

THE ETHICS OF ENTRAPMENT: A DIRTY HANDS PROBLEM?

Abstract

In this paper we focus on a possible framework for analysing the morality of legal entrapment (which we define based on our previous work): the dirty-hands model. We take as our starting point Christopher Nathan's criticism of the model (when applied to undercover policing). We have two aims throughout the paper. Our primary aim is to see if the model applies at all to legal entrapment; our secondary aim is to establish whether, if the model applies, Nathan's criticism hold for it. Regarding the first aim, we present three possible versions of the model, loosely taking our inspiration from Nathan's remarks and using the work of Janos Kis. We argue that the first two accounts don't apply well to legal entrapment and the last, perhaps more plausible but still questionably applicable account has no room for morally wrong acts. Regarding the second aim, we argue that Nathan's criticism of the model is not so forceful once we take account of all the resources available in the dirty-hands model. We end the paper with a brief conclusion and offer some concluding – and critical - remarks regarding a closely related alternative to the dirty-hands model as the right framework of analysis: admirable immorality.

Keywords

Undercover policing; legal entrapment; dirty-hands model; moral dilemmas; moral wrongness; moral phenomenology; admirable immorality; Christopher Nathan; Janos Kis

I. Dirty hands and undercover policing: Nathan's criticism

Christopher Nathan (2017) has recently argued against “the dirty hands model” of undercover policing. Here is how he describes the model (Ib. 371):

“The view often attributed to Machiavelli is that power inevitably involves doing some things that are wrongs, arising from genuine moral dilemmas. We must accept this moral residue, but we also do better not to dwell on our misdeeds. On this view, committing moral wrongs is part of the core of undercover work. The best we can do is to embrace the values we gain: in this case, the reduction of crime and the increase in security. It retains, nonetheless, a tragic element, since it is necessary that the work is performed, and those who perform it commit wrongs, thereby performing a sacrifice.”

However, Nathan introduces the model only to side-line it as wrong (Ib.):

“A public that takes on board this view of manipulative policing will correctly feel that it puts wrongful acts at the centre of police practice. The wrongs may be justified by appeal to necessity, but unease will remain. Furthermore, one can reasonably expect that the effects of an internalisation of a dirty hands ethic by agents of a practice that is inherently secretive would be to encourage further secretiveness. A belief on the part of its agents that the practice is not wrongful is more conducive to public justification.”

Since his interest lies elsewhere¹, Nathan does not ask the question whether the dirty-hands model applies to undercover policing in the first place. We think this is a crucial and interesting

¹ In his paper Nathan argues that undercover policing does not necessarily wrong its targets because people can make themselves *liable* to deception and manipulation. He only uses the dirty-hands model as the starting point for his discussion (as the model that he thinks we should be steering away from for the reasons quoted in the text).

question to ask given the influence of the model both in policing (or, more broadly, law-enforcement) and in elsewhere (for example, in politics or in journalism). Whether Nathan's criticism holds can only be asked once the question of application is settled. With this paper we would like to contribute to both sides of this discussion.

However, we will qualify our focus in one important way. Our interest in this paper won't be undercover policing *per se* but something closely related to it: *legal entrapment*. Arguably, some form of entrapment is often involved in undercover policing and undercover policing, as Nathan's own paper testifies, is a crucial policing method.² Hence it is natural to inquire whether the dirty-hands model can be applied to the case of legal entrapment, the morality of which, from other angles, is often discussed. This seems to be a natural move also because the way Nathan describes the model appears to fit well with the structure of legal entrapment. In particular, many argue that legal entrapment is morally wrong for various reasons (cf. Howard 2016, 26), while it is clear that the declared aim of legal entrapment is the prevention, detection or reduction of crime.

Thus, as noted, our primary aim in the paper is to settle the question of application of the dirty-hands model to legal entrapment. Our secondary aim is to see whether Nathan's criticism of the model holds in the case of legal entrapment. Let us take a closer look at these critical points. Nathan argues that on the dirty-hands model of undercover policing

- 1) The public would correctly feel that morally wrongful acts are at the centre of police practice;
- 2) Despite the justification of these acts, public unease will remain;
- 3) The police becomes, as a result of internalization of this ethic, even more secretive;

²There were 1,229 undercover police officers operating in England and Wales in 2014, and there were 3,466 authorized undercover operations in England and Wales. For details, see *An inspection of undercover policing in England and Wales* (HMIC 2014) <<http://www.justiceinspectors.gov.uk/hmicfrs/wp-content/uploads/an-inspection-of-undercover-policing-in-england-and-wales.pdf>> accessed 6 June 2020.

- 4) In contrast, the belief on the part of the police that undercover policing is not wrongful would be more conducive to publication justification.

Our focus will be on 1)-3). Notice that 1), at its core, is a theoretical claim about the right form of the dirty-hands model, whereas 2) and 3) are practical predictions about what will happen in real life if the model is used as the right framework of understanding undercover policing. These are thus rather different claims and should not, we think, be discussed together. Instead, we think the best is to first settle the issue whether we must indeed construe the dirty-hands model as relying on 1). Since this is a conceptual question, moreover, this work can be done as part of our discussion of the question of application (to decide whether something is applicable, we need first to know what it is that we apply). Once this is settled, we can ask the question whether 2) and 3) have any truth to them. To foretell, we hold that 1) relies on a particular dilemmatic understanding of dirty hands that is controversial and does not apply well to legal entrapment, whereas the truth of 2) and 3) is questionable whether or not 1) is correct.

Here is how we are going to proceed. In section II, we will provide a definition of legal entrapment based on our previous research. This, in section III, is followed by a longer discussion of moral dilemmas and their connection to the problem of dirty-hands. We emerge from this discussion with three accounts of legal entrapment as a kind of moral dilemma involving dirty-hands. However, we argue that the first two accounts don't apply well to legal entrapment and the last, perhaps more plausible – but still questionably applicable - account has no room for morally wrong acts. In section IV, we look at how Nathan's second and third critical points fare. We show that it is far from clear that these criticisms stand. We then end, in section V, with a brief conclusion and offer some concluding – and critical - remarks regarding a closely related alternative to the dirty-hands model as the right framework of analysis: admirable immorality.

II. Legal entrapment to commit a crime: a definition

We approach this challenge in two stages. First, we provide a categorization of different forms of entrapment. This will help us single out the specific form of entrapment we are interested in: legal entrapment. Second, we then zoom in on legal entrapment by providing a detailed definition in the form of jointly necessary and sufficient conditions. Throughout this section we will rely on previous published work (*redacted*).

Cases of entrapment involve a party that intends to entrap, whom we call the ‘agent’, and a party that is entrapped, whom we call the ‘target’. Let the terms ‘party’, ‘agent’, and ‘target’ encompass both individuals and groups. We draw two distinctions, which cut across each other, concerning acts of entrapment. The first concerns the status of the agent; the second concerns the act that the target performs and that the agent procures.³

Legal entrapment occurs when the agent is a law-enforcement officer, acting (lawfully or otherwise) in their official capacity as a law-enforcement officer, or when the agent is acting on behalf of a law-enforcement officer, as their deputy. When, on the other hand, the agent is neither a law-enforcement officer acting in that capacity, nor the deputy of such an officer, acting in their capacity as deputy, we have *civil* entrapment.

We distinguish between procured acts of criminal and of non-criminal types. An investigative journalist might entrap a politician into performing a morally compromising act that is not a crime, in order that the journalist might expose the politician for having performed the act. When the act is non-criminal but is morally compromising (whether by being immoral, embarrassing or socially frowned upon in some way), we are dealing with ‘moral’ entrapment (using the word ‘moral’ in a wide sense). When the act is of a criminal type, we have ‘criminal’ entrapment.

³ Our notion of procurement is admittedly technical. We understand procurement to involve the agent in having an intentional influence, via directly related communicative acts, on the target’s will. See (*redacted*) for further discussion.

Accordingly, we classify acts of entrapment via the following two-dimensional matrix.

	A	B
1. Is the agent acting (permissibly or otherwise) in their capacity as a law-enforcement agent or their deputy?	Yes	No
2. Is the act that the agent intends the target to commit of a type that is criminal?	Yes	No

We thus get four types of entrapment:

Type 1 = 1A + 2A = legal entrapment to commit a crime

Type 2 = 1B + 2A = civil entrapment to commit a crime

Type 3 = 1B + 2B = civil moral entrapment

Type 4 = 1A + 2B = legal moral entrapment

Our interest is in Type 1 entrapment – henceforth: legal entrapment - since this is the form of entrapment that Nathan’s subject matter encompasses.⁴

How is it best to define Type 1 entrapment? Elsewhere we argue for and defend the following definition:

- (i) a law-enforcement agent (or the agent’s deputy), acting in an official capacity as (or as a deputy of) a law-enforcement agent, *plans* that the target perform an act;
- (ii) the act is of a type that is criminal;
- (iii) the agent *procures* the act (using solicitation, persuasion or incitement);

⁴ Although Type 1 entrapment will be our focus throughout the paper, occasionally we will also remark on other forms of entrapment (especially Type 2 or 3).

- (iv) the agent intends that the act should, in principle, be *traceable* to the target either by being *detectable* (by a party other than the target) or via *testimony* (including the target's confession), that is, by *evidence that would link the target to the act*;
- (v) in procuring the act, the agent intends to be enabled, or intends that a third party be enabled, *to prosecute* (or threaten to prosecute) the target for having performed the act.

Entrapment (without the qualifier) differs from this definition in that conditions (i) and (ii) are more inclusive. Namely, corresponding to our categorization above, a general definition of entrapment is not restricted to law-enforcement agents and involve acts that are not criminal (but immoral in the broad sense characterized above). There are, of course, many who would take issue with this definition and we interpret, defend and argue for its various conditions elsewhere (*redacted*). In this paper, however, we proceed on the assumption that these points need not be rehearsed again since the purpose of the paper is not to defend our concept of entrapment.⁵

III. Moral dilemmas, dirty hands, and legal entrapment

We now have a suitable notion of legal entrapment at hand. The next step is to connect it to a suitable account of dirty hands. How to understand the problem of dirty hands? To begin, it is useful to rehearse the main elements of Nathan's description of the dirty-hands model. According to the model, Nathan writes, the following features of a situation must be in place:

- Moral wrongs are committed;

⁵ Also, it is useful to distinguish entrapment from, first, *virtue testing* and, second, *mere temptation*. Virtue testing is, in a sense, the weakest of the three notions: it need not involve either procurement or tempting. Temptation and entrapment differ principally in that the former need not involve procurement. While we hold that all temptation is virtue testing, we deny that all virtue testing is temptation; we deny that all temptation is entrapment, and we deny that all entrapment involves temptation. We work out and present the relevant differences in (*redacted*).

- Genuine moral dilemmas are involved;
- A moral residue is involved that we must accept;
- We gain the values of reduction in crime and increase in security;
- We are encouraged not to dwell too much on our misdeeds;
- The overall picture is tragic, though, since a moral wrong must be committed (a ‘sacrifice’ must be made) in order to pursue these values.

Nathan, given, perhaps, that his focus is elsewhere, does not dwell on the concepts he uses. We take a ‘moral dilemma’ to refer to a choice situation in which the agent is confronted with moral oughts and whatever course she takes will be, in some sense, morally unacceptable. We deliberately do not take stance on what ‘moral oughts’ are: they can be duties, obligations, reasons. We, furthermore, do not equate ‘morally unacceptable’ with ‘morally wrong’ in order not to foreclose the important possibility that, even if acting with dirty hands involves facing moral dilemmas, these dilemmas are not best characterized in terms of moral wrongness. Nathan also does not specify what he means by ‘moral residue’, an issue we will clear up in subsequent discussion.

The core of Nathan’s depiction of the model is that acting with dirty hands involves a tragic moral dilemma in which moral wrongs are committed. That is, on a natural interpretation, the phenomenon of dirty hands is presented as a side-product of a tragic understanding of moral dilemmas. There are many ways to spell out this core idea (and further ways are available that discard at least one element of the overall picture). Let us begin with the one that, we think, reproduces the most of Nathan’s depiction above: *the tragic account* (TRAGIC). We should mention that in formulating these accounts we shall rely heavily on Kis (2008, Chapter 9)’s work on dirty-hands. Our contribution will be to see how and if at all, the different accounts of dirty-hands can be applied to the case of legal entrapment. This also means that unlike Kis and

others in the literature, we will only question these interpretations if this directly serves our primary purpose: to make sense of the dirty-hands model, which we will then try to apply to the case of legal entrapment.

The tragic account (TRAGIC)

Nathan says that situations involving dirty-handed acts retain a tragic element “since it is necessary that the work is performed, and those who perform it commit wrongs, thereby performing a sacrifice.” There are two claims made here, we think, under the heading ‘necessity’. One, Nathan assumes that in law-enforcement practice necessity and corresponding last resort conditions are met in the case of entrapment (cf. Bovève 1991 on media ethics or Coady 2008 on politics). This may not be obvious in real-life cases but we can accept it as a requirement that law-enforcement practice should conform to. Two, and perhaps intended as comprising the previous point, one is morally required to do the work: the necessity is also of a moral kind.⁶ This then leads to familiar slogans that aim to bring out the paradoxical nature of the eschewing situation: that sometimes it is right to do what is wrong (Bülow and Helgeson 2018) or that sometimes whatever we do is wrong (Walzer 1973).⁷

However, slogans don’t help us understand the underlying structure of the situation. Let us, therefore, first formalize the account. We could say the following happens in such tragic situations where ‘S’ is the agent (in our case, the law-enforcement officer or his/her deputy), ‘ought’ refers to a moral requirement driven, for example, by reasons or duties, ‘a’ is to entrap and ‘b’ is to not entrap. Three assumptions have to be put in place first (Kis 2008, 238 whose formal depiction, throughout this section, we follow verbatim):

⁶ If this is not true, if the necessity/justification is of a non-moral kind, then we don’t see how Nathan could be talking about a *moral* dilemma. Then instead we have a conflict of morality with something else. See also our remarks on the first version of admirable immorality at the end of the paper.

⁷ See Coady (2008) and Curzer 2006) on Walzer (1973). See Alexandra (2000) on policing. Portmore (2007) would say that these slogans describe two different moral dilemmas: prohibition-dilemmas and obligation-dilemmas. We opt not to use this distinction – we take it that Kis’s (2008) formulations can incorporate both.

A1: *S* ought to do *a*, and *S* ought to do *b*;

A2: *S* can satisfy each of the two oughts separately, but

A3: *S* cannot satisfy both oughts together.

Now, this gives us moral conflict but it doesn't give us a moral dilemma.⁸ For that we must add a further assumption:

A4^{TR}: Neither “*S* ought to do *a*” overrides “*S* ought to do *b*” nor “*S* ought to do *b*” overrides “*S* ought to do *a*”. Both oughts emerge from their encounter undefeated.

This last assumption, A4^{TR} is crucial also because it partially explains the tragic (and paradoxical) nature of the situation: there is no way the agent can do the right thing without doing also something wrong. However, this is only a partial explanation. According to Kis and others (e.g. Gowans 1994 whom Kis cites), moral dilemmas are tragic largely because of the lost innocence of the people involved. The idea is that the dilemma is *inescapable*: one finds oneself in the dilemmatic situation due to no fault of one's own. That is, one goes into the situation innocently but, because of the nature of the choice involved, one cannot come out of it innocently. That is, given the undefeated oughts on both sides *and* the way of entering into the situation, TRAGIC has tragical implications:

I1^{TR}: Whichever ought *S* should choose to disregard, *S* violates a valid, in-force ought;

I2^{TR}: The situation described by A1-A4^{TR} is such that *S* may become involved in it innocently,

I3^{TR}: once in it, however, *S* has no choice of coming out of it innocently.

⁸ Some, such as Hare (1981), would even deny the existence of moral conflicts.

I4^{TR}: It will be appropriate for *S* to feel guilty about what she does.

To sum up, on this reading moral dilemmas have three layers of the tragic. The first is encapsulated in I1^{TR}, the second is to be found in I2^{TR} and I3^{TR}. Once these three implications are in place, I4^{TR}, the third layer, falls in place. However, when attempting to apply this interpretation of moral dilemmas to the case of legal entrapment, all three layers can be called into question. We will focus on the first two layers. The third layer is arguably not supported by the phenomenology of these cases, i.e., most if not all law-enforcement agents don't feel guilty about what they have done⁹. Of course, this does not rule out that it would be *appropriate* for law-enforcement agents to feel guilt. However, given that the appropriateness of such reactive moral emotions is determined by the moral structure of the case (what these agents all things considered morally ought to do), the third layer of the tragic is parasitic upon the previous two layers.

Regarding the first layer, it is far from clear that A4^{TR} would be an assumption that holds of all or even just most cases of legal entrapment. Certainly, law enforcement officers might be doing what is *normally* considered wrong overall (i.e. all-things-considered), after all, for example, they often tempt and deceive people, not necessarily career-criminals, into doing something criminal. Still, in the particular context of legal entrapment, this is far from clearly so given all the good that will be achieved through the act (in terms of increased public security). The literature appears to be far from a consensus on what is the overall correct normative verdict on entrapment; what is not, to our knowledge, suggested is that there would *not* be a clear

⁹ Even if they do feel *bad* about what they did, there can be other explanations of this fact. The perhaps best alternative that is particularly fitting for police work (as well as for media, see Bovève 1991) is to emphasize the uncertainty – the *doubt* – that accompanies such work, including acts of entrapping: many officers ex post might be haunted by doubt that they have done the right thing. This explanation, moreover, typically accompanies accounts that do not accept the existence of moral dilemmas and want to explain its phenomenology in a way that does not assume that moral wrongs were committed. See Hare (1981) and Nielsen (2000) for canonical expositions of this approach. There can be other explanations of possible bad feelings (regret and remorse, in particular) and we will appeal to them in what follows.

normative conclusion. In short, even on a charitable approach, there is reason to hold that A4^{TR} doesn't hold in many cases of legal entrapment. A case by case analysis would be needed; TRAGIC cannot, without further evidence, be taken to apply to legal entrapment across the board.¹⁰

Regarding the second layer, it is not clear how the inescapability requirement would apply to cases of legal entrapment. After all, legal entrapment involves law-enforcement officers (or their deputies) making the choice of entrapping their targets. This choice could hardly be construed as inescapable – these officers, we can reasonably assume, made an informed choice when they became officers and their choice of opting to entrap their targets is also free and informed. The stereotypical case of an inescapable choice, that is often used in discussing TRAGIC, is Sophie's choice (based on Styron 1979) who had to choose which of her two children was to be sacrificed (and if she refused the choice, both children were taken away to the gas chamber). The choice was imposed upon Sophie by Dr Mengele: she did not somehow create the choice situation and she had no innocent way out of it either. Sophie, thus, truly loses her innocence no matter what she does; but can we say anything remotely similar of law-enforcement officers who freely enter their profession as well as engage in entrapment?¹¹

Lastly, and turning now away from the layers of tragicness, Nathan, probably following Walzer (whom he references in a footnote), talks about a *moral residue* as resulting from the

¹⁰ Cf. Alexandra (2000) on typical cases of noble cause corruption; Walzer (1973) is unclear on exactly what situations are covered but his focus is mostly on emergencies, which, by their nature, are rare situations.

¹¹ Two possible replies might be offered (thanks to *redacted*). Innocence comes in degrees: while Sophie fully loses her innocence, law-enforcement officers only lose some degree of innocence. This might be true but is it enough to put the second layer of the tragic in place? We doubt it. The notion of inescapability might also be questioned. Sophie was no doubt innocent in the choice situation described, but did she indeed enter it innocently? She has made all sorts of prior choices, we can presume: Why did she not hide from the Nazis? Why did she not try to flee? Insofar as we have free will, perhaps no choice we make is truly inescapable...But does this 'help' law-enforcement officers? We could say that a choice situation is inescapable if one enters into it while not having created it, not having intended it, or not having even foreseen it (this might be too demanding, but let us go along with it for now). Will this do? Not likely. It does explain why Sophie's choice is inescapable and perhaps also shows that a law-enforcement officer's career choice did not itself make the choice to entrap escapable. However, the very act of choosing to entrap remains escapable also on this test. This remains the case even if entrapment is a last resort tool of policing and even if the officer is somehow driven to do it as a matter of necessity (cf. William's 1981 on practical necessity).

wrongness of, in our case (we assume), the entrapping act. But this is a rather awkward construction. If the agent has done wrong, then the agent, as per I4^{TR}, is guilty of wrongdoing and this isn't a moral remainder somehow, we think. Could the moral residue be referring to the moral phenomenology of these cases, i.e. to what the agent experiences or at least what would be appropriate to experience? However, we don't think it is plausible to construct the moral residue in question as merely phenomenological. Nathan himself writes about "this moral residue" where 'this' refers back to the moral wrongdoing and not to the feelings about such wrongdoing. In other words, the residue should be located in the moral structure of entrapment and not (merely) in its phenomenology.

Thus, we think TRAGIC, despite its large degree of fittingness to what Nathan says, is not a good way to take for advocates of the dirty-hands model of legal entrapment. At the same time, the problems above suggest an alternative, known again to us from the literature on moral dilemmas (in particular, from the writings of Bernard Williams), that might therefore be a better choice for our purposes: the *moral residue account* (RESIDUE).

The moral residue account (RESIDUE)

Williams (1973) has argued that what happens in moral dilemmas is that, contrary to A4^{TR}, one ought overrides the other oughts present but that the overridden, defeated oughts are not silenced: they 'stick around', their force doesn't evaporate.¹² In particular, these defeated oughts give rise to *derivative oughts* to compensate for and/or to repair the damage done: this is the moral residue that TRAGIC was lacking. As Kis (2008, 251) puts it: "The defeated ought has no action-guiding force in the immediate context of the situation in which the choice is being

¹² RESIDUE is supported naturally by a picture of competing *pro tanto* oughts the balancing of which gives us an all-things-considered ought-judgment. See Dancy (2004) for the broader meta-ethical picture; see also Alexandra (2000) who depicts ordinary cases of noble cause corruption exactly this way. Note also that this meta-ethical view of reasons is a feature RESIDUE shares with TRAGIC: both accounts presuppose that competing reasons compete with each other based on their strengths; alternative behaviours of reasons are not considered (such as silencing or excluding – see (*redacted*) on the former, Raz (1999) on the latter).

made, but it has action-guiding force in the context of a later choice that emerges in virtue of *S*'s action." That is, the decision situation is more complex than in TRAGIC. We have agent *S* who, if she chooses *b*, must then do c_a (where this represents compensating for and/or repairing the harm caused by failing to do *a*), and if she chooses *a*, must then do c_b (same as c_a just with *b*). The decision is more complex because, when deciding how to act, *S* must not only decide whether to do *a* and *b*, but also consider whether it is feasible for her to do c_a and c_b .

Before we move on to the question of application, this time an important general problems has to be considered. For it appears that RESIDUE's claim to a moral residue vanishes on closer analysis. After all, if *S* makes the right choice by choosing to act on the most forceful ought and then also compensates the victims of her choice (for failing to take the other courses of action open to her), no moral residue remains.¹³ *S* simply did the right thing, on both levels (acting and then compensating); the moral universe remains intact and she comes out of the situation (morally) innocent.

There is a way around this problem but the needed move also suggests that the proper form of RESIDUE whose application to legal entrapment we intend to query is not the one we started out with. What is the extra move? As Kis points out, if there are *irreparable* damages involved in a choice situation, that is, if either c_a or c_b would be such that *S* is not able to carry them out, then a moral residue would *necessarily* remain. This appears to restore the tragic character of *S*'s choice-situation since there would be no way for her to fully satisfy the requirements that apply to her. To formalize, we would get the following depiction of the moral residue account. Start with our original depiction of moral conflict:

A1: *S* ought to do *a*, and *S* ought to do *b*;

¹³ When applied to legal entrapment, this means that we presuppose that in entrapment cases compensation is not in principle impossible. Given that, e.g., courts standardly order compensation by the police, we take it that this assumption is defensible.

A2: *S* can satisfy each of the two oughts separately, but

A3: *S* cannot satisfy both oughts together.

We should add, following this version of RESIDUE, a fourth assumption (Kis 2008, 253):

A4^{MR}: At least *a* involves a non-eliminable moral residue, and *b* either involves a non-eliminable moral residue or the requirement of doing it is not overriding.

However, replacing A4^{TR} with A4^{MR} produces its own problems. First, as Kis (Ib. 254) points out, if a damage is irreparable, i.e., if it *cannot* be repaired, then, by the ought-implies-can principle, it *ought not* be repaired.¹⁴ This means that the ‘rediscovered’ tragic element in RESIDUE begins to fade away again: if the damage is repairable, no residue need remain; if the damage is irreparable, a residue remains but it oughtn’t be - because cannot be – repaired and/or compensated for. In either case, no derivative ought remains in the final moral analysis that could be violated, triggering a tragic conclusion.

Second, Kis (Ib.) argues that this doesn’t rule out that it would nevertheless be appropriate for *S* to feel bad about her failure to compensate the victim(s) of her act. True, this feeling should not be guilt, perhaps not even regret or remorse; still, *S* can think of her act as morally reprehensible and feel accordingly. This gives a thinly tragic analysis: whatever *S* does, it is appropriate for her to feel bad about her choice of act. Is this enough? We don’t think so. Very little of the tragic aspect of the original situation remains as far as the moral analysis of entrapment is concerned. And we have already argued that cashing out the supposedly present moral residue exclusively in terms of moral phenomenology is not an adequate solution.

¹⁴ Of course, the principle has its detractors. See e.g. Stocker (1990). Still, most accept some version of the principle and, we think, common-sense is also on our side. In any case, given the nature of this paper, this is not the place to discuss problems with the principle.

Furthermore, third, the original version of RESIDUE was applicable to legal entrapment across the board. But the present version is overly restrictive: only those instances of legal entrapment come under it that involve irreparable damage. These surely are rare – an irreparable damage is always a very great damage, and most cases of legal entrapment are too mundane to involve such damage. In short, what we gain in ‘tragicness’ in changing focus to irreparable damage, we lose in scope of application.¹⁵

Lastly, fourth, there is good reason to think that with RESIDUE we would also not be able to preserve the dilemmatic nature of dirty hands. As Kis points out, reference to irreparable damage cannot be what constitutes moral dilemmas because hard choices of the form A1-A3 depicts that are not moral dilemmas can also involve irreparable damage. Just imagine cases when you allow someone to die by deciding to save another and your choice is perfectly well supported by moral reasons. RESIDUE should be able to distinguish this case from genuine moral dilemmas but it cannot. However, this goes against our intention to find a dilemmatic reading of dirty-hands that we can apply to legal entrapment.

Driven by the failures of TRAGIC and RESIDUE, Kis (2008) proposes a third account – the *dirty hands account* (DIRTY) - that he claims is superior to the other two. The question, besides its general plausibility (which we won’t take up), is whether it is applicable to the case of legal entrapment.

The dirty hands account (DIRTY)

The standard problem for any account of moral dilemmas - in the framework of which we have so far tried understand Nathan’s dirty hands model - lies in their paradoxical nature: that the agent is supposed to do something that is both morally unacceptable (wrong) and morally acceptable (right) at the same time. Both TRAGIC and RESIDUE try to make sense of this in

¹⁵ In fact, this gets worse if we consider that, on A4^{MR}, the choice to not entrap (‘b’) should be either not overriding (not obvious, to say the least) or should cause irreparable damage (certainly not the case across the board).

their own way. However, notwithstanding more general problems with them, their applicability to legal entrapment appears, or so we have argued, questionable. DIRTY is the third – and last - attempt we shall discuss that connects dirty hands to moral dilemmas. We start with the usual three assumptions depicting moral conflict:

A1: *S* ought to do *a*, and *S* ought to do *b*;

A2: *S* can satisfy each of the two oughts separately, but

A3: *S* cannot satisfy both oughts together.

The fourth assumption, as before, is the one that fulfils the task of accounting for the dilemmatic nature of the situation including dirty-handed acts (Kis ib. 264). The central concept of this assumption is *moral reprehensibility*:

A4^{DH}: At least *a* is morally reprehensible, and *b* is either morally reprehensible or it is not morally overriding.

How this helps to provide a new account of dirty hands that is also dilemmatic follows from the way Kis arrives at A4^{DH}. His argument consists of three steps (Ib. 260-3). First, he holds that an act can be right to do, hence morally acceptable, in certain circumstances and nonetheless morally reprehensible in the same circumstances. This is possible because some acts such as murder and betrayal are such that they have *essential properties* that make them morally reprehensible irrespective of circumstances (the idea is borrowed from Marcus 1996). Some concepts, like ‘murder’ or ‘betrayal’, are such that their descriptive content cannot be determined in separation from evaluative criteria: the acts they describe cannot be identified in

morally neutral terms.¹⁶ Murder and betrayal always remain morally reprehensible even if in the given situation they are, all things considered, the morally right thing to do. Still, as Kis (Ib. 260) points out, it can happen that the given act has the essential properties that make it morally reprehensible only on certain descriptions of the act.¹⁷ For example, pulling the trigger of a gun can be described both as killing and as not killing depending on whether the gun is loaded or not. In short, an act is “morally reprehensible if it meets a description under which it has essential properties that make it reprehensible, and if the particular circumstances under which it is not committed does not make it inappropriate to meet that description.” (Ib.)

The way is now open to hold that certain acts are morally acceptable (because morally right) and morally unacceptable (because morally reprehensible) at the same time. Of course, and this is the second step in Kis’s argument, we still have to prove the first part of this claim. At this point, Kis resorts to a certain version of deontology: *threshold deontology* (Nagel 1979).¹⁸ The idea is that we can evaluate an act in two ways: from a consequentialist point of view according to the states of affairs it produces; from a deontological point of view according to the way it treats its object. It is the latter that can make the act morally reprehensible or dirty-handed: if it fails to treat its object the way it should be treated. Now, deontological constraints, on the threshold morality view, normally constrain consequentialist concerns: they exclude them from being valid reasons. However, beyond a certain threshold (e.g. avoidance of great harm), the consequentialist considerations take over. Still, even in such cases, the deontological concerns remain in place as evaluative considerations: it is morally right to avoid great harm but inappropriate treatment remains inappropriate treatment. Hence the act, albeit morally right, remains morally reprehensible.

¹⁶ This is not the same as Williams (1985)’s thick (vs thin) ethical concepts where the *descriptive* aspect is crucial.

¹⁷ But see Alvarez and Hyman (1998) for a view that might conflict with the claim made in the text.

¹⁸ Nagel is perhaps the most well-known and earliest proponent of this view; see also his (1986). For a discussion in a specifically political context, see Coady (2008) on what he calls ‘balanced exceptionalism’. Of course, it is a big question what exactly ‘inappropriate treatment’ consists in. Different deontological theories answer this question differently. We don’t need to decide this matter here.

Finally, third, although these acts are morally acceptable and unacceptable at the same time, they are not blameless and blameworthy at the same time. This is because the chosen act is morally right, hence no blame is appropriate. The act is morally reprehensible but the proper response to this is not blame but *regret* or *remorse*. And the outsider's ('our') proper response are not resentment and indignation but fear and pity.¹⁹

No doubt, many theoretical questions can be asked about DIRTY, but as before, we try to stay clear of the general debate on the morality and phenomenology of moral dilemmas and of dirty-hands.²⁰ Our interest is primary in the application of DIRTY to legal entrapment. This is particularly so here since DIRTY appears to be a good candidate to use for our purposes. It has no place for wrongness, hence it is immune to the initial step in Nathan's criticism of the dirty-hands model. At the same time, it retains the dilemmatic nature of the choice situation and it even keeps some element of the tragic as well as a moral remainder in it (in those cases where the essentially reprehensible nature of the act derives from causing (perhaps irreparable) harm to someone). But is DIRTY applicable to legal entrapment - Kis's interest after all is in politics, not in policing? Let us go through his first two steps above (the third step, as before, follows from affirmative answers to the first two, thus we do not engage with it).

Do acts of legal entrapment have essential properties that make them morally reprehensible because they involve ways of treating their targets inappropriately? There is no explicit discussion specifically of the moral reprehensibility of legal entrapment in the literature. However, there is much discussion of entrapment's wrongness and some of the proposed candidate wrong-making features might serve our purposes. Some talk about

¹⁹ As Kis (1998, 265) notes, the appropriate phenomenology of dirty-handed acts lies between cases of faultless involuntary contributions to accidents (where only what Williams (1991) calls agent-regret is appropriate) and blameworthy wrongdoing (where guilt and blame are appropriate).

²⁰ The perhaps most pressing issue about Kis's presentation of DIRTY is the way consequentialist and deontological reasons relate to each other. Although strangely Kis doesn't name him, DIRTY bases their relationship on a particular interpretation of Raz's (1999) notion of exclusionary reasons. Raz's account is notoriously controversial and many interpretations exist. See Adams (2020) for a recent discussion.

manipulation (Dillof 2004), autonomy-undermining temptation (Hughes 2004), even sadism (Carlson 2007) with respect to entrapment; others argue that entrapment disrespects the targets by subverting their moral capacities (Howard 2016). These are all objections of a deontic kind where the wrongness of entrapment lies in some form of inappropriate treatment of the target of entrapment. But do any of these features constitute *essential properties*? This is a more difficult question to answer. But perhaps an affirmative answer is on the cards. For example, Howard (2016, 25) writes that

“entrapment is wrong, in part, because it *subverts* the moral capacities of entrapped persons. To subvert an agent’s moral capacities is to interfere with the agent’s practical reasoning in ways that increase the likelihood she will culpably choose to act wrongly. Such activity, I contend, is incompatible with respect for that agent. Specifically, it is incompatible with a core regulative attitude that respect enjoins: an attitude of support for the successful operation of others’ moral capacities.”

It is possible, we reckon, to construe Howard’s candidate feature – subversion of the moral capacities of entrapped persons – as a property that entrapping someone “has in all possible circumstances or, if you like, in all possible worlds” (Marcus 1996, 25). No doubt, the entrapment in question has to be qualified – since it is crucial for Howard that the target “will culpably choose to act wrong”. But probably the specific concept of *legal* entrapment can be analysed in this way. We prefer to leave this question – and similar queries related to other wrong-making candidate features - open.

Still, even if entrapment has the requisite essential properties, it also to be shown that entrapping someone is the right thing to do in the given circumstances. Kis, as noted, appeals to threshold deontology at this point. However, threshold-based views are always tricky since

where does the threshold lie? What we think is clear from Kis's own view is that the 'safest' way to surpass the threshold is if we can prove that great (although not necessarily irreparable) harm is at stake. But, similarly to what we said regarding RESIDUE, that arguably may not hold in many cases of legal entrapment; surely, there are many mundane cases when the criminal entrapped is not a person who could have caused 'great harm' in the future. On the other hand, it should also matter if the deontic 'mistreatment' is not very severe, i.e., if the inappropriate treatment through entrapment doesn't amount to a significant deontic violation – then perhaps the consequentialist aim doesn't have to be set very high.²¹ In sum, as with the question of having the requisite essential properties, it is fairly unclear how restricted in scope DIRTY would be if applied to legal entrapment.

Let us sum up this long section. Nathan has put forward the idea that the dirty-hands model of undercover policing is best described with certain features: morally dilemmatic, tragic, with a moral residue, in particular. We have made an attempt to see if there is indeed a way to flesh out Nathan's remarks into a full-fledged account that is also applicable to legal entrapment. We can't say that we have found a clear answer. All three accounts we considered – TRAGIC, RESIDUE, DIRTY – are restricted in scope: they can't cover all or even nearly all cases of legal entrapment. Besides, especially TRAGIC and RESIDUE have further serious difficulties of application. Perhaps DIRTY is the account that is best applicable to legal entrapment. However, crucially, DIRTY has no place for moral wrongness in it, which, by appearance, is what Nathan's subsequent argument against the dirty-hands model places the most weight on. In short, in probably some or perhaps many cases, the dirty-hands model simply cannot be made to apply to legal entrapment, whereas in those cases when it does apply, it might do so along the lines of DIRTY that lacks the feature Nathan's criticism relies on: moral wrongness.

²¹ This assumes that deontic violations are gradable. This is controversial even for consequentialists, although see Peterson (2013) who argues for deontic degrees in a consequentialist framework.

IV. Dirty-handed acts, public justification and the police

Still, understood along one of the above accounts, the dirty-hands model might apply to some instances of legal entrapment. What is more, given what might be at stake - irreparable or great harm to be done if entrapment is not used, for example – it could be particularly important to see if Nathan’s argument against the model holds. Recall his reasoning. Given this model of undercover policing:

- 1) The public would correctly feel that morally wrongful acts are at the centre of police practice;
- 2) Despite the justification of these acts, public unease will remain;
- 3) The police becomes, as a result of internalization of this ethic, even more secretive.

In contrast:

- 4) The belief on the part of the police that undercover policing is not wrongful would be more conducive to publication justification.

Now, it is clear from our discussion so far that the dirty-hands model doesn’t have to construe dirty-handed acts as morally wrong: one of the accounts, DIRTY has no place for moral wrongness. However, Nathan could say that the relatively fine analytical difference between a morally wrongful act and, say, a morally reprehensible act that is not morally wrong is not enough to refute his 2) and 3) above. This might be so. After all, it is the common public and not the analytical philosopher who stands in the focus of Nathan’s predictions in these two points and we cannot now run large-scale empirical experiments to see how people react to a police practice that is evaluated along these lines publicly.²² Moreover, we take it that the

²² In fact, Nathan’s points could, arguably, have a bite even if the public’s perception of dirty-handed acts being morally wrong/reprehensible were totally wrong. However, this would create an entirely different situation and, we take it, it is not merely a slip of the tongue when Nathan writes that the public is *correct* in their perception of the morality of these acts.

dialectical context for Nathan's criticism is methodological: if one can come up with an equally good moral framework for analysing - in our case - legal entrapment and this framework doesn't have 2) and 3) as consequences (because 4) is true in it), that framework is better. Thus, it would be good to see what an advocate of the dirty-hands model could say to these criticisms.

Here is a possible response.²³ One can try to turn around Nathan's reasoning. He emphasizes the *consequences* of embracing the dirty-hands ethic for public as well as police morale. But one could point out that that very ethic he targets has resources in it to control these detrimental effects and that these resources centre exactly on the notion of what Nathan finds badly effected: public justification. That is, *exactly because* the acts involved are dirty-handed, public accountability, in institutional as well as in non-institutional form, is placed at centre stage in the dirty-hands model. The basic idea is simple: in the very model that Nathan targets, there is centre place for public accountability that can counteract the negative empirical tendencies emphasized by Nathan.

Paradoxically, and contrary to what Nathan suggests, public justification of dirty-handed acts would be easy to argue for if we held that these acts are morally wrong (as they are, if covered, on TRAGIC and RESIDUE). If this was the case, we would expect that the agent feels guilt and we would be expected to feel indignation; blame and condemnation were also appropriate. Both TRAGIC and RESIDUE would take these to be the appropriate reactions and would require some form of public penance as a result. In short, if 1) is true, 2) and 3) wouldn't at all obviously follow simply because it is part of the dirty-hands model to put the relevant *public* controls in place.

However, if, as proposed by DIRTY, (justified) dirty-handed acts are not morally wrong but 'only' morally reprehensible, then, the relevant reactions by the agent are regret and

²³ Arguably, there is another, more radical response: to hold the true morality must not be publicly affirmed but must be kept secret. It is clear Nathan implicitly assumes a publicity condition regarding moral rules that govern our – in this case, law-enforcement – conduct. However, not everyone accepts this. For a recent defence of such 'esoteric morality', see Lazari-Radek & Singer (2010).

remorse, and by the observer compassion, fear and pity. Consequently, although - since a (deontic-constraint-violating) mistreatment has occurred - the agent would owe explanation to the target (victim) of her acts, specifically public accountability would not follow. The agent would be, in some sense, accountable to her own conscience, but not to the public.

This way of approaching the reaction to dirty-handed acts is in line with some traditional accounts of the morality of the political leader. Machiavelli (1988) famously argues that the good prince should learn to be bad and on this ‘renaissance model’ (as Walzer 1973 calls it), the agent has no inner life (recall Nathan: ‘not to dwell on his misdeeds’) – we don’t know what is happening there and, Machiavelli seems to suggest, we don’t need to know. Max Weber (1994) goes one better than this by depicting the political leader as suffering internally due to the choices she must make. Still, this is an entirely private experience. Finally, Walzer himself suggests what he calls a ‘catholic model’: the suffering (a la Weber) should be socially expressed; the politician should be socially allowed (required?) to purify himself of his sins, repent and achieve salvation. There should be a secular public authority to offer this opportunity, although Walzer doesn’t specify this.

This is naturally not enough for us. However, Kis (2008, Chapter 8) forcefully argues that for several reasons public justification is needed even if dirty-handed acts are not morally wrong. Although his focus is on politics, most of these reasons are also applicable to policing and entrapment. First, there is the problem of *moral corruption*. On the one hand, in policing (or in politics) we need - if we go along with the idea that there are morally justified dirty-handed acts - people who are willing to dirty their hands. On the other hand, we don’t want that these people dirty their hands *too easily*. “Power corrupts and absolute power corrupts absolutely” – is also true in policing. It shouldn’t get too easy for police officers to dirty their hands: this threat of moral corruption is very much prevalent in the literature on noble cause corruption (e.g. Alexandra 2000, Miller 2016 who speak about the moral negligence, arrogance

and insularity of police officers). This is further underlined by a second reason that we want to mention here since it also applies to policing: *uncertainty*. No police officer can be sure that when they entrap, their reasons are indeed good. Hence, just as with moral corruption, we do not want to make it easy for police officers to act on their reasons, however good they are conceived to be. The two considerations also connect: those who are more easily inclined to dirty their hands, are also more likely not to care about the fact that they might be wrong.

It is arguable that the best solution is therefore *public accountability*. Typical elements of a liberal (constitutional) democracy can all be mentioned here: freedom of press, freedom of speech, independent courts and so on. The media, of course, is particularly important.²⁴ But in the case of the police, other, less general institutional measures are also important: all the ways of overseeing police work, both internal and external; the extensive discretionary rights of police officers that allow them to reflect upon their practice on a case by case basis; their original authority coming directly from the law that makes them legally accountable for their actions; all sorts of other procedural barriers on police work; and so on.

If this need for public accountability stands, then 2) in Nathan's argument is much less of a danger. The point of public accountability is exactly to make dirty-handed acts (relatively) rare occurrences; to make sure that they are costly enough to be considered carefully before carried out. This can be because once publicly known, the acts become reasons to end someone's career, for example, because the public doesn't accept such acts despite their moral justification (the acts can also be made illegal, of course, as a result of such public uproar or even 'ostracization'). At the minimum, public accountability should bring transparency and critical public scrutiny into law-enforcement practice.

Step 3) in Nathan's is trickier since he explicitly refers to increased secrecy, i.e., increased willingness to avoid public scrutiny. But this misses our point and the nature of the

²⁴ In fact, relevant for journalistic entrapment, the media itself is arguably subject to a 'publicity condition' concerning e.g., the methods they use; see Bovée (1991).

present reasoning. While Nathan puts forward a slippery slope kind of argument, i.e., a claim about what (morally wrong) acts will cause what (secretiveness, less willingness to subject yourself to public justification, less willingness to dwell upon your deeds), we are saying that to endorse the moral justification of dirty-handed acts is *only* possible if parallel to this, we, as a society, also put in place a structure that makes sure that such acts are costly, rare, and well justified. It is then exactly this very structure of public accountability that will make sure that we do not begin our descent down the kind of slippery slope Nathan describes.

V. Summary and concluding remarks on alternatives

In this paper we have focused on a possible framework for analysing the morality of legal entrapment (which we have defined based on our previous work): the dirty-hands model. We took as our starting point Nathan's (2017) criticism of the model (when applied to undercover policing). We had two aims throughout the paper. Our primary aim was to see if the model applies at all to legal entrapment; our secondary aim was to establish whether, if the model applies, Nathan's criticism hold for it. Regarding the first aim, we have presented three possible versions of the model, loosely taking our inspiration from Nathan's remarks and using the work of Kis (2008). We have found that the first two accounts – we called them TRAGIC and RESIDUE - don't apply well to legal entrapment and the last, perhaps more plausible but still questionably applicable account – DIRTY- has no room for morally wrong acts. Regarding the second aim - which we have investigated since at least in some important instances one or more of these accounts might well be applicable to legal entrapment - we have argued that Nathan's criticism of the model is not so forceful once we take account of all the resources available in the dirty-hands model.

Could there be an alternative to the dirty-hands model that is also not too far away from its 'spirit'? There are of course, as we noted, many accounts that either claim that entrapment

is wrong or that, like Nathan, hold that it is right, at least in some cases. The question is whether there are theories out there that somehow keep the paradoxical, morally complex analysis that the dirty-hands model arguably offers. In ending, let us, briefly, offer one alternative with some accompanying, and even briefer, critical considerations (mainly terminating in open questions to be pursued outside this paper).

The candidate in question is admirable immorality. To quote from Curzer (2002, 229): “Acts are *admirably immoral* when they are (a) somehow great, (b) morally wrong, and (c) these two features are intrinsically connected.” In contrast, “People have *dirty hands* when they perform acts that are both morally required and morally repugnant.” That is, Curzer’s view of dirty-hands is similar to DIRTY in that it doesn’t take dirty-handed acts to be morally wrong, whereas he emphasizes that instances of admirable immorality have moral wrongness in them. Admirable immorality would thus be a way of reintroducing moral wrongness in our analysis of legal entrapment while also keeping the paradoxical complexity (and tragicness) of the original dirty-hands analysis. Finally, Curzer’s requirement (a) might also be fulfilled in those instances of legal entrapment when irreparable or great harm can be avoided in this way.

But is admirable immorality applicable to legal entrapment? This is not obvious. Curzer (2002; cf. 2006) discusses three candidates for admirable immorality. The first arises from conflicts between morality and *other value systems* (religious, aesthetic etc.) Take e.g. Abraham’s choice to sacrifice Isaac: this could be construed, as Kierkegaard did (according to Curzer), as a choice between acting immorally (sacrificing Isaac) and acting sacrilegiously (not sacrificing Isaac).²⁵ What matters, from our point of view, is that cases of legal entrapment can hardly be construed as choices between morality and some other value system. What would that alternative value system be? Perhaps some kind of societal prudence – but then this could be understood as a utilitarian, hence moral, consideration, we submit.

²⁵ For a recent analysis of Kierkegaard that takes a roughly similar line, see McDaniel (2020). Michael Slote (1983) promotes the same idea as a way of showing that morality is not overriding.

Curzer's second candidate construes admirably immoral acts as conflicts between morality and certain *role moralities*. Some of these role moralities are obviously immoral: think of the thief or the mafioso. But other role moralities are not as such immoral although they can give rise to immoral acts: think of the role of the leader who, like Churchill or Agamemnon, might have a single-minded devotion to the national interest. Curzer mentions lawyers, so could we list also policemen here? In fact, some construe policing as a role-based practice; hence, bringing in role moralities appears natural here. There are two problems with this proposal in the present context. One is that whether the idea that a specific role morality is indeed best construed as somehow different from morality (without the qualifier) is unclear, both in general and in the specific case of legal entrapment. For example, Alexandra (2000) talks about a technical division of labour in which policemen pursue a specific kind of role morality which consists in occasionally making their hands dirty – understood roughly along the lines of RESIDUE. This kind of construal wouldn't give us admirable immorality. Also, one could, as Cooper (2012) does, look at what policemen do in the case of noble cause corruption²⁶ – and by extension perhaps, legal entrapment – as a conflict of roles: between, e.g., the role of protector and the role of respecting procedural rights. This way of interpreting legal entrapment could again be understood as a dirty-hands model, along the lines of RESIDUE, perhaps.

The third case of admirable immorality Curzer discusses is that of *virtue-virtue* conflicts. He gives the Roman commander Manlius's choice to hand over his son to the courts to be sentenced to death as an example. However, Curzer himself notes that he understands this conflict as a case when admirable immorality meets dirty hands. In particular, he argues that although it is tempting to construe the conflict along the lines of TRAGIC, he prefers to remove moral wrongness from the equation and embrace the idea, as on DIRTY, that an act can be

²⁶ Noble cause corruption is corruption in the service of a good end where 'corruption' is understood to be morally wrong. In the case police practice, it can consist in such acts as fabricating evidence, beating up suspects, committing perjury and so on. The classic example comes from the main character in the movie *Dirty Harry*. See Miller (2016) for details.

morally repugnant while not morally wrong. Whether he succeeds in earning the right to this move is a question²⁷, but for us there is the further issue whether legal entrapment could fit the bill. As Curzer sees it, what we should see then in the case of legal entrapment is a conflict between the ethics of care and the ethics of justice, which he in turn takes to be simply descriptions of the virtues of justice and care. But is it obvious that what happens in the case of legal entrapment is that these two virtues conflict? We don't think so. Just to mention one rather obvious problem. While perhaps entrapping the target could be construed as a form of justice, the virtue of care comes typically into force with respect to concrete, particular individuals whom we especially care about. Who would these be in the case of legal entrapment, or more precisely, in the case of choosing not to entrap?

Thus, we are not convinced that legal entrapment is best construed as a case of admirable immorality, although many of the questions above are genuinely open to further elaboration. As we have shown, attempting to a dirty-hands type of analysis is also probably not the best way to go, although some instances of legal entrapment might well be covered in this way. At the same time, there is something in the moral complexity of (legal) entrapment that makes these avenues of analysis attractive and hopefully subject of further research.

Acknowledgements. [...]

²⁷ As is perhaps clear from what follows in the text, unlike Kis, Curzer establishes this difference not by using the concept of essential properties but by arguing for a conflict of virtues.

REFERENCES

- Adams, N. P. (2020). 'In Defence of Exclusionary Reasons.' *Philosophical Studies*, forthcoming (online first).
- Alexandra, A. (2000), Dirty Harry and Dirty Hands. In Tony Coady, Steve James, Seumas Miller and Michael O'Keefe (eds.) *Violence and Police Culture* Melbourne: Melbourne University Press, 2000, pp. 235-248
- Alvarez, M. & Hyman, J. (1998), Agents and Their Actions. *Philosophy* 73 (284): 219-245.
- Bülow, W., Helgeson, G. (2018), Hostage Authorship and the Problem of Dirty Hands, *Research Ethics*, 14(1): 1-9
- Carlson, A. (2007). Entrapment, Punishment, and the Sadistic State. *Virginia Law Review* 93: 1081-1134.
- Dillof, A.M. (2004). Unraveling Unlawful Entrapment. *Journal of Criminal Law and Criminology* 94: 827-896.
- Cooper, J.A. (2012), Noble Cause Corruption as the Consequence of Role Conflict in the Police Organisation, *Policing and Society* 22(2): 169-184
- Curzer, H.J. (2002), Admirable Immorality, Dirty Hands, Care Ethics, Justice Ethics, and Child Sacrifice, *Ratio (new series)*, 15(3): 227-244
- Curzer, H.J. (2006), Admirable Immorality, Dirty Hands, Ticking Bombs, and Torturing Innocents, *The Southern Journal of Philosophy*, 44: 31-56
- Dancy, J. (2004). *Ethics Without Principles*. Oxford: Oxford University Press.
- Gowans. C. (1994), *Innocence Lost*. Oxford: Oxford University Press.
- Hare, R.M. (1981), *Moral Thinking*, Oxford: Clarendon Press
- Howard, J.W. (2016), Moral Subversion and Structural Entrapment, *The Journal of Political Philosophy* 24(1): 24-46

- Hughes, Paul M. (2004) 'What is Wrong with Entrapment?'. *Southern Journal of Philosophy*, 42 (1), 45–60.
- Kis, J. (2008), *Politics as a Moral Problem*, Budapest, Hungary: CEU Press
- Lazari-Radek, K. & Singer, P. (2010), 'Secrecy in Consequentialism: A Defence of Esoteric Morality', *Ratio* 23: 34-58
- Machiavelli, N. (1988). *The Prince*. Eds. Q. Skinner and R. Price. Cambridge: Cambridge University Press.
- Marcus, R.B. (1996), More About Moral Dilemmas. In H.E. Mason: *Moral Dilemmas and Moral Theory*. New York: Oxford University Press, pp. 23-36.
- McDaniel, K. (2020), Teleological Suspension in *Fear and Trembling*. *Philosophy and Phenomenological Research* 100 (2): 425-451.
- Miller, S. (2016), Noble Cause Corruption in Policing, in. S. Miller (ed.), *Corruption and Anti-Corruption in Policing – Philosophical and Ethical Issues*, Springer Briefs in Ethics, 39-51
- Nagel, T. (1979), War and Massacre, in. T. Nagel: *Mortal Questions*, Cambridge University Press, pp. 53-74
- Nagel, T. (1986). *The View from Nowhere*. New York: Oxford University Press.
- Nathan, C. (2017), Liability to Deception and Manipulation: The Ethics of Undercover Policing, *Journal of Applied Philosophy* 34(3): 370-388
- Nielsen, K. (2000), There Is No Dilemma of Dirty Hands, in. R. Rynard and D. Shugarman (eds.), *Cruelty and Deception*, Broadview Press, pp. 139-155
- Portmore, D. (2007), Consequentializing Moral Theories. *Pacific Philosophical Quarterly* 88: 39-73.
- Raz, J. (1999). *Practical Reasons and Norms* (2nd ed.). Princeton: Princeton University Press.
- Slote, M. (1983), Admirable Immorality, *Goods and Virtues*, Oxford: Oxford University Press.
- Peterson, M. (2013), *The Dimensions of Consequentialism*, Cambridge University Press.

Stocker, M. (1990), *Plural and Conflicting Values*. Oxford: Clarendon Press.

Walzer, M. (1973), Political Action: The Problem of Dirty Hands, *Philosophy and Public Affairs* 2: 160-180

Weber, M. (1994), Politics as a Vocation, in M. Weber: *Political Writings*, Cambridge University Press

Williams, B. (1973), Ethical Consistency, in. B. Williams: *Problems of the Self*, Cambridge University Press.

Williams, B. (1985). *Ethics and the Limits of Philosophy*. Cambridge, Mass.: Harvard University Press.

Williams, B. (1991). Moral Luck in *Moral Luck*. Cambridge: Cambridge University Press.