealized.[6] This conclusion follows from the nature of idealization and *ROBUST*.

How? I shall argue that because the ideal agent has desires for what to do in an extensive range of situations featuring the circumstances of justice, to be able to exercise her instrumental rationality in a *ROBUST* way, she must have that cooperative desire (in all such situations), for that desire is what allows her to be instrumentally rational in a *ROBUST* way. And *ROBUST*, I have

[6] What is 'cooperation'? Good question. I shall only assume that cooperation involves several agents trying to achieve some end together (cf. Regan, 1980, p. 129). The reader is free to fill in with more.

argued, follows from *REASONS INTERNALISM*, so the 'must' here is not normative. It is explained by the features of idealization.

Now, it is well known that, *prima facie*, it need not always be better for individually self-interested agents to abide by the rules of justice. Gyges, Fooles, and Sensible Knaves populate the history of philosophy. These characters are sometimes better at satisfying their desires than the virtuous are. Nevertheless, having desires for what to do in the circumstances of justice, *A*+ benefits from participating in human societies, including benefiting just from living in a society in general.

In fact, being able to enjoy the good of cooperation is a matter of *A*+'s idealized instrumental rationality. Whatever else instrumental rationality requires, it requires taking the best means one believes there are to one's ends, where 'best' should be understood weakly, as whichever means is the one to take in one's circumstances. And the two main idealizing conditions of the ideal agents are that they, first, are supposed to have the features constitutive of agency fully functional or manifested when acting, and, second, that their background conditions are the right ones for their functionality.

But to be able to be fully functional or manifest her capacity for instrumental rationality in actions, an ideal agent must be able to take the best means she believes there are to her ends. Similarly, the ideal agent must be able to take the relevant means to satisfy some different desires, since the agent can have multiple, conflicting or changing, desires. This means that insofar as the ideal agent has desires for what to do in the circumstances of justice, the goods of social interaction are necessary background conditions to ensure that her psychology is functional or possible to manifest when acting. And this is because, in situations featuring the circumstances of justice, social interaction generates more and better means both relative to the ideal agent's existing desires and relative to other possible desires she may have. The former is usually the case, and the latter is always the case, for even if agents do not have desires that are better satisfied using means available in social interaction, they can always acquire such desires. So in the circumstances of justice, social interaction is a background condition for the ideal agent's instrumental rationality.

Then we may draw a distinction. Either the ideal agent has a final desire to engage in cooperative schemes, at least given that other agents also do so, or not. If she does, all is well with her when she participates in social interaction. She will happily do so. But assume instead that she lacks such a desire. She need not necessarily be a disinterested maximizer like Gyges, the Foole, or the Knave; she can be anyone who doubts the value of any kind of cooperative arrangements but still benefits from them. What matters is that she lacks the final desire for cooperation.

If *A+* lacks that desire to cooperate, however, but still benefits – again *ex hypothesi* – from those schemes with respect to her instrumental rationality, then she is essentially a free rider.[7] Free riders will, in many situations, be punished by the other participants in the cooperative schemes. In the extensive range of situations featuring the circumstances of justice for which *A+* must have desires, there are no doubt some where that happens. However, in situations where free riders would be punished, an agent with a final desire to cooperate would be able to be more fully instrumentally rational, whereas an otherwise ideal agent who lacks that desire would not. The other agents would not, plausibly, punish an agent with a desire to cooperate.

But then comes the magic trick. An ideal agent who lacks the desire to cooperate is not able to be fully instrumentally rational with respect to taking the best means to satisfy her desires in a *ROBUST* way. *ROBUST*, I wrote, says that *A+* must have psychological dispositions and capacities that remain the same over minor changes in the circumstances she may inhabit or otherwise have desires for what to do in., where a 'minor' change is a change in *A+*'s circumstances which is such that, if *A+* had been sensitive to it, it would undermine her being ideal.

Assume, then, that *A+* inhabits some situation featuring the circumstances of justice. If *A+* were to lack a desire to cooperate, in many such situations, she would be punished as a free rider so that she no longer would be able to be instrumental rational – she might even become literally incapacitated, for example by being killed. Clearly, that would undermine her ideal rationality. So *A+* will only be *ROBUST*-ly disposed to be instrumentally rational if she has a desire to cooperate with other agents in situations where she may be punished. But her psychology should be *ROBUST*, and if she has the pro-cooperative desire, she will be able to maintain her ideal rationality. And as she may be punished in this way in all situations featuring the circumstances of justice – *ex hypothesi*, as she always may be overpowered – it follows that her psychology is *ROBUST* only if she has a desire to cooperate in *all* situations featuring the circumstances of justice.

To be clear, this is not to say that the ideal agent would be *more* instrumentally rational if she were to have a desire to cooperate. It is possible that she would be able to be instrumentally rational in at least some situations even without the desire to cooperate. Rather, with it, she is able be robustly instrumentally rational, which is needed for the manifestation of her capacities and dispositions to the extent which makes her ideal. It is *ROBUST* which does the magic trick here.

How should we characterize the desire that *A+* must have to be ideal? For a start, the desire cannot allow her to cooperate when that seems instrumentally best and free ride when that

[7] Note that the ideal agent here would be a free rider with respect to the means she can take to her ends. That should be enough to run the argument, for I presume that making use of possibilities that other agents' work give her is enough to annoy some others. But it should also be possible to run the argument, *mutatis mutandis*, with the agent free-riding with respect to her desire satisfaction.

seems instrumentally best. As the agents in the circumstances of justice are of roughly equal power, she would be able to be punished when attempting to trick others by free riding whenever she would be found out.

Would an *ideal* *A*+ always be potentially found out and punished? Yes. *A*+ cannot be smart enough to always be able to trick others. If we were to idealize her to that extent, to ensure a balance of power between agents in the circumstances of justice, we would have to idealize the other agents too. And we must do so, because that balance of power is an aspect of the circumstances of justice. This means that it will, in principle, always be possible for others to find out and punish even an ideal agent.

Furthermore, the desire to cooperate plausibly has to be final, and not merely instrumental, or else it would not be very robust. A merely instrumental desire to cooperate is unreliable and likely end up punished, since whether or not it is rational to enact often will be up for grabs, given the agent's other desires. For the same reason, the final desire must be strong enough for *A*+ to act on it, or else she would not be able to be taken seriously by other agents.

With these considerations in mind, I take it that, to be fully ideal, *A*+ must have a fairly strong final cooperative desire with a content which suggests that *A*+ cooperates with others to satisfy *A*+'s other important desires. Moreover, as an enabling condition for the successful and robust exercise of that desire, *A*+'s psychology must be *sensitive* to her situation. Sensitivity, in turn, imposes two conditions on her psychology. First, (*i*) *A*+'s desire to cooperate must be in one sense disjunctive; it recommends cooperation if others cooperate *or* acting on *A*+'s other desires if they do not. Second, (*ii*) *A*+ must not (otherwise) have important anti-social desires that would impede the exercise of the pro-cooperative desire.[8]

A+ is subject to these two extra sensitivity conditions because the desire to cooperate would not be possible to exercise successfully or robustly without them. First, cooperating with all agents, independently of their motives, would put the agent at risk of either being harmed by cooperation or a sucker's pay-off. On many occasions, this would undermine the rest of her psychology. But her psychology is supposed to be *ROBUST*. Hence, the desire to cooperate must be disjunctive.

Second, we could ask what would happen if *A*+ were to cooperate on anti-social desires. By 'anti-social desires', I mean (final) desires for goals the satisfaction of which would significantly impede others' abilities to satisfy their own desires. For example, they might be desires to hurt others so that they cannot satisfy their other desires. If the desires to hurt others so that they

[8] I write 'important desires' rather than 'desires' here because it is possible that we still can have some weak anti-social desires that do not matter for our actions or reasons. Such desires need not be ruled out.

cannot satisfy their other personal desires are satisfied, then those who are hurt cannot satisfy their own desires when cooperating. But then, the other agent(s) would not desire to cooperate with *A*+, given (*i*). It is obvious that if such desires are known among potential co-operators – which they will be by at least some co-operators in the circumstances of justice – aiming to cooperate on the desires will not have others wanting to cooperate with the agent who has them. So for the desire to cooperate not to be self-undermining, condition (*ii*) puts limits on *A*+'s other desires.

It seems plausible to think, then, that a desire to cooperate must feature in *A*+'s idealized psychology. But how exactly does the desire to cooperate feature there? This takes us to a discussion of (5). Premise (5) says that if *A*+ must have a desire to cooperate with other cooperative agents, *A*+ possesses a desire to cooperate with other cooperative agents.

There are two possibilities here. Either the desire to cooperate is ('just') an extra desire of *A*+'s, or the desire is part of *A*+'s instrumental rationality. I prefer the former view. Building more conditions into instrumental rationality would be very clunky. For then instrumental rationality would require coherence between desires and means-beliefs *and* a particular desire, making it much less theoretically elegant than adding the desire to *A*+'s psychology. Nevertheless, the desire should still be added to the ideal agent.[9]

Then we arrive at premise (6). How can we go from a desire to cooperate to moral norms? Following Smith (1994, ch. 5), I suspect that there are two properties that are the strongest marks of the moral – universal prescriptivity and conventional recognizability.[10] The former means that the norm must have prescriptive force for all, and the latter that it should be able to be recognized as moral via some moral convention. Using the desire to cooperate, we can explain two fundamental moral norms with these properties. They lie at the heart of a full account of moral norms.

First, according to *REASONS INTERNALISM*, reasons have their sources in the desires of ideal agents. With their desires to cooperate, idealized agents all have a reason-explaining desire to sensitively cooperate to satisfy their other important (and respective) desires. So they all have a reason which suggests that they cooperate with other cooperating agents. Moreover, they also lack anti-social desires via the second condition, (*ii*), on the cooperative desire. Because reasons internalism says that the desires of an ideal agent explain our reasons, it follows that we all have a

[9] Moreover, it is plausible that ideal agents are *constituted* by their psychologies (cf. Leffler, 2019). If that is so, it follows that the desire is partially constitutive of the ideal agent. This assumption is, however, not necessary for the my argument here.

[10] Based on a comprehensive literature review, Forcehimes and Semrau (2018) list four potential moral/non-moral distinctions: (*i*) moral reasons are not merely social, but have stronger force than that, (*ii*) moral reasons do not depend on individual commitments (but are categorical), (*iii*) moral reasons are responsibility-implying, and (*iv*) moral reasons are altruistic. Here, (*i*) and (*ii*) look like ways to try to spell out the intuition of universal prescriptivity, whereas (*iii*) and (*iv*) are ways to spell out conventional recognizability.

reason to cooperate to satisfy our other respective important desires – except anti-social desires, which ideal agents are ruled out from having.[11]

Since reasons are prescriptive, and all ideal agents have this pro-cooperative desire, guaranteeing that we all have the same reason to cooperate, it follows that we all have the same universally prescriptive reason. Anyone whose reasons can be explained by *REASONS INTERNALISM* has a reason to cooperate with other cooperative agents.

Moreover, the reason is recognizably moral, for a fundamental pro-cooperative reason seems like the kind of thing we want to count as moral. We have a reason to cooperate in our social interactions so that we can act on our other important desires, which means that we have a reason to simultaneously *benefit* each other – we can act on reasons set by our respective desires – *recognize* each other's ends – since those are what we have reason to cooperate on – and *respect* each other as setting ends – for that is what important desires set. Moreover, since the desire to cooperate extends to all other cooperating agents, beneficence, recognition, and respect are mutual between all cooperating agents, so one may well argue that it is *fair*. These are familiar moral themes. Hence, the reason to cooperate explains a fundamental moral norm.

Second, there is also another moral norm that can be explained by the conditions that enable the cooperative desire. This norm stems from condition (*ii*). Condition (*ii*) rules out cooperating on anti-social desires, for it rules out anti-social desires on part of the ideal agent. Given reasons internalism, it follows that it rules out some potential reasons. Moreover, it does so universally, since all ideal agents have it, and it is clearly recognizably moral, because it seems to explain a norm against anti-sociality in its own right. Hence, it explains a moral norm – but not a moral norm based on a reason; rather, it *rules out* some potential reasons.

To exemplify this, consider Bernard Williams' (1995) case of a husband who abuses his wife but lacks any motivation to stop (even after being idealized). Assume that the husband has a final desire to abuse her and lacks desires not to do so.[12] On Williams' view, the husband lacks an internal reason to stop because he cannot be so motivated. On my view, however, he would not have a reason to start in the first place, because his desire is anti-social. On any plausible interpretation of what 'abuse' is, the abuse limits his wife's abilities to satisfy her own end-desires when this desire is satisfied – perhaps out of physical pain, but more likely out of the psychological impact of such actions. This means that the idealized counterpart of the husband, whose desires

[11] I just wrote: 'reason to cooperate to satisfy (…)'. Does this mean that agents have a reason to cooperate to *actually* satisfy each other's respective desires, or to cooperate in a way which *allows* them to satisfy their other respective desires? I think this is an issue in first-order ethics that my theory does not answer. Here it is enough to say that we have a reason to cooperate.

[12] This is, of course, not a very realistic interpretation of all cases of abuse. But I am not after realism here, I am just after illustrating my point.

give him reasons, will lack that desire. So he cannot have a reason to abuse his wife grounded in the desire to do so.

Can we say even more about the moral norms we have on this theory, beyond the two fundamental ones? Yes. In virtue of the structure of the theory, we can also explain an additional type of moral norms. This type is based on reasons that are necessary means to cooperate, and hence for acting on our fundamental reason to cooperate, insofar as we are involved in social interactions. Given the reason-grounding desire to cooperate, a necessary means for cooperating is to cooperate to satisfy those other desires. Insofar as we are involved in cooperation-inducing situations, then, cooperating to satisfy them is a necessary means for living up to the central cooperative norm, for there is no other way to do it than through these desires (cf. Strandberg, 2019). These necessary means are *secondary moral norms*. Secondary moral norms are universally prescriptive because everyone has them (in the right situations) and are (at least usually) recognizably moral, but they are more contingent than other norms, for people's desires may, of course, vary.

There can, therefore, be all kinds of moral norms, depending on what people in various social settings desire to do. But regardless of our distance to others, we have reason to cooperate with them, *qua* cooperating agents. This gives a basis for a conception of society as a system of mutual cooperation. And, importantly, it gives us the conclusion (C) in the argument from idealization.

(3) Law

I have now presented and briefly defended an argument from idealization which suggests something about the nature of morality. What does it have to do with the nature of law? The two main competing theories about the nature of law in philosophy of law are natural law theories and legal positivism. To put one difference between the theories very crudely, natural law theories feature the idea that there is a theoretically important necessary connection between moral norms and legal norms, whereas legal positivists deny this claim.[13] But natural law theorists tend to believe that law has its basis – perhaps its source or ground – in morality.

Because natural law theories take law to be necessarily connected to morality, natural law theories have, at least, two fundamental and well-known problems just on their face. One is to

[13] I add the 'theoretically important' qualifier here as positivists may accept that there are several necessary connections between law and morality. Even arch-positivist H.L.A. Hart thought there were two: law often depends on morality in interpretations, and moral and legal principles are similarly public (Hart, 1961, ch. 9). But these connections are not explanatorily important.

explain which moral norms there are that plausibly can play their role as being necessarily connected to, and even be the bases of, law. Call this the *problem of moral foundations*. To be sure, natural law theorists have tried to solve it. Perhaps most famously, John Finnis (2011) is the paradigmatic defender of such norms, and he defends an Aristotelian-*cum*-Thomist conception of the morality that underlies law. But whether that framework does any important explanatory work well is very much an open question.

Another major problem is the *problem of bad law*. Perhaps it is the case that laws, in many societies, indeed are moral. But that cannot always be the case. What do we make of, for example, Nazi law if law is supposed to be necessarily connected to, and indeed based on, morality?

It is here that the cooperation-based conception of morality just defended comes in helpful. To see why, we need to take explore the theoretical structure of natural law theories further. At least one significant strand of natural law theories is, to use Jonathan Dancy's (2018) term, *focalist* in structure.[14] This means that the theories are structured so that one takes a certain class of cases or instances of the phenomenon one is trying to analyze to be focal (or 'central', or 'paradigmatic', or 'ideal') and other instances of the phenomenon one is analyzing to be less focal (or 'central', or 'paradigmatic', or 'ideal'). Despite being similar to the focal case, and in some sense belonging to the same kind of phenomenon, they deviate from it in some interesting way.[15]

There are several distinct kinds of focalist theories, and their structures are the same regardless of what they are supposed to be theories of. Most notably, one may either take the focal case(s) to be the function of the phenomenon one is discussing, or hold a disjunctivist view. On the former type of theory, the phenomenon one discusses has a function, but various instances may live up to it to higher or lesser extents. If the function of the eye is to see, seeing eyes may be focal cases, but blind ones are not. On the latter type of theory, the phenomenon one discusses comes in distinct kinds, one of which is more fundamental than the others. The non-fundamental kinds may then also deviate from the fundamental kind. The fundamental kind of perception may, perhaps, be veridical, whereas hallucinatory perception is a non-fundamental kind of perception (cf. Martin, 2004).

Natural law theorists need not, for present purposes, decide which type of focalist theory is preferable. For regardless of which type of theory is better, the cooperation-based moral norms from the last section can contribute. In particular, the secondary moral norms I have discussed seem like excellent candidates for sometimes *being* laws. They are socially recognized, and if they

[14] Dancy does, in fact, refer back to natural law theorist Finnis when outlining this type of theory (Dancy, 2018, pp. 105-106).

[15] I formulate focalism slightly differently from Dancy: he does not like the term 'paradigmatic', and prefers to say that the non-focalist cases depend on the main one rather than deviate from it. For present purposes, nothing of interest turns on these changes.

are also formally recognized, for example by being codified by being written, they look a lot like actual laws. Indeed, I hypothesize that if some moral norms of that kind are instantiated, it is very likely that at least some of them actually are laws.

If that is right, we can provide fairly easy solutions to the problems of moral foundations and bad law. The former problem can be solved by counting at least some moral norms – usually, the codified or written ones – as the moral foundations of a natural law theory. If the argument from section (2) is correct, we can see how a particular kind of norms seems likely to play the role in the theory.

Moreover, the problem of bad law can be solved by a manoeuvre that natural law theorists often make anyway, namely, by appealing to the focalist structure of the theory (cf. Finnis, 2011, pp. 351-366). For regardless of whether one goes for a functionalist or a disjunctivist focalist theory, one can count laws that are not secondary moral norms as not living up to the function of law – which one may construe as 'instituting secondary moral norms in a society' or suchlike – or the fundamental kind of law – when that in itself is a certain subset of the secondary moral norms there are.

Admittedly, this manoeuvre is only as plausible as what one posits the function or fundamental kind of law to be. But that is a point that speaks in favour of the agreement-based conception of morality as a basis for law that I have defended here. On more traditional natural law theories, the relation between the norms that lie at the basis of law and how we usually understand law can be murky. Why would human social institutions have a lot to do with mind-independent or even God-given moral facts, for example? By contrast, the contractualist elements of the reasons internalist picture I have presented clearly shows how secondary moral norms have a social element from the start, which, as I have stressed, makes them very intuitive candidates for lawhood.

(4) Conclusion

In this chapter, I have presented a surprising version of a natural law theory. In section 1, I introduced reasons internalism. Then, in section 2, I sketched an argument from idealization for a cooperation-based theory of moral norms. In section 3, I indicated how these socially accepted norms, in the right circumstances, might be slotted into a natural law account of laws and yield unexpectedly plausible results.

Does that mean I have defended a natural law theory? Sort of. I am attracted to the picture I have presented, but I am not sure about why one should prefer a theory with a focalist structure

to one which is more straightforward. So I recommend the reader to treat this paper as exploratory. It fundamentally aims to paint a picture and show why it has some attractions. But whether or not that picture is accurate will have to be determined elsewhere.

References

Chang, R. 2009. Voluntarist Reasons and the Sources of Normativity. In: Sobel, D. and Wall, S. eds. *Reasons for Action*, New York, NY: Cambridge University Press, pp. 243-271.

Chang, R. 2013. Grounding Practical Normativity: Going Hybrid. *Philosophical Studies*. 164(1), pp. 163-187.

Dancy, J. 2018. *Practical Shape: A Theory of Practical Reasoning*. Oxford, UK: Oxford University Press.

Forcehimes, A. and Semrau, L. 2018. Are There Distinctively Moral Reasons? *Ethical Theory and Moral Practice*. 21(3), pp. 699-717.

Finnis, J. 2011. *Natural Law and Natural Rights*. 2nd Edition. Oxford, UK: Oxford University Press.

Hart, H.L.A. 1961. *The Concept of Law*. Oxford, UK: Oxford University Press.

Hume, D. 1978. A Treatise of Human Nature. In: Selby-Bigge, L.A. and Nidditch, P.H. eds. *A Treatise of Human Nature*. 2nd edition. Oxford, UK: Oxford University Press.

Hurtig, K.I. 2006. Internalism and Accidie. *Philosophical Studies*. 129(3), pp. 517-543.

Joyce, R. 2001. *The Myth of Morality*. Oxford, UK: Oxford University Press.

Leffler, O. 2019. *The Constitution of Constitutivism*. PhD Dissertation, University of Leeds, UK.

Martin, M.G.F. 2004. The Limits of Self-Awareness. *Philosophical Studies*. 120(1-3), pp. 37-89.

Peter, F. 2019. Normative Facts and Reasons. *Proceedings of the Aristotelian Society*. 119(1), pp. 53-75.

Rawls, J. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Regan, D. 1980. *Utilitarianism and Cooperation*. Oxford, UK: Clarendon Press.

Smith, M. 1994. *The Moral Problem*, Oxford, UK: Blackwell Publishing.

Smith, M. 1995. Internal Reasons. *Philosophy and Phenomenological Research*. 55(1), pp. 109-131.

Smith, M. 2012*a*. Agents and Patients: Or, What We Learn about Reasons for Action by Reflecting on Our Choices in Process-of-Thought Cases. *Proceedings of the Aristotelian Society*. 112(3), pp. 309-331.

Smith, M. 2012*b*. A Puzzle about Internal Reasons. In Heuer, U. and Lang, G. eds. *Luck, Value and Commitment: Themes from the Philosophy of Bernard Williams*. Oxford, UK: Oxford University Press, pp. 195-218.

Strandberg, C.S. 2019. An Ecumenical Account of Categorical Moral Reasons. *Journal of Moral Philosophy*. 16(2), pp. 160-188.

Williams, B.A.O. 1981. Internal and External Reasons. In: Williams, B.A.O. *Moral Luck: Philosophical Papers 1973-1980*. Cambridge, UK: Cambridge University Press, pp. 101-13.

Williams, B.A.O. 1995. Internal Reasons and the Obscurity of Blame. In: Williams, B.A.O. *Making Sense of Humanity and Other Philosophical Papers 1982-1993*. Cambridge, UK: Cambridge University Press.