Undercover Policing and ‘Dirty Hands’: The Case of Legal Entrapment

Abstract

Under a ‘dirty hands’ model of undercover policing, it inevitably involves situations where whatever the law-enforcement agent does is morally problematic. Christopher Nathan argues against this model. Nathan’s criticism of the model is predicated on the contention that it entails the view, which he considers objectionable, that morally wrongful acts are central to undercover policing. We address this criticism, and some other aspects of Nathan’s discussion of the ‘dirty hands’ model, specifically in relation to legal entrapment to commit a crime. Following János Kis’s work on political morality, we explain three dilemmatic versions of the ‘dirty hands’ model. We show that while two of these are inapplicable to legal entrapment, the third has better prospects. We then argue that, since the third model precludes Nathan’s criticism, a viable ‘dirty hands’ model of legal entrapment remains an open possibility. Finally, we generalize this result, showing that the case of legal entrapment is not special: the result holds good for policing practices more generally, including such routine practices as arrest, detention, and restraint.

Keywords Dirty Hands • Legal Entrapment • Moral Dilemmas • Police Ethics • Proactive Law Enforcement • Undercover Policing
1 Dirty Hands and Undercover Policing: Nathan’s Criticisms

It is often said that politics is a dirty business. More philosophically, but still somewhat imprecisely, the ‘dirty hands’ model of political power asserts that exercising it inevitably involves situations where whatever the politician does is morally problematic. One might think that this analysis extends to the exercise of state power in general. Policing (or, more broadly, law enforcement) seems like an obvious example, particularly when it deploys coercive, violent, covert, or undercover methods or practices that arouse moral unease. The police arguably use such methods to pursue valuable ends, like the prevention, detection, and reduction of crime. Policing thereby creates situations when perhaps whatever police officers do will have morally problematic aspects. The ‘dirty hands’ model might therefore be considered a fitting theoretical framework within which to analyse the morality of some central aspects of reactive and proactive policing.

Following Christopher Nathan (2017; cf. 2022: 30–32), we focus mainly upon a type of proactive policing that arouses moral unease and to which the ‘dirty hands’ model might be considered applicable, namely undercover policing. Nathan (2017: 37) introduces, but rejects, the ‘dirty hands’ model of undercover policing:

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.[…]
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.

Nathan does not examine whether the ‘dirty hands’ model applies to undercover policing in the first place. Since Nathan’s criticisms of the ‘dirty hands’ model of undercover policing hold only if the model applies, it is pertinent to examine whether the model does indeed apply.

We do this by focusing on a specific method of proactive undercover law-enforcement, namely legal entrapment to commit a crime (henceforth, ‘legal entrapment’). Our first aim, then, is to assess whether the ‘dirty hands’ model applies to legal entrapment. We show that there is a version of the model that can be applied to legal entrapment. Our second aim is to assess whether, when this account is applied to legal entrapment, Nathan’s criticisms of the model hold good. We argue that they do not. Finally, our third aim is to show that the case of legal entrapment is not special: our results extend to policing practices more generally, including such routine reactive practices as arrest, detention, and restraint.

In section 2, we provide a definition of legal entrapment based on our previous work. In section 3, we adapt to the case of legal entrapment work by János Kis on political morality and dirty hands. This enables us to set out three accounts that (as in the case of Nathan’s

---

1 Nathan (2017) argues that undercover policing does not necessarily wrong its targets because people can make themselves liable to deception and manipulation. The ‘dirty hands’ model is merely the starting point for his discussion (as a model that he thinks we should steer away from).

2 Undercover policing, as Nathan’s own paper testifies, is often seen as a crucial policing method, and it is becoming increasingly ‘normalized’ (HMIC 2014; cf. Loftus 2019 on covert surveillance). While the morality of entrapment is often discussed from other angles, it is interesting to consider the viability of applying the ‘dirty hands’ model to it, and whether and how this might interact with the more familiar arguments in the literature that make a moral case for or against legal entrapment. This is a natural development from Nathan’s depiction of how the model might apply to undercover policing in general.
interpretation of the ‘dirty hands’ model) understand the ethics of legal entrapment in terms of a kind of moral dilemma involving dirty hands. We argue that the first two accounts, adapted from Kis, are inapplicable to legal entrapment. The third, which is based on, but significantly different from, Kis’s formulation, has, we argue, better prospects of applying. It, however, leaves no room for wrongful acts, as opposed to acts that merely have bad aspects, as part of the picture: it therefore precludes Nathan’s fundamental criticism of the ‘dirty hands’ model (i.e. the criticism of it on which his other criticisms rely), namely his contention that on the ‘dirty hands’ model a wrongful act is inevitable. In section 4, we consider a possible way of reinstating Nathan’s fundamental criticism so that it appeals to badness rather than wrongfulness. Our response to this attempt enables us to contest Nathan’s other criticisms of the ‘dirty hands’ model, and to extend to other policing practices our main result, namely that there is a version of that model that both applies to legal entrapment and resists Nathan’s criticisms. Section 5 is a concluding summary.

2 Legal Entrapment to Commit a Crime: A Definition

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 (also called ‘state entrapment’) occurs when the agent is a law-enforcement officer, acting (lawfully or otherwise) in their official capacity as a law-

3 Our notion of procurement is 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.
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 (also called ‘private 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.

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

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

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
Type 1 entrapment, now called ‘legal entrapment’ for short, is the kind that is relevant to our discussion. Elsewhere (redacted) we argue that an act is one of legal entrapment if and only if it meets the following five conditions:

(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);

(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 to threaten to prosecute) the target for having performed the act.⁴

3 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 are we to understand the problem of dirty hands? To begin, it is useful to recall the following main elements of Nathan’s description of the ‘dirty hands’ model:

- Moral wrongs (i.e. impermissible actions) are committed.
- Genuine moral dilemmas are involved.

⁴ Entrapment (without the qualifier) differs in that conditions (i), (ii), and (v) are more inclusive.
- A moral residue is involved that we must accept.
- We, and undercover agents themselves, are encouraged not to dwell too much on their misdeeds.
- The overall picture is tragic, despite a reduction in crime and an increase in security, because a moral wrong is unavoidable: a ‘sacrifice’ must be made to achieve these gains.\(^5\)

Let us clarify some important terms here. We take ‘moral dilemma’ to refer to a choice situation in which the agent is confronted with moral demands and whatever course is taken is morally problematic. We do not equate ‘morally problematic’ with ‘morally wrong’: this is so as 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. We will explain the meaning of ‘moral residue’ shortly.

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 an offshoot of a tragic understanding of moral dilemmas. Following Kis (2008, Chapter 9) on dirty hands and political morality, we explain three dilemmatic accounts of the ‘dirty hands’ model.\(^6\) We make some modifications, which we explain as we go along, to Kis’s discussion, and we somewhat simplify his presentation. Our modifications are substantial in the case of the third dilemmatic account of the ‘dirty hands’ model. We consider how, if at all, each of these three dilemmatic accounts of dirty hands can be applied to the case of legal entrapment. Unlike Kis and others in the literature on dirty hands and/or moral dilemmas, we question an account only when this directly serves our primary

\(^5\) The sacrifice involved does not consist in forgoing something that is of value but in acting contrary to a value: in other words, it constitutes a kind of moral damage.

\(^6\) Of course, there are other ways of conceptualizing dirty hands: see Coady (2018) for a good summary. Focusing on Kis’s discussion, however, serves our dialectical purposes well.
purpose, to examine the applicability or inapplicability of the ‘dirty hands’ model to the case of legal entrapment.

Of the three dilemmatic accounts, let us begin with the one that, we think, most closely matches Nathan’s depiction of the ‘dirty hands’ model: the tragic account (TRAGIC). We begin each of the three subsections that follow with an indented summary of the account at issue.

3.1 The Tragic Account (TRAGIC)

The agent, $S$, is bound by two moral demands that cannot simultaneously be satisfied. Whichever demand $S$ disregards, $S$ violates a valid, in-force demand. The dilemmatic situation is inescapable, in that even if $S$ becomes involved innocently $S$ cannot come out of it innocently. However $S$ acts, $S$ will act impermissibly and incur guilt.

Recall that Nathan (2017: 37) 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 necessity claims here. The first claim is one of instrumental necessity. Nathan assumes that undercover policing methods, even if they are methods of last resort, are necessary means towards their ends (cf. Bovée 1991 on media ethics, and Coady 2008 on politics, for a similar analysis). While it is doubtful that the actual deployment of undercover policing typically meets this condition (cf. Loftus 2019 on covert surveillance), we assume, for the sake of argument, that the necessity towards such ends is a normative constraint on uses of undercover methods: that is, we assume that they are permissible only when instrumentally necessary towards their ends. The second form of necessity to which Nathan seems to allude is moral necessity or obligation: when the use of
undercover methods is instrumentally necessary, the police have a moral duty to use them. This then leads to familiar slogans that aim to bring out the paradoxical nature of the resultant 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).

Slogans, however, do not help us to understand the underlying structure of the situation. Let us, therefore, first formalize the account, closely following Kis (2008: 238). We could say that the following happens in such tragic situations. Let \( S \) be the agent (in our case, the law-enforcement officer or their deputy), let \( a \) be a course of action that involves entrapment, and let \( b \) be an option that does not. In order to generate a dilemmatic situation, an analysis must begin with the following three assumptions:

**A1:** There is a moral demand that \( S \) should perform \( a \), and there is a moral demand that \( S \) should perform \( b \).

**A2:** \( S \) can satisfy each of the two demands separately.

**A3:** \( S \) cannot satisfy both demands together.

This gives us a situation of moral conflict, but not yet a situation of moral dilemma. For that we require a further assumption:

---

7 Unless the second type of necessity is moral, we do not see how the ‘dirty hands’ model of undercover policing would involve, as Nathan takes it to involve, a moral dilemma.

8 See Coady (2008) and Curzer (2006) on Walzer (1973), as well as Alexandra (2000) on policing. Portmore (2007) would say that these two slogans describe two different kinds of moral dilemma: prohibition-dilemmas and obligation-dilemmas. We think that the three versions of the ‘dirty hands’ account that we discussed, after Kis (2008), can incorporate both kinds of moral dilemma.

9 The literature on moral dilemmas often reads as though these must involve conflicting obligations or duties. By contrast, a conclusion from the discussion in Railton (1996: 159) is that moral dilemmas ‘can arise apart from the framework of obligation or in cases in which obligation is indeterminate’. Under the influence of Railton’s discussion, Kis (2008: 238) characterizes moral dilemmas as involving conflicting oughts. We have instead characterized dilemmas in terms of conflicting moral demands: we take this to be closer to the spirit of Railton’s discussion, and a better way of attaining generality.

10 Some, such as Hare (1981), would even deny the existence of moral conflicts.
A4^{TR}: Of the two demands, neither overrides the other: both emerge undefeated.

A4^{TR} is also crucial because it partially explains the tragic (and paradoxical) nature of the situation: there is no way that the agent can do the right thing without also doing something wrong. This, however, is only a partial explanation. According to Kis (2008: 239) and others (e.g., Gowans 1994, cited by Kis), moral dilemmas are tragic largely because of the agent’s lost innocence. The idea is that dilemmas are (in a technical sense) inescapable: one finds oneself in the dilemmatic situation through 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. Given the moral demand on each side of the dilemma and that the dilemma is inescapable, TRAGIC has four tragic implications. To paraphrase Kis (2008: 239), these are:

1)^{TR}: Whichever moral demand S chooses to disregard, S violates a valid, in-force moral demand.

2)^{TR}: The dilemmatic situation is such that S may become involved in it innocently.

3)^{TR}: Once in the dilemmatic situation, S has no opportunity to come out of it innocently.

4)^{TR}: Whether S performs a or b, it will be appropriate for S to feel guilty.

On this account, moral dilemmas involve three layers of the tragic. The first is encapsulated in 1)^{TR} (and based on A4^{TR}), and the second in the conjunction of 2)^{TR} and 3)^{TR}. Once these are in place, 4)^{TR}, the third layer, falls into place.

The question that concerns us, in relation to each of the three dilemmatic accounts of the ‘dirty hands’ model, is whether it applies to the case of legal entrapment (and, if so, whether Nathan’s criticisms of that model apply specifically to the case of legal entrapment). In relation to TRAGIC, the applicability of each of the three layers of the tragic can be called into question.
We focus on the first two. The third layer might not be supported by the phenomenology of these cases, i.e., if law-enforcement agents that entrap typically do not feel guilty about what they have done.\footnote{Even if they do feel bad about what they did, there can be other explanations of this fact. Perhaps the best alternative that is particularly fitting for police work (as well as for investigative journalism, see Bovée 1991) is to emphasize the uncertainty that accompanies such work, including acts of entrapping: many officers might afterwards 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 the 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. Other explanations of bad feelings (regret and remorse, in particular) are possible, and we will appeal to them in what follows.} Of course, this does not rule out the view that it would be appropriate for them to feel guilty. Since the appropriateness of such reactive moral emotions is nevertheless determined by the moral structure of the case, the third layer of the tragic is parasitic upon the previous two layers.

With respect to the first layer, it is far from clear that $A_{4TR}$ (hence $I_{1TR}$) holds true for most cases of legal entrapment. Certainly, law-enforcement officers that entrap 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 always career criminals, into doing something criminal. Still, in the context of legal entrapment, it is far from clear that $A_{4TR}$ holds true: given all the good that might be achieved through an act of entrapment (e.g., in terms of long-term crime prevention), it might be that the moral demand not to entrap is sometimes overridden. While the literature on legal entrapment contains conflicting views about its permissibility, the literature does not suggest (at least to our knowledge) that no entrapment scenario gives rise to a clear normative conclusion about entrapping as against not entrapping. In short, even on a charitable approach, there is reason to hold that $A_{4TR}$ is not generally true of scenarios of legal entrapment. Rather than taking TRAGIC to apply across the board, it seems that a case-by-case analysis would be needed.\footnote{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.}
With respect to the second layer, the inescapability requirement is of dubious applicability to cases of legal entrapment. After all, legal entrapment involves law-enforcement agents (or their deputies) choosing to entrap their targets. Although a law-enforcement agent might be ordered by a superior to entrap a target, the choice to entrap in this situation can hardly be construed as inescapable (and certainly not on the part of the superior). In addition, both the superior and the subordinate, we can reasonably assume, made, or at least could have made, an informed choice when they became law-enforcement agents. It might be that some agents stumble into being asked to entrap in jurisdictions in which entrapment is generally discouraged; still, this does not make their choice, in any relevant sense, inescapable.\(^\text{13}\) It is reasonable to conclude, then, that the decision to entraps, at least typically, a free and informed one. By contrast, the stereotypical case of an inescapable choice, which is often used in discussing TRAGIC, is Sophie’s choice (from Styron 1979). Sophie had to choose which of her two children was to be sacrificed, and if she refused the choice both children were to be taken to the gas chamber. The choice, moreover, was imposed upon her by Dr Mengele: Sophie neither created the choice situation nor had an innocent way out of it. Sophie, thus, truly loses her innocence no matter what she does. It is doubtful, to say the least, that we can say anything remotely similar of law-enforcement officers that engage in entrapment.\(^\text{14}\)

\(^{13}\) Of course, one could say that there were inescapability of a weak kind still present in this choice: policing needs to be done—a social choice, if you will—so someone must dirty their hands. While this may be true in politics, however—we might need politicians, and some of them, on occasion at least, might have to dirty their hands—it is far from clear that we need entrapment or undercover policing in general. Besides, there is something (morally) problematic about making an individual choice inescapable on the basis of a (presumed) social choice. Thanks to (redacted) for raising this worry.

\(^{14}\) Two possible replies might be offered (thanks to redacted). Innocence comes in degrees: while Sophie fully loses her innocence, law-enforcement officers lose only 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. Some might also question the notion of inescapability: 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, and some might ask: why did she not hide from the Nazis? Why did she not try to flee? This—besides being extremely naïve—misses the point, we contend. Say that a choice situation is inescapable if one enters into it not having created it, not having intended it, and not having even foreseen it. Given this, there is a clear difference between entrapment and Sophie’s choice. The latter is inescapable, whereas the former is most likely not: it is hard to believe that those recruited into the police could not at least foresee, upon reflection, what might await them in the service. Police officers receive training, and this training involves information on the different aspects of the job. This applies even more to those working
Lastly, to turn away from the levels of the tragic, Nathan (after Walzer 1973) associates a moral residue with the ‘dirty hands’ situation. Although TRAGIC is arguably the best fit for Nathan’s depiction of the model, this particular remark ill-fits TRAGIC. If the agent has done wrong, then the agent, as per I4\textsuperscript{TR}, is guilty of wrongdoing, and this does not seem to us to constitute a moral remainder. Could the moral residue be referring to the moral phenomenology of these cases, i.e., to what the agent experiences, or, at least, to what it would be appropriate to experience? We do not consider it plausible to construe the moral residue as merely phenomenological. Instead, we think, the relevant residue should be located in the moral structure of entrapment, and not (merely) in its phenomenology. More generally, we suggest, it is the moral structure of a scenario that determines what it is apt for a person in that scenario to feel. Nevertheless, Nathan’s mention of moral residue suggests an alternative dilemmatic account that might be considered applicable to legal entrapment: the moral residue account (RESIDUE).

3.2 The Moral Residue Account (RESIDUE)

\( S \) is bound by two moral demands, to perform \( a \) and to perform \( b \), that cannot simultaneously be satisfied. The demand to perform \( a \) overrides the demand to perform \( b \), but \( b \)'s normative force does not evaporate: rather, it gives rise to a derivative requirement that the target of \( S \)'s act must receive redress.

Williams (1965) has argued that what happens in moral dilemmas is that, contrary to A4\textsuperscript{TR}, one moral demand overrides the others, but the overridden demands are not silenced. They ‘stick undercover: they receive specialist training, and are recruited from the ranks of ‘ordinary’ police officers. Besides, the very act of choosing to entrap remains escapable also on this test. This remains the case even if entrapment is a method of last resort, and even if the officer is somehow driven to do it as a matter of necessity (cf. Williams 1981 on practical necessity).
around’; their force does not evaporate. In particular, these defeated demands generate *derivative demands* to compensate for, or to repair, the damage done: this is the moral residue that TRAGIC lacks. 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 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. If agent S opts to perform *b*, then *S* must perform *c*<sub>a</sub> (where this represents compensating for, or repairing the harm caused by, having failed to perform *a*). If, on the other hand, *S* opts to perform *a*, then *S* must then perform *c*<sub>b</sub> (where this represents compensating for, or repairing the harm caused by, having failed to perform *b*). The decision is more complex because, when deciding how to act, *S* must not only decide whether to do *a* or *b*, but also which, if either, of *c*<sub>a</sub> and *c*<sub>b</sub> is a feasible option (Kis 2008: 252).

Before we address the question of application, let us this time highlight an important general problem. This is that it appears that RESIDUE’s claim to a moral residue vanishes on closer analysis (Kis 2008: 252). After all, if *S* makes the right choice by choosing to act on the overriding demand and also compensates the victims of this choice (for the failure to take the other courses of action that were initially open), then no moral residue remains. *S* simply did the right thing, on both levels (acting and then compensating): the moral universe remains intact, and *S* comes out of the situation (morally) innocent.

---

15 RESIDUE is supported naturally by a picture of competing *pro tanto* reasons the balancing of which gives us an all-things-considered ought-judgement. See Dancy (2004) for the broader meta-ethical picture (which is partially rooted in the notion of *prima facie* duties in Ross (1930)); see also Alexandra (2000), who depicts ordinary cases of ‘noble cause’ corruption in exactly this way. Note also that this meta-ethical view of reasons is a feature that RESIDUE shares with TRAGIC: both accounts presuppose that reasons compete with each other based on their strengths. DIRTY, as we explain later, seems to follow a different route.
While there is a way around this problem, the way around suggests that the proper form of RESIDUE whose application to legal entrapment we intend to query is not the one with which we started out. What is the way around the problem? As Kis (2008: 253) points out, if there is irreparable damage involved in a choice situation, then a moral residue would necessarily remain (Kis calls this ‘non-eliminable moral residue’). This appears to restore the tragic character of S’s choice-situation, since there would be no way for S fully to satisfy the requirements that apply. To formalize, the following account of the ‘moral residue’ account would hold. Let us start with the original account of moral conflict:

A1: There is a moral demand that S should perform a, and there is a moral demand that S should perform b.

A2: S can satisfy each of the two demands separately.

A3: S cannot satisfy both demands together.

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

A4MR: Performing a involves a non-eliminable moral residue, and either performing b involves a non-eliminable moral residue or the demand to perform it is not overriding.

Replacing A4TR with A4MR produces its own problems, however. First, as Kis (2008: 253) points out, if damage is irreparable, i.e., if it cannot be repaired, then, if ‘ought’ implies ‘can’,16 it is not the case that it ought to be repaired. This means that the ‘rediscovered’ tragic element in RESIDUE becomes diluted. If, on the one hand, the damage is reparable, then no residue

16 Of course, the principle is controversial: see e.g. Stocker (1990).
need remain. If, on the other hand, the damage is irreparable, then a residue remains but redress is not required (because it is impossible). In either case, no derivative moral requirement remains that could, if violated, trigger a tragic dénouement.17

Secondly, Kis (2008: 253) argues that this does not rule out the possibility that it would be appropriate for $S$ to feel bad about having failed to compensate the victim(s) of $S$’s act. This feeling should not be guilt, and perhaps not even regret or remorse. Still, $S$ can think of the act as morally problematic and feel bad about this. This gives a thinly tragic analysis: whatever $S$ does, it is appropriate for $S$ to feel bad about the chosen act. As far as the moral analysis of entrapment is concerned, however, this leaves us with very little of the tragic aspect of the situation as Nathan originally described it.

Furthermore, thirdly, while the original version of RESIDUE was generally applicable to legal entrapment, the present version applies only to those instances of it that involve irreparable damage. Arguably, however, most cases of legal entrapment are too mundane to involve irreparable damage.18 In short, what we gain in ‘tragedy’ by changing focus to irreparable damage, we lose in scope of application.19

---

17 One might deny this (thanks to redacted). For, while the doctrine that ‘ought’ implies ‘can’ rules out strict compensation of exactly the damage caused (along the lines of an eye-for-an-eye in compensation terms), it does not rule out that we ought to do as much as possible, i.e., to do what is within our power, to compensate. We are not sure this works morally, although it might explain the way in which courts operate. (Perhaps this is the logic behind court decisions mentioned in the next footnote.) It is true that moral demands can be attenuated: we are not required to save the world alone. Normally, however, this is because we are in the situation together with others: morality, in many respects, is a co-operative enterprise, so, perhaps, we are required to do only our fair share (cf. Murphy 2000), or, more demandingly, what is within our power (cf. Singer 1971). It is doubtful that these considerations apply to compensation for harm that is now in the past, is not ongoing, and was caused by one individual, or a group of individuals, rather than by society (or even humanity) as a whole.

18 This point is perhaps supported by the fact that in jurisdictions that work with the notion of entrapment, some form or other of entrapment remedy is the norm (redacted). The courts or prosecutors can presumably apply an entrapment remedy because the damage involved is reparable. It is of course possible that courts might attempt to compensate, repair, or remedy very great or ‘priceless’ sorts of damage as well, but this would be a legal, not a moral compensation: the damage would be repaired, remedied, or compensated for within the (positivistic) nature of the legal system, but not in the practical—and, consequently, moral—sense under consideration here.

19 In fact, this gets worse if we consider that, on A4\textsuperscript{108}, the choice not to entrap should be either not overriding (which is not obvious, to say the least) or should cause irreparable damage (which is certainly not generally the case).
Lastly, fourthly, there is good reason to think that RESIDUE fails to preserve the
dilemmatic nature of dirty hands. As Kis (2008: 255–256) shows, reference to irreparable
damage cannot be what constitutes moral dilemmas because hard choices of the form depicted
by A1–A3 that are not moral dilemmas can also involve irreparable damage. For example, a
rescuer might allow someone to die by deciding to save another and the choice might be
perfectly well supported by moral reasons. RESIDUE fails to distinguish such cases from
genuine moral dilemmas.

RESIDUE, we conclude, is not a viable dilemmatic account of the ‘dirty hands’ model
apt for application to the case of legal entrapment. Let us turn, then, to the third account drawn
from Kis’s discussion—the ‘dirty hands’ account (DIRTY). In this case, we significantly
modify the details.

3.3 The ‘Dirty Hands’ Account (DIRTY)

S is bound by two moral demands, to perform a and to perform b, that cannot
simultaneously be satisfied. Although the demand to perform a overrides the demand to
perform b, performing a remains morally bad. Hence, while it is right for S to choose to
perform a, what S does remains morally bad.

The above summarizes our version of this account, which differs significantly from Kis’s. We
now explain Kis’s version, and then, using it as background, flesh out the details of our version
and explain why it diverges, in some important respects, from Kis’s. In setting out his account,
we start with the usual three assumptions to depict moral conflict:
A1: There is a moral demand that $S$ should perform $a$, and there is a moral demand that $S$ should perform $b$.

A2: $S$ can satisfy each of the two demands separately.

A3: $S$ cannot satisfy both demands 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 2008: 264). Moral reprehensibility is, on Kis’s account, the central concept that features in this assumption (adapted from Kis 2008: 267):

A$^{DH}4$: Performing $a$ is morally reprehensible, and performing $b$ is either morally reprehensible or the demand to perform it is not overriding.

Kis (2008: 260–263) explains how A$^{DH}4$ is meant to help secure a dilemmatic account of the ‘dirty hands’ model. This involves five central ideas. The first concerns the aforementioned notion of moral reprehensibility. Kis holds that an act can be right, hence morally acceptable, in certain circumstances, and nonetheless morally reprehensible in the same circumstances. This is possible, says Kis, here introducing his second central idea, because some acts, such as murder and betrayal, have essential properties that make them morally reprehensible irrespective of the circumstances. Some concepts, like murder or betrayal, Kis says, have descriptive content that cannot be determined in separation from evaluative criteria: the acts they describe cannot be identified in morally neutral terms. Murder and betrayal, Kis

---

20 This idea is based on what we think is a debatable reading of Marcus (1996).
asserts, always remain morally reprehensible even if in the given situation they are, all things considered, the morally right thing to do.\textsuperscript{21}

The third idea is that \textit{threshold deontology} is an appropriate approach to normative ethics.\textsuperscript{22} At its basis, the idea is that we can evaluate an act in two ways: according to the states of affairs it produces (the viewpoint typically associated with consequentialism), and according to how it treats its object (the viewpoint typically associated with deontology). It is the latter that is crucial: if an act fails to treat its object as it should be treated then this makes it morally reprehensible. Now, action-based constraints, on the ‘threshold morality’ view, often outweigh concerns related to consequences: the putatively good consequences that the action would bring about are outweighed by the constraints relating to how the action would treat its object. Nevertheless, beyond a certain threshold (e.g. avoidance of great harm) considerations of the consequences override those that relate to the intrinsic nature of the action. Importantly, even in such cases, the concerns pertaining to the nature of the action remain in place as \textit{evaluative} considerations: it is morally right to avoid great harm, but what is, with respect to the nature of the action, inappropriate treatment, nevertheless remains inappropriate treatment.\textsuperscript{23} Hence the act, albeit morally right because justified by its consequences, remains morally reprehensible: a moral residue is left, even though we act rightly. Kis’s fourth central idea is that this makes

\textsuperscript{21} As we will soon explain in the main text, we consider Kis’s views about the essential properties of acts debatable, extraneous to DIRTY \textit{per se}, and extraneous to its defensibility. Kis (2008: 260) also includes commitments concerning actions and their individuation that we do not share.

\textsuperscript{22} Nagel (1979) is perhaps the best-known and earliest proponent of this view; see also his (1986). For discussion in a specifically political context, see Coady (2008; 2018: section 7) on what he calls ‘balanced exceptionalism’. For a general discussion, with references, see Alexander & Moore (2021: section 4).

\textsuperscript{23} It is a tricky question, and is not taken up in Kis’s presentation of DIRTY, exactly how threshold deontology sees the inter-relation of reasons relating to the consequences and reasons relating to the intrinsic nature of the action. Although Kis does not name him, the theory seems to base their relationship on a particular interpretation of the notion of exclusionary reasons from Raz (1999). Basically, beyond the threshold, reasons relating to the consequences exclude reasons relating to the intrinsic nature of the action; this is why their normative force evaporates leaving behind only evaluative considerations. Raz’s account, however, is notoriously controversial, and there are conflicting interpretations of it. See Adams (2021) for a recent discussion. This, moreover, also has a linguistic dimension. Although we follow Kis in talking about ‘overriding’ and ‘outweighing’ in the rest of this subsection, to speak strictly, this might not be correct: for Raz, exclusion is not the same as overridingness. Note also that this usage arguably fits RESIDUE better than DIRTY. Still, given that these complex metaethical matters cannot be taken up in this paper, we follow Kis’s example.
DIRTY paradoxical: whatever the agent does will be morally acceptable and unacceptable at the same time.

Finally, we come to Kis’s fifth idea: although the acts the reprehensibility of whose natures is overridden by the goodness of their consequences in dilemmatic situations are morally acceptable (because right) and unacceptable (because reprehensible) at the same time, they are not blameless and blameworthy at the same time. The justified nature of the act means that no blame is appropriate. The act is morally reprehensible, but the proper response to this is not blame, but regret or remorse. An observer’s proper responses are not resentment and indignation, but fear and pity. Kis (2008: 265) depicts the appropriate phenomenology of dirty-handed acts as lying between cases of faultless involuntary contributions to accidents (where only what Williams (1976) calls ‘agent-regret’ is appropriate) and blameworthy wrongdoing (where guilt and blame are appropriate).

While Kis’s formulation of DIRTY would, we think, apply to legal entrapment, our interest is in a version of DIRTY that expunges some commitments from Kis’s formulation. Getting rid of these commitments has, we think, two main virtues: first, it renders the amended account more plausible; secondly, it expands the number, and the variety, of morally interesting situations to which DIRTY can be applied.24

We amend Kis’s first commitment so that the notion of reprehensibility is replaced with the idea that an act can have a morally bad aspect. Thus, on our version of DIRTY, an act can be right, hence morally acceptable, in certain circumstances, but nonetheless have a morally

24 In a revised edition of the book, available only in Hungarian, Kis (2017: 290-2) discards threshold deontology in favour of a distinction borrowed from Scanlon (2008: Chapter 1). On this revised approach, moral reprehensibility concerns what Scanlon calls the ‘critical question’ of how the agent relates to the dirty-handed deed before and after it: what her attitudes ought to be about the act. Moral permissibility, by contrast, concerns what Scanlon calls the ‘deliberative question’ of whether the act should be, or should have been, done. Kis does this because he wants to avoid the paradoxical (because seemingly inconsistent) nature of his earlier position (cf. footnote 27). Our discussion is—although, we surmise, consistent with this deployment of Scanlon’s distinction—not influenced by this change to Kis’s position. Our new formulation of DIRTY, based on Kis (2008), shows that a threshold-deontological version of DIRTY is available that attains the ends that lead Kis (2017) to abandon the appeal to threshold deontology. As we show in the main text, our version of DIRTY already avoids the paradox and, as in Kis (2017), the appeal to the idea of moral reprehensibility too.
bad aspect that, in the same circumstances, does not evaporate. Our next move is to specify further the nature of such aspects. Here, too, we disagree with Kis, now concerning his second central idea: essential properties. Besides holding that serious critical questions can be asked about Kis’s understanding of essentialism, we also reckon that if a viewpoint in normative ethics can forswear a metaphysical commitment, while still doing its normative-ethical work, then it is better for it to do so. In any case, we think that it is not the idea of an act’s having essential properties that is, at bottom, relevant to deontology. Instead, it is the familiar idea that, apart from having extrinsic consequences (which they bring about), some acts have in themselves, as tokens of their types, morally-relevant intrinsic features (such as being a promise, the breaking of a promise, or a lie). An appropriate example might be that of a field surgeon that, to a patient’s great extrinsic benefit, in order to prevent a tragic outcome for the patient, and under orders, violates the patient’s autonomy by administering a treatment to which the patient has expressly not consented. This act has, we surmise, overwhelming extrinsically good consequences (in that, so to speak, it saves life or limb), while it is also intrinsically morally bad (in that it violates the patient’s autonomy).

It is thus this notion of intrinsic moral badness that we propose to use instead of Kis’s notion of essential moral reprehensibility. We are now in a position to replace Kis’s A4DH with our own proposal based on the previous three points:

---

25 A venerable approach to metaphysics regards substances as things of the right kind to have essences; acts would not count, according to this tradition, as things of the right kind. Furthermore, it is clear that Kis is working with a modal account of essence, according to which all and only those properties that an entity has ‘in all possible worlds’, or as a matter of necessity, are of its essence. That conception of essence has often been contested, especially since Fine (1994).

26 It is part of the ‘consequentializing’ agenda (e.g., of Portmore 2007) to break this distinction down by claiming that the intrinsic features, too, are states of affairs that the acts cause to obtain. Our account of DIRTY is, we suspect, incompatible with that agenda. Our interest, however, is not in backing DIRTY per se, but in showing that it does provide an applicable ‘dirty hands’ model of legal entrapment. If consequentializing and DIRTY are indeed in conflict, we here remain neutral observers of the conflict.
A$^{DH^*}$: Performing $a$ has a morally bad aspect, and performing $b$ either has a morally bad aspect or the demand to perform it is not overriding.

We next deploy this notion within Kis’s third idea: threshold deontology. The example above of the field surgeon is a case at hand: given the overwhelming positive consequences (even when balancing them against the negative consequence of pain), the threshold is reached, and administering the treatment becomes the morally right thing to do, despite the fact that doing so is intrinsically morally bad because it violates the patient’s autonomy (since the soldier has not consented to the treatment).

Furthermore, given A$^{DH^*}$, we diverge from Kis’s view that an act can be both morally acceptable (because right) and morally unacceptable (because reprehensible) at the same time.$^{27}$ We supplant Kis’s notion of moral reprehensibility, and his subsequent paradoxical deployment of the notion of moral unacceptability, with the notion that, in the dilemmatic situation, a threshold-deontological account sees one course of action as right but intrinsically bad. In the situations to which our amended version of DIRTY is meant to apply, the following situation arises. When the agent acts rightly, intrinsic moral badness is nevertheless produced. When the agent acts wrongly, an intrinsic moral good is promoted, but a more significant extrinsic ill is brought about. In short, on DIRTY, while the agent can choose to act rightly, they cannot avoid doing something morally bad at the same time. This left-over intrinsic moral badness is a moral residue that acting rightly leaves behind.

Finally, turning to Kis’s fifth central idea, from a phenomenological point of view, how should the agent then regard their action? Here we think that Kis’s phenomenological picture can be weakened somewhat. We agree that in DIRTY it can be appropriate for the agent to feel

$^{27}$ This paradox, or seeming paradox, might be thought to be an undesirable feature of Kis’s account. This perhaps raises bigger questions, which we do not propose to address, about the appropriate philosophical methodology to use when approaching a seeming paradox.
remorse at having to take the course of action in question, even though taking that course is permissible (and perhaps even mandatory) in the circumstances, and that this remorse is something more than mere agent-regret (Williams 1976), though less than guilt. On our account, however, remorse would be appropriate only in situations in which the agent’s chosen course involved a grave moral ill. Similarly, observers in such serious cases can react with fear and pity as Kis proposes. In other, less serious, situations, perhaps some kind of moral disappointment would be more appropriate (cf. Brady 2010; Menges 2020).28

For our modified version of DIRTY to apply to legal entrapment, the act of entrapment must be intrinsically morally bad in the sense explained above. Wherein lies the intrinsic moral badness of legal entrapment? Here is a possible candidate. Howard (2016: 25) argues that entrapment (whether legal or civil):

*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 […] is incompatible with respect for that agent. Specifically, it is incompatible with […] an attitude of support for the successful operation of others’ moral capacities.

While we are inclined to think that there are other morally bad aspects of legal entrapment, some of which may be agent-centred rather than target-centred,29 we agree with Howard that

---

28 Exactly how the phenomenological analysis plays out depends on where and how we draw the threshold, as well as on in what the badness of entrapment consists (when DIRTY is applied). We say more about these presently. Our point here is only that our notion of intrinsic moral badness commits us to less on the phenomenological side, and thus allows for more phenomenological space, than does Kis’s notion of essential moral reprehensibility.

29 On this distinction, and on some agent-centred objections to entrapment, see *redacted*. For details of further objections, some of which focus on features of the act of entrapment relating to the intrinsic nature of the action, see Dillof (2004), Hughes (2004), Carlon (2007), Howard (2016: 26), and *redacted*. 
such subversion is involved in legal entrapment, and that it is morally undesirable.\textsuperscript{30} While Howard sees the subversion that is involved as sufficient to make entrapment wrong, rather than just bad, within the version of DIRTY that we are entertaining it is seen merely as a bad aspect. When is this badness outweighed, rendering legal entrapment permissible? As in Kis’s discussion, this occurs when great harm is at stake. Now, there are many mundane cases where the target of legal entrapment is unlikely to have otherwise proceeded to cause great harm; legal entrapment is unlikely to be justified in such cases. On the other hand, the level of subversion inherent in a particular act of legal entrapment might also be relevant: if this is minor, then entrapment may be morally justified even for minor benefits.\textsuperscript{31}

While the extent of DIRTY’s applicability as justification for real-life cases of legal entrapment is thus unclear, this does not restrict its theoretical applicability: only those cases of entrapment are permissible where the intrinsic moral badness of the action is outweighed by the (overall) goodness of the consequences (and hence one in-force moral demand overrides another). Even in these cases, the account is non-consequentialist. This is because it remains the case that the action of legal entrapment has intrinsic negative moral value that the gains do not nullify. DIRTY not only explains why legal entrapment is sometimes permissible, but also why it is aptly regarded (as it often is by the courts and scholars) as a method of last resort. Legal entrapment is still bad, and the fitting response to having done it is, under at least some circumstances, remorse.

\textsuperscript{30} Notice that it is crucial for Howard that the target ‘will culpably choose to act wrongly’. Under the prevalent view, in US federal and state courts, of the ‘entrapment defense’, a defendant that the court finds to have been put on trial as a result of an act of legal entrapment is regarded as innocent of the offence because not legally culpable for it. If that view were right, then Howard’s objection to legal entrapment would not actually apply to cases of it. (Although Howard’s focus is on moral culpability, he is talking just about moral offences that are also legal crimes, and his argument would seem to apply equally to legal culpability.) Like some courts elsewhere (see redacted), Howard is therefore implicitly rejecting the specific entrapment doctrine that has become orthodox in the US. We also reject that doctrine. Moreover, in saying that entrapment subverts the target’s moral capacities short of nullifying culpability, Howard would be able to avail himself of the distinction (explained in redacted) between causing the target to commit an act (e.g. by the administration of a mind-altering drug) and procuring the act (as we there understand it).

\textsuperscript{31} This connects to the discussion between those that advocate for a fixed threshold and those that argue for a sliding-scale form of threshold deontology. See Alexander & Moore (2021: section 4).
No doubt, many theoretical questions can be asked about DIRTY, but, as before, we have tried to stay clear of the general debate on the morality and phenomenology of moral dilemmas and of dirty hands. Our interest has been primarily in the application of DIRTY to legal entrapment. Unlike TRAGIC and RESIDUE, DIRTY appears to be a good candidate for this. It has no place for the idea that whatever the agent in a ‘dirty hands’ scenario does will be morally wrong. It is therefore immune to Nathan’s fundamental criticism of the ‘dirty hands’ model. At the same time, it retains not only the dilemmatic nature of the choice situation, but some element of the tragic (when the intrinsic moral badness qualifies as a grave moral ill), and a moral remainder as well.

3.4 Summary

Our substantive discussion took as its starting point Nathan’s criticisms of the ‘dirty hands’ model of undercover policing (Nathan 2017: 37). Nathan makes four assertions about what would happen if the ‘dirty hands’ model of undercover policing were to be publicly endorsed, which we now enumerate:

1) The public would correctly feel that morally wrongful acts were at the centre of police practice.

2) Despite the justification of these acts, public unease would remain.

3) The police would become, because of internalization of this ethic, even more secretive.

4) By contrast, if there were a better model for understanding the morality of undercover policing that would yield the belief on the part of its practitioners (the police) that it were not wrongful, that would be more conducive to publication justification.

32 For relevant critical discussion, with references, see Coady (2018: section 7) and Alexander & Moore (2021: sections 4–5).
2) and 3) are social-scientific predictions. 1) is too, but it incorporates a purely theoretical assertion that we have been subjecting to philosophical scrutiny: namely, that the ‘dirty hands’ model takes wrongful acts to be at the centre of police practice. This is Nathan’s fundamental criticism of the model.

We have highlighted Nathan’s characterization of the ‘dirty hands’ model (specifically, of undercover policing) as morally dilemmatic, tragic, and involving a moral residue. We have examined whether there was a way to flesh out Nathan’s remarks into a fully-fledged account that was applicable, in particular, to legal entrapment. In doing so, we have deployed three accounts of the ‘dirty hands’ model of politics that we have adapted from Kis (2008). We have argued that the first two, TRAGIC and RESIDUE, are inapplicable to the case of legal entrapment. We have proposed a significantly modified version of the third account, DIRTY, and we have argued that (regardless of its other merits or demerits) it is applicable to legal entrapment. Crucially, however, DIRTY (whether in Kis’s original form or in our modified version) rejects the idea that whatever the agent in a ‘dirty hands’ scenario does will be morally wrong. Nathan’s case against the ‘dirty hands’ model of undercover policing was predicated upon the crucial contention that, under that model, undercover policing did involve moral wrongdoing. Of the three accounts here surveyed, then, only DIRTY remains generally applicable to legal entrapment. DIRTY precludes Nathan’s fundamental criticism because it sees dirty-handed acts as not wrongful but merely as morally bad.

4 Policing, Dirty Hands and Public Justification

At this point, a possible rejoinder is that Nathan’s fundamental criticism can be revised by replacing the appeal to moral wrongness (from his characterization of the ‘dirty hands’ model)
with an appeal (from our version, amending that of Kis, of DIRTY) to what is intrinsically *morally bad*. Accordingly, the ‘dirty hands’ model would be characterized via the following list (with changes italicized):

- Acts with *morally bad aspects* are committed.
- Genuine moral dilemmas are involved.
- A moral residue is involved that we must accept.
- We, and undercover agents, are encouraged not to dwell too much on the *morally bad aspects* of their acts.
- The overall picture is tragic, despite a reduction in crime and an increase in security, since a *moral evil is unavoidable*: a ‘sacrifice’ must be made to achieve these gains.

The amended criticism could then be summarized as follows, with only 1) and 4) having to be changed):

1*) The public would correctly feel that *acts with morally bad aspects* were at the centre of police practice.

2) Despite the justification of these acts, public unease would remain.

3) The police would become, because of the internalization of this ethic, even more secretive.

4*) By contrast, if there were a better model for understanding the morality of undercover policing that would yield the belief on the part of its practitioners (the police) that it *did not involve acts with morally bad aspects*, that would be more conducive to publication justification.
We think, however, that neither 1*) nor 4*) is a serious problem, and, partly for the reasons that support this suggestion, neither 2) nor 3) is a serious objection.

Let us start with 1*). The first thing to notice here is that the ‘dirty hands’ model, under our modified version of DIRTY, applies, at least in principle, not only to the comparatively exotic, and certainly specialized, domain of undercover or covert proactive policing, but also to any forceful or coercive elements of policing that are necessary to the practice of law-enforcement. Among the many such forceful and coercive elements of policing, consider such routine reactive practices as arrest, detention, and restraint. Although they may be for the greater good of their targets, and/or of society, these practices, given their forceful or coercive natures, can justifiably be resorted to only when the probability is relatively low that their ends can effectively and efficiently be achieved by non-forceful and non-coercive means. This is because when an act is forceful or coercive, that is an intrinsically morally bad aspect of it.

Consequently, unlike in the case of the original version of Nathan’s criticism, which was predicated on the idea that morally wrong acts are central in the ‘dirty hands’ model, the contention that acts with morally bad aspects are at the centre of police practice is neither specific to the case of undercover policing nor, we suggest, particularly contentious. Indeed, the observation that even everyday police practice involves morally bad aspects is innocuous, realistic, and compatible with conscientious, but morally careful, policing. To deny it would seem naïve. Moreover, the contention that policing involves acts with morally bad aspects is not something that is distinctive of the ‘dirty hands’ model: it arguably follows from any realistic picture of the ethics of everyday policing. In short, our modified version of DIRTY provides an account of the ‘dirty hands’ model that applies not only to legal entrapment, but to every method or practice of policing, whether undercover or not, that inevitably involves
intrinsically bad aspects. These considerations suggest that 1*) is true but innocuous: it is ineffectual as a criticism of the ‘dirty hands’ model of undercover policing.

Turn now to 4*). The relationship between dirty-handed acts and public justification has two sides. On the one hand, there is the side, emphasized by Nathan, of the agent and the institution of which they are a part. The agent gets dirty hands; thereby, so does their institution. The resulting demand for public justification puts a heavy burden on the agent’s institution (in this case, the police). Now, if wrongful acts are crucial to undercover policing, then of course this makes public justification of it difficult (albeit, perhaps, not necessarily impossible). Nathan is right about this. On our version of DIRTY, however, the relevant acts are not morally wrong, but only morally bad. This makes public justification easier, and perhaps much easier, than Nathan envisages. In fact, given the above-noted wide scope and generality of DIRTY, it is difficult to see what one could propose as a realistic and non-naïve contrasting alternative.

It is, however, the other side of the relationship that is even more important for the evaluation of 4*): the side of the public to whom justification is owed. If policing did not centrally involve acts with morally bad—or, for that matter, morally wrong—aspects, then it would not, as against more typical forms of labour, be in special need of moral and political (as against economic) public justification. That is, in an important sense, 4*) gets things backwards, and Nathan’s objection can be inverted. He emphasizes the consequences of embracing the ‘dirty hands’ ethic for public morale, as well as for police morale. The very ethic he targets, however, has resources in it to control these detrimental effects, and these resources centre exactly on the notion that Nathan finds detrimentally affected: public justification. That is,

---

33 This is even more so if one endorses some form of philosophical anarchism. For then states, and thus their police force, are illegitimate exactly because whatever they do violates our autonomy (Wolff 1970), or because we have not consented to them (Simmons 2001).
exactly because the acts involved are dirty-handed, public accountability, in institutional as well as in non-institutional form, is placed centre stage in the ‘dirty hands’ model.

It is not difficult to understand why this is so. If an agent commits a morally bad—let alone morally wrong—act, then the agent owes, at a minimum, an explanation-cum-justification to the targets (or victims) of their act. Now, from this it does not automatically follow that this must take a public form. Still, there are good reasons to think that it should do so (cf. Kis 2008: Chapter 8). First, there is the problem of moral corruption. On the one hand, in policing we need—if we go along with the idea that there are morally justified dirty-handed acts—people that are willing to dirty their hands. On the other hand, we do not want these people to dirty their hands too easily. Lord Acton’s apothegm ‘power tends to corrupt, and absolute power corrupts absolutely’ (Acton 1887) is also true in policing (cf. Alexandra 2000 and Miller 2016.). This is further underlined by a second reason: uncertainty. No police officer can be sure that when they entrap, for example, their reasons are indeed good. Hence, just as with moral corruption, we do not want to make it too easy for police officers to act on their reasons, however good they take them to be. The two considerations also connect: those that are more easily inclined to dirty their hands are also more likely not to care about the fact that they might be acting wrongfully. In short, public justification, for good reasons, is an essential part of a ‘dirty hands’ model, and not an external constraint or demand on it.

Two further conclusions concern how the modified version of DIRTY undermines 2) and 3). In relation to 2), to the extent that the public is at ease with the idea that policing is

---

34 Typical elements of a liberal (constitutional) democracy can all be mentioned here: freedom of the press, freedom of speech, independent courts, and so on. The media, of course, are particularly important. Moreover, 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.

35 In the history of political thought, public justification has not always been the norm. The prince of Machiavelli (1998) has no inner life, and certainly is not a model for public justification. The politician of Weber (1994) suffers internally only. Although the ‘Catholic’ model of Walzer (1973) proposes social expression, this is primarily for the repentance of sins in order to achieve salvation. There is also, of course, the radical utilitarian tradition that holds that the true morality must not be publicly affirmed, but must be kept secret (Lazari-Radek & Singer 2010).
necessary at all, the public will be, or at least ought rationally to be, at ease with the idea that the police will engage in forceful or coercive acts: for policing is, as the term, both broader (in that, for example, it includes the activities of the intelligence services) and narrower (in that it excludes, for example, educational and outreach work undertaken by the police), ‘law-enforcement’ suggests, an inherently forceful and coercive endeavour. Contrary to 3), the modified DIRTY account makes increased secrecy by the police less likely than would a contrary account that denied that morally bad acts were an inevitable aspect of policing. For, given that such acts are inevitable aspects of policing, and that the acknowledgement of this is crucial to the need for public (moral and political) justification of the activities of police forces, the rational and strategically appropriate attitude to those acts, on the part of the police, is to admit their presence and to justify them. Secrecy, on the contrary, would be both irrational and counter-productive in respect of public justification.

5 Conclusion

We have focused on the ‘dirty hands’ model as a possible framework for analysing the morality of undercover policing, specifically in relation to legal entrapment. We took as our starting point Nathan’s (2017) criticisms of the model (when applied to undercover policing). We had two main aims. Our primary aim was to see if the model applied at all to legal entrapment; our secondary aim was to establish whether, if the model applied, Nathan’s criticisms of it held. With respect to the first aim, we presented three possible versions of the model, loosely taking our inspiration from Nathan’s remarks and making extensive use of Kis (2008). We found that the first two accounts, TRAGIC and RESIDUE, were inapplicable to the case of legal entrapment. We argued that the third, DIRTY, applied, but had no room for morally wrong acts as part of the scenario. When Nathan’s fundamental criticism was amended, so as to deploy the
notion of the intrinsically morally bad in place of is the notion of the morally wrong, we found that DIRTY was immune to all four of Nathan’s criticisms. In short, once we consider the range of resources available to the ‘dirty hands’ model, Nathan’s critique of that model is either precluded or loses its force.

Acknowledgements […]

References


