ENTRAPMENT AND ITS ETHICS: A DIRTY HANDS PROBLEM?

Daniel Hill, Stephen McLeod, Attila Tanyi

I. **Classification of entrapment** (from Hill, McLeod, Tanyi 2018)

"Agent" – the entrapping party
"Target" – the party that the entrapping party intends to entrap

	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

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

My interest is in Type 1 but I will also remark occasionally on media entrapment (Type 2 or 3).

II. **Definition of legal entrapment** (from Hill, McLeod, Tanyi 2018)

Legal entrapment to commit a crime occurs whenever:

(i) a law-enforcement agent (or their deputy), acting in their capacity as (or as a deputy of) a law-enforcement agent, *plans* that the target commit an act;
(ii) the planned act is of a type that is criminal;
(iii) the agent *procures* the act (by solicitation, persuasion or incitement);
(iv) the agent intends that the target's 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 a third party to be enabled, *to prosecute or to expose the target* for having committed the act.

Entrapment (without the qualifier) differs from this definition in that condition (ii) is more inclusive. Also, it is useful to distinguish entrapment from, first, *mere (virtue) testing* (where (i) is not present) and, second, *mere temptation* (where (iii) is not present since there is no communicative act involved).

III. **Dirty hands and legal entrapment: Nathan's criticism**

Nathan (2017) has recently argued against "the dirty hands model" of undercover policing. Here is how he describes the model (NB: in fairness, this is not Nathan's main target in his article, his focus is elsewhere):

"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." (371; Italics are mine)

My idea is to see whether the model can also be applied to the case of legal entrapment. This seems to be a natural move since the way Nathan describes the model appears to fit well with the structure of legal entrapment as defined above. In particular, many argue that legal entrapment is morally wrong for various reasons (see a recent suggestion and a list of other reasons in Howard 2016), while it is clear that the declared aim of entrapment is prevention/detection/reduction of crime.

However, Nathan himself introduces the model only to side-line it as wrong:

"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 *publication justification*." (Ib.; Italics are mine)

Nathan's critical points:

1. (Morally) wrongful acts would be at the centre of police practice.
2. This can cause public unease.
3. The police becomes, as a result of internalization of this ethic, even more secretive and wary of public justification.

I would like to see if this criticism can be extended to legal entrapment.

I think these are all relevant and important points but I also think that 1 relies on a particular understanding of dirty hands, whereas the truth of 2 and 3 largely depend on 1 ['largely' is actually not correct – Nathan makes it sound as if without 1, 2 and 3 wouldn't arise as problems, but it is not clear whether his use of 'wrongness' is based on proper conceptual discrimination or 'wrong' is more of an umbrella term for anything morally inappropriate or unacceptable].

I would like to see – with a very open-mind! – whether we must indeed construe the dirty-hands model as relying on 1 and whether 2 and 3 have any truth to them (whether or not 1 is correct) in the case of legal entrapment. I certainly would like to get to the end of my analysis of 1 and I hope to say few words on 2 and 3 as well.

IV. How to understand the problem of dirty hands?

It is useful to rehearse the main elements of Nathan's description of the dirty hands model:

- Moral wrongs are committed;
- Genuine moral dilemmas are involved [I take this to refer to a choice situation in which the agent is confronted with moral oughts/reasons (these can be duties etc.) and whatever course she takes that will be, in some sense, morally unacceptable];
- 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.

I would like to argue for three points: (1) Nathan takes the dirty hands model to be a case of moral dilemma and this isn't obviously true; (2) Even if the dirty hands model is a case of a moral dilemma, only on a certain understanding of moral dilemmas do we get the claim that moral wrongs are committed; (3) Although there are alternatives to dirty hands that we also have to consider (since they keep moral wrongness in the picture), this doesn't help Nathan out.

I begin with (2) and then move on to (1) and, finally, (3). (Arguably, (1) has only an instrumental role: it leads us to my preferred picture of moral dilemmas.) Here are four different ways of understanding moral dilemmas (I follow in large part Kis 2008, others are separately indicated). Let us start with the one that fits best Nathan's description above.

<u>The tragic account (TRAGIC)</u>
"Sometimes it is right to do what is wrongful." (Bülow and Helgeson 2018) or "Whatever you do is wrong." This does appear to be tragic; in fact, many would say that this brings incoherence/inconsistency into the structure of morality.

When Walzer (1973) introduced the dirty hands problem (in politics), he emphasized the paradoxical nature of the phenomenon (this is a matter of interpretation, but see e.g. Coady 2008 or Curzer 2006). Others, such as Alexandra (2000), describe more extreme instances of noble cause corruption in this way (namely, the original Dirty Harry problem).

How would this look like more formally? Something like this (from Kis 2008):

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.

(A1-3 presupposes, when applied to legal entrapment, that there is no other way of achieving the good end (reduction/prevention/deterrence of crime, security) but by entrapping. But such necessity and corresponding last resort conditions are standard in the literature on entrapping, I take it. See e.g. Bovèe 1991 on media ethics or Coady 2008 on politics. However, in real life this might well restrict the number of cases of entrapment that can be analysed this way.)

This gives us moral conflict but it doesn't give us TRAGIC. For that we must add:

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.

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

However, there are several problems with this interpretation:

1. It is far from clear that TRAGIC would be the correct description of all or even just most cases of legal entrapment. Sure, law enforcement officers might be doing what is *normally* considered wrong *overall (all-things-considered)*, but in the particular context, this is far from clearly so given all the good that will be achieved through the act. In short, there is reason to hold that A4^{TR} doesn't hold in many or even most cases of legal entrapment. [Cf. Alexandra on typical cases of noble cause corruption; Walzer is unclear on exactly what situations we can consider here to be covered but his focus is mostly on emergencies.]
2. Nathan, probably following Walzer (whom he references in a footnote), talks about a *moral residue* as resulting from the wrongness of, in our case, the entrapping act. But this is a

rather awkward construction. If the agent has done wrong, then the agent, as per $I3^{TR}$ and $I4^{TR}$, is not innocent but guilty and this isn't a moral remainder somehow, I think. [Cf. Coady 2008: Arguably, the moral residue here might be referring more to the phenomenology of these cases, i.e., what the agent experiences (the pulling of opposite forces, the psychological pain whichever way she decides). But even then, it is far from clear that the feeling of guilt is the best way to go in accounting for this residue. We should also consider that we are often not capable of high degree of emotional discrimination – to decide whether we feel guilt, remorse, regret, culpable etc.]

3. A moral general problem with TRAGIC is that it is not compatible with ought-implies-can (OC): If *S* cannot do *a*, then it is not the case that *S* ought to do *a*. Now, coupled with $I1^{TR}$, "the implication is that *S* must be capable of doing *b*. However, A3 entails that if *S* does *a*, she is not capable of doing *b*. Suppose that *s* ought to do *b* even if she does *a*. Then, in virtue of OC, *S* can do *b*. In virtue of A3, however, *S* cannot do *b*. Therefore, *S* both can and cannot do *b*." (Kis 2008, 242)
4. TRAGIC also severs the ordinarily confirmed connection between blame and responsibility. Given that *S* cannot but violate an in-force ought, *S* cannot act blamelessly. However, the ordinary view is that when one has no option but to do something morally wrong, one is not to be blamed for one's choice.

Thus, I think TRAGIC, despite its fittingness to what Nathan says, is not a good way to take for advocates of the dirty-hands model. 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 be a better choice.

The moral residue account (RESIDUE)

Williams (1973) has argued that what happens in moral dilemmas is that one ought overrides other oughts but that the overridden, defeated oughts are not silenced: they 'stick around', their force doesn't evaporate. In particular, these defeated oughts give rise to *derivative oughts* to compensate, to repair. "The defeated ought has no action-guiding force in the immediate context of the situation in which the choice is being made, but it has action-guiding force in the context of a lager that emerges in virtue of *S*'s action." (Kis)

In short, the decision situation is more complex than in TRAGIC. We have *S* who, if she chooses *b*, must then do c_a (where this represents compensating for/repairing the harm caused by failing to do *a*), and if she chooses *a*, must then do c_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. [For example, RESIDUE can in this way handle tie-breaking situations: if the oughts involved are equally strong, the compensation requirement can break the tie.]

RESIDUE doesn't encounter the problems that beset TRAGIC. In particular, it appears to conform to the way I've depicted the structure of legal entrapment in my first problem with TRAGIC above. [RESIDUE is supported naturally by an intuitionist/Rossian (e.g. Dancy today) picture of competing *pro tanto* oughts the balancing of which gives us an all-things-considered ought judgment. See also Alexandra who depicts ordinary cases of noble cause corruption exactly this way.]

However, RESIDUE faces other difficulties:

1. There appear to be no place for wrongness in it – hence its tragic character can be questioned. 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. *S* simply did the right thing, on both levels (acting and then compensating); the/her moral universe remains intact and she comes out of the situation (morally) innocent.

2. One could try to get around this problem by claiming that somehow even acting on the overriding ought amounts to acting wrongly (Coady 2008 interprets Walzer in this way, e.g.) But this produces just another version of my second problem with TRAGIC: I find it difficult to accept that this is the relevant moral residue in question, esp. given the new kind of framework (competing *pro tanto* oughts) employed here.
3. There is also the question of how exactly compensation/reparation would look like in legal entrapment cases (even more interesting is how it would look like in civil entrapment cases). On RESIDUE, this is crucial, of course – could we encounter instances of legal entrapment when compensation is not feasible?

This last problem is more important than it looks. For, as Kis points out, 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, contrary to my first problem above, the tragic character of *S*'s choice situation (not entirely, since on RESIDUE, *S* would still not be considered to have done the wrong thing). We would get the following depiction of the moral residue account:

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.
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.

But replacing A4^{TR} with A4^{MR} produces its own problems:

1. As Kis points out, if a damage is irreparable, i.e., if it *cannot* be repaired, then, by OC, it *ought not* be repaired. 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, residue remains but it oughtn't be (because cannot be) repaired/compensated for. In either case, no (derivative) ought remains in the final analysis that could be violated, triggering an at least somewhat tragic conclusion.
2. Kis does argue 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 that of guilt, perhaps not even of regret or remorse, but still, *S* can think of her act as morally reprehensible and feel accordingly. Perhaps this is true. But even if it is true, it gives only a very thinly tragic analysis: whatever *S* does, it is appropriate for her to feel bad about her choice of act. Is this enough?
3. We should not forget that we are interested in applying RESIDUE to legal entrapment. The original version of RESIDUE was eminently applicable (as I've pointed out). But the present version is overly restrictive: only those instances of legal entrapment come under it that involve irreparable damage. These surely are very 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.
4. Kis shows that reference to irreparable damage cannot be what constitutes moral dilemmas because hard choices 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).

V. **Must dirty handed acts be responses to moral dilemmas?**

If we side-line TRAGIC and RESIDUE, the obvious choice is to go for a non-dilemmatic reading of moral dilemmas. Well-known such attempts are Hare's (1981) or Nielsen's (2000). One obvious challenge to these views is that they cannot explain the phenomenology of moral dilemmas and thus of dirty hands. They of course try and one way they do it is particularly fitting for police work (as well as for media, see Bovèe 1991): to emphasize the *uncertainty* – the *doubt* – that accompanies such work, including acts of entrapping. Have I really done the right thing? Am I really going to achieve the good end? And so on. It is this sense of doubt that advocates of this approach want to identify with the 'bad feeling' that accompanies dirty-handed acts.

We can formalize this view (DOUBT) as follows (following Kis 2008):
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.
A4^{MD}: *S* is uncertain whether the requirement to do *a* overrides the requirement to do *b*, or whether the requirement to do *b* overrides the requirement to do *a*, or else whether both requirements are non-overridden.

Based on this, the implication, regarding phenomenology, is then claimed to be this:

I1^{MD}: Even if *S* should permissibly do *a* (or *b*), it may be appropriate for her to be haunted by doubts as to whether doing *a* (or *b*) was not wrong after all.

However, whether or not this is the correct account of the phenomenology of moral dilemmas (and, as mentioned, there are serious doubts), it doesn't help to criticize the dirty hands model along the lines Nathan suggests. In fact, from our point of view, DOUBT throws the baby out with the bathwater. On DOUBT, there are no moral dilemmas and no hands get dirty: there is one right way to act. There is also no moral residue and corresponding compensatory duty. Hence, since my idea is to see the application of the dirty hands model to legal entrapment, this cannot be the way to go.

Kis (2008) proposes a way out (in the case of politics) that can keep the dilemmatic nature of the choice situation without introducing wrongness. It also has space for some account of what is tragic about the choice. In short, it offers a way to answer Nathan's criticism. The question is whether it is applicable to the case of legal entrapment.

But, first, here is the account:

<u>The dirty hands account (DIRTY)</u>
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.
A4^{DH}: At least *a* is morally reprehensible, and *b* is either morally reprehensible or it is not morally overriding.

The implication concerning the agent's reactive emotions to the situation:

I1^{DH}: Even if what *S* does is no worse than any of its alternatives, it is appropriate for *S* to feel remorse about her action.

How does Kis gets here? He argues for three points.

1. 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.
2. Consequently, such acts are morally acceptable (because right) and morally unacceptable (because reprehensible) at the same time. This is possible because we can endorse what some call *threshold deontology* (Nagel 1979) or balanced exceptionalism (Coady). 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 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, the consequentialist considerations take over (e.g. avoidance of great harm). Still, even in such cases, the deontological concerns remain in place as evaluative considerations: it is right to avoid great harm but inappropriate treatment remains inappropriate treatment. Hence the act, albeit right, remains morally reprehensible.

3. 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 act is 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.

As I say above, DIRTY appears to be a good candidate to use for our purposes. It has no place for wrongness, hence it is immune to Nathan's criticism. 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 (through the act being morally reprehensible and remorse/regret being appropriate morally).

But is DIRTY applicable to legal entrapment (Kis's interest is in politics, not in policing)? Let us go through his three points above to at least locate the questions to be asked:

1. Do acts of legal entrapment have essential properties that make them morally reprehensible because they involve ways of treating their targets inappropriately? Judged from the extensive literature on the wrongness of entrapment, the answer might well be affirmative. Some talk about manipulation, deception, even sadism with respect to entrapment; others argue that entrapment disrespects the targets by subverting their moral capacities (Howard 2016). However, while these are candidate wrong-making feature, our question is whether entrapment has *essential properties* that make it reprehensible along these lines. To me it seems that, at least on some views of entrapment's 'wrongness', this question can be answered in the affirmative. But I do not have argument to this effect…(I much like Howard's deontic candidate, moral subversion.)
2. This is also not easy. Threshold-based views are always tricky since where does the threshold lie? What I think is clear from Kis's own view is that the 'safest' way to surpass the threshold is if we can prove that great harm is at stake. That arguably may not hold in many cases of legal entrapment (a point. I've already made); 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 (this, of course, assumes, controversially I suppose, that deontic violations are, like consequentialist assessments, gradable) – then perhaps the consequentialist aim doesn't have to be set very high.
3. This is fine if 1 and 2 turn out to be fine.

VI. Moral wrongness without dirty hands?

Thus, perhaps Kis's proposal can help us to give an account of dirty hands that fits legal entrapment without putting the wrongness of entrapment in the centre of analysis. [As I say above, though, this is a rather strong 'perhaps'.]

But could Nathan's criticism be saved in some other way without appealing to dirty hands? What alternatives are there and how do they relate to the dirty-hands model? I see two options.

Admirable immorality

To quote from Curzer (2002), “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.” [I quote Curzer on dirty-hands only to show that his account is similar to the above in that it doesn’t take dirty-handed acts to be morally wrong.] Thus, if one can show that instead of being dirty-handed, acts of legal entrapment are instances of admirable immorality, one could bring moral wrongness back in the picture.

However, this is hard to show. Curzer (2002, 2006) lists 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, as a choice between acting immorally (sacrificing Isaac) and acting sacrilegiously (not sacrificing Isaac). [Of course, we are now setting aside the idea of a God-based morality such as divine command theories. This could be done for good reasons, though.] Michael Slote (1983) promotes this idea as a way of showing that morality is not overriding. This doesn’t have to concern us here. 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 be? Legality? But, of course, many argue that entrapment is illegal…

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 also mentions lawyers, so could we list also policemen here? In fact, some (many? most?) 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 above. 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 corruption – and by extension, legal entrapment – as a conflict of roles: between the role of protector and the role of respecting procedural rights (e.g.). This way of interpreting legal entrapment could again be understood as a dirty-hands model, along the lines of RESIDUE above.

The third case of admirable immorality Curzer discusses is that of *virtue-virtue* conflicts. He gives 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. Moreover, he argues that although it is tempting to construe the conflict along the lines of TRAGIC above, he prefers to remove moral wrongness from the equation. Whether he succeeds in this is a question, but for us there is the further issue whether legal entrapment could fit the bill and I don’t think so. Ethics of care vs. ethics of justice in the case of legal entrapment?

Multi-dimensional consequentialism

Thus, I am not convinced that legal entrapment is best construed as a case of admirable immorality. But there may be a second way to keep moral wrongness without endorsing the dirty-hands model. Peterson (2013) promotes a version of consequentialism – he calls it *multi-dimensional consequentialism* – that handles moral dilemmas understood along TRAGIC differently. Namely, Peterson works out a theory on which there are deontic degrees – an act can be somewhat right and somewhat wrong at the same time (the master argument given by him is that this is the only way to avoid what he calls ‘deontic leaps’). Applied to TRAGIC, this would mean that, indeed, all options in a moral dilemma are morally wrong but they are not equally wrong; instead, they are wrong to some (varying) extent only.

Two problems with this theory. One is general, see my work with Andric (2016) in which we criticize Peterson’s master argument and also provide an alternative that is more orthodox than his proposal. The other is specific: would it be enough for Nathan’s criticism if it turned out that options in a moral dilemma are wrong but, say, only to a small extent?

VII. **Dirty-handed acts, public justification and the police**

Let us go back to Nathan’s critical points:

1. (Morally) wrongful acts at the centre of police practice.
2. This can cause public unease.
3. The police become even more secretive and wary of public justification if they internalize a dirty-hands ethic. (They will learn ‘not to dwell on their mistakes’.)

So far, my strategy was to refute 1 by showing that the dirty-hands model doesn’t have to construe dirty-handed acts as morally wrong. However, Nathan can try to say that the relatively fine analytical difference between a morally wrongful act and, say, a morally reprehensible act is not enough to refute his 2 and 3. This might be so – to investigate its truth, it seems to me, would require large-scale empirical experiments to see how people react to a police practice that is evaluated along such lines publicly. Moreover, I take it that the dialectical context for Nathan’s criticism is methodological: if we 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, that framework is better. [Nathan’s own proposal concerns liability.]

I am not able to assess here either of these claims. What I can do, however, is to try to turn around Nathan’s points: It seems to me that, based on the literature on dirty-hands, 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 (and again follows Kis): there is strong need for public accountability that can counteract the negative tendencies emphasized by Nathan. Perhaps the following four points are the most important and, with some modifications, applicable to entrapment:

1. Public justification of dirty-handed acts would be easy to argue for if we held that these acts are morally wrong. 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. However, if, as suggested earlier, (justified) dirty-handed acts are not morally wrong but ‘only’ morally reprehensible, then, as foretold, the relevant reactions by the agent are regret and 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, accountability would not follow. The agent would be, in some sense, accountable to her own conscience, but not to the public.
2. 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 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 name it.

3. This is naturally not enough for us. However, Kis 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 I want to mention here since it also applies to policing: *uncertainty*. We saw this already: no police officer (or journalist for that matter) can be sure that when they entrap, their reasons are indeed good (technically: not every motivating reason is also a normative reason). Hence, just as with moral corruption, we do not want to make it easy for police officers to act on their reasons. The two reasons 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.
4. 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. (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.) 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 Nathan's point 2 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). At the minimum, public accountability should bring transparency and critical public scrutiny.

Nathan's point 3 is trickier since he explicitly refers to increased willingness to avoid public scrutiny. But this misses my point. 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 willing to dwell upon your deeds), I am saying that to endorse the moral justification of dirty-handed acts is only possibly if parallel to this, we also put in place a structure that makes sure that such acts are costly, rare, and well justified. It is then exactly this very (institutional) structure of public accountability that will make sure that we do not begin our descent down on the kind of slippery slope Nathan describes.

REFERENCES

Alexandra, A. (2000), Dirty Harry and Dirty Hands, Tony Coady, Steve James, Seumas

Andric, V., Tanyi, A. (2016), Multi-Dimensional Consequentialism and Degrees of Rightness, *Philosophical Studies* 173(3): 311-331

Miller and Michael O'Keefe (eds.) *Violence and Police Culture* Melbourne: Melbourne University Press, 2000, pp. 235-248

Bülow, W., Helgeson, G. (2018), Hostage Authorship and the Problem of Dirty Hands, *Research Ethics*, 14(1): 1-9

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

Hare, R.M. (1981), *Moral Thinking*, Oxford: Clarendon Press

Hill, D., McLeod, S., Tanyi, A. (2018), The Concept of Entrapment, *Criminal Law and Philosophy* 12(4): 539-554

Howard, J.W. (2016), Moral Subversion and Structural Entrapment, *The Journal of Political Philosophy* 24(1): 24-46

Kis, J. 2008, *Politics as a Moral Problem,* Budapest, Hungary: CEU Press

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

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

Slote, M. (1983), Admirable Immorality, *Goods and Virtues*, Oxford: Oxford University Press,

Peterson, M. (2013), *The Dimensions of Consequentialism*, Cambridge University 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