**Philosophical Methodology and Conceptions of Evil Action**

Andrew J. Latham, Kristie Miller and James Norton

**Abstract**

There is considerable philosophical dispute about what it takes for an action to be evil. The methodological assumption underlying this dispute is that there is a single, shared folk conception of evil action deployed amongst culturally similar people. Empirical research we undertook suggests that this assumption is false. There exist, amongst the folk, numerous conceptions of evil action. Hence, we argue, philosophical research is most profitably spent in two endeavours. First, in determining which (if any) conception of evil action we have prudential or moral (or both) reason to deploy, and second, in determining whether we could feasibly come to adopt that conception as the single shared conception given our psychological make-up and the content of the conceptions currently deployed.

**Keywords:** Evil; Folk Concepts; Conceptual Engineering; Philosophical Methodology; Conceptual Analysis.

1. **Introduction**

Theorists of evil disagree, along a number of dimensions, about what it takes for an action to be evil.[[1]](#footnote-1) They disagree about the content of the folk *concept* of evil.

In what follows we will call each of the competing ways of spelling out the concept of evil, competing *conceptions* of evil. Here, we depart somewhat from the usual way in which the concept/conception distinction in drawn. Typically, conceptions are held to be determinate specifications of an indeterminate concept.[[2]](#footnote-2) Thus a single concept can have a number of competing conceptions, each of which make that concept determinate in different ways. One way of thinking of contested concepts such as the concept of evil action, is in terms of a single shared concept and a number of competing conceptions.[[3]](#footnote-3) We wish to preserve the idea that competing conceptions are competing conceptions of some shared notion; that’s why they’re *competing* conceptions. So we suppose that when theorists spell out different accounts of evil action, they are offering different (and competing) *conceptions* of evil action. Given the way debates then proceed, we think the right way to understand disputes amongst these theorists is as descriptive disputes about which of these *conceptions* is the one (or closest to the one) deployed by the folk. This is in sharp contrast to a *normative* project of evaluating these competing conceptions to see which, if any, it would be normatively—*prudentially* or *morally*—good for us to deploy: a project known as conceptual ethics.[[4]](#footnote-4)

The assumption underlying these descriptive disputes is that there is a single, shared, *conception* of evil deployed by sufficiently culturally similar philosophers and folk alike. We will say that an individual *possesses* a concept or conception if they are able to use that concept or conception: if the concept/conception is able to be a constituent of their thoughts. By contrast, an individual *deploys* a concept/conception if the way she thinks about the world is structured through the use of that concept/conception.[[5]](#footnote-5) The aim is to determine which conception this is. If there is a shared conception of evil, then it follows that there is a single *concept* of evil. Hence disputants should also be seen as making the claim that there is a single shared (determinate) concept.

This paper addresses the question of whether there is a single shared *conception* of evil action, leaving open whether, if there is not, there might still be a single shared indeterminate concept. It is in this regard that we depart from the usual way of thinking of the connection between concepts and conceptions, which entails that there can be competing conceptions only if there is some shared indeterminate concept which these conceptions make determinate. We make no such assumption. For all we say, there might be *nothing* that is shared amongst all disputants, which is made determinate in different ways by different conceptions. Instead, it might be that these are all conceptions of evil action because of some shared *role* these conceptions play, rather than because there is some overlapping content shared between all of them, which is the shared indeterminate concept of evil action. In fact, the data we collect suggests that there probably is no such shared content, and hence no shared indeterminate concept. Our principal aim, however, is to investigate whether there is any shared *conception*, not whether there is *anything at all* that is shared, which could constitute a shared indeterminate concept.

We think the descriptive project is important, not least because theorists disagree, along a number of dimensions, about which is the correct conception. If we want to avoid engaging in merely verbal disputes about which actions are evil, then we need to know which, of the various candidate conceptions articulated by theorists, is closest to the conception currently deployed. Moreover, in order to adjudicate the feasibility of bringing a community to deploy some alternative conception—such as one recommended by conceptual ethicists—we need to know how those recommended conceptions are similar to, and different from, the conception we do deploy.

This paper presents empirical evidence (§2) we gathered about the content of the folk conception of evil action. On the basis of this evidence we argue (§3) that there is no single, philosophically interesting, answer to the question of which of the competing conceptions is descriptively the best. That does not make the various competing conceptions uninteresting: far from it. We suggest (§4) that for each of these conceptions—and other engineered conceptions besides—we should not ask whether the conception articulated is one that *is* deployed (or is closest to one or other that is deployed) but rather, whether it is a conception we *ought* to deploy. While there is not scope to answer that question here, we articulate a number of possible social roles one might have reason to want a conception of evil action to play, and say something about which of those roles the various theoreticians’ conceptions are well placed to play. We then briefly consider the feasibility of bringing the broader community to deploy certain conceptions of evil action given the content of the conceptions of evil action that are actually deployed.

1. **Investigating conceptions of evil action**

Theorists of evil disagree, along a number of dimensions, about what it takes for an action to be evil. They disagree over whether wrongfulness is a necessary feature of evil action (Card 2002; Garrard 1999; Formosa 2008; Russell 2007) or not (Calder 2013). While some (Russell 2014; Arendt 1951) insist that evil actions must be *extreme* wrongs, others allow that evil actions can be relatively minor wrongs (Garrard 1998; Kekes 1988:13). Further, while some take blameworthiness and extreme wrongness to be jointly sufficient and necessary for an action to be evil (Russell 2014) others think that evil actions must be qualitatively distinct from ordinary wrongdoings by having some additional feature such as the perpetrator gaining pleasure in doing wrong (Steiner 2002) or the perpetrator having particular malicious motivations (Calder 2013; Kekes 2005; Thomas 1993; Steiner 2002). Others disagree, arguing that an action can be evil regardless of which psychological states it issues from (Card 2010). There is even disagreement about whether, necessarily, an action that is evil must be harmful. Some think it must (Card 2002; Kekes 2005), while others disagree (Calder 2002:56; Garrard 2002:327).

Some theorists hold that actions are evil only if there exist supernatural entities or forces, since evil actions are, necessarily, those that issue from such forces (Clendinnen 1999:79–113; Cole 2006). Others disagree (Card 2010:10–17; Russell 2007; Haybron 2002b; Kekes 2005; Arendt 1951; Garrard 2002:325).

In what follows we call these various attempts to spell out the content of the folk conception (and hence shared folk *concept*) of evil action *theoreticians’ conceptions* (leaving open whether some, or all, map onto the folk conception(s). We reserve the expression *engineered conceptions* for those conceptions that have, as necessary or sufficient conditions, those conditions proposed by theoreticians, but not in a combination any actual theoretician has proposed. We leave it open whether some, or all, of these conceptions map onto the folk conception(s).

In order to investigate the content of the folk conception of evil action we focussed on six conditions that theorists have taken to be severally, or jointly, necessary or sufficient for an action to satisfy the folk conception of evil action. Importantly, while competing conceptions of evil action are not always explicitly framed in terms of necessary and sufficient conditions, the dialectic between theorists usually involves trading intuitions regarding whether, for some condition C, C is necessary, sufficient, or neither, for some action to count as evil. It does not follow that participants in these debates think any of these conceptions can be easily and exhaustively spelled out in terms of necessary and sufficient conditions. But it does reveal that they take conceptions to have associated with them *some* sufficient or necessary condition or conditions..[[6]](#footnote-6)

*Demonic powers* (dP): The action issues from an agent who has been given supernatural powers by some supernatural force.

*Demonic control* (dC): The action issues from an agent who has been given supernatural powers by some supernatural force, and, in addition, the supernatural force is controlling the agent’s actions.

We distinguish demonic powers from control. It is clear that for Cole (2006) there must be some kind of supernatural influence for an action to count as evil. He thinks that an agent who has such powers “does not act on her own initiative” (2006:6), and so it might seem most natural to say that he thinks that dC is necessary for an action to be evil. However, he also talks of actions that issue as a result of a “force that choses to work through” (2006:6) an agent, which is perhaps closer to dP. Given that there is some ambiguity here between whether the possession of supernatural powers by an agent is necessary, or said agent being controlled by the supernatural force that provided the powers is necessary, we tested both conditions.

*Morally wrong* (mW): The action in question was morally wrong.

*Extreme wrong* (xW): The action in question was extremely morally wrong.

*Moral blameworthiness* (mB): The action in question is one for which the agent in question is morally blameworthy.[[7]](#footnote-7)

*Motivated by malice* (mV): The agent was motivated by malice, to perform the action in question.

To test which, if any, of these conditions or combinations thereof, is necessary, or sufficient, for an action to satisfy the folk conception of evil action, we presented 112 fluent English participants over the age of 18 and residing in the US with a range of vignettes and asked them to report whether a scenario was one in which the action was evil. [[8]](#footnote-8) [[9]](#footnote-9) Hence they are asked to *deploy* their conception. Non-extreme moral wrong scenarios involved stealing a hat, while extreme moral wrong scenarios involved murdering the entire population of a village (for an example of each see Table 1, below).

Table 1. *An example of a non-extreme moral wrong and an extreme moral wrong scenario. In the scenarios the perpetrator, Phoebe, also satisfies the conditions of being morally blameworthy, and motivated by malice.*

|  |
| --- |
| Non-Extreme Moral Wrong Evil Scenario:  Jane and Phoebe know one another, but are not close friends. Jane has a hat that Phoebe does not really like, but she knows that Jane is very fond of the hat. One day, Jane puts the hat down and Phoebe picks it up. Jane asks if anyone has seen her hat, and Phoebe responds that she has not. Phoebe takes the hat home and is pleased to discover that, as she expected, Jane is very upset at its loss. |
| Extreme Moral Wrong Evil Scenario:  Jane and Phoebe know one another, but are not close friends. Jane grew up in a little village that she really likes. Jane and Phoebe work at a bio-tech company and both have access to some deadly pathogens. One day Jane puts down a deadly pathogen, and Phoebe picks it up. Phoebe picks up the pathogen, because she knows that if she releases it in the village where Jane grew up, she will kill all its inhabitants and this thought gives her great pleasure. Phoebe takes the pathogen to the village and releases it, where it spreads throughout the village, killing everyone who lives there. |

The scenarios were generated in a systematic manner by combining a pre-worded exemplar of each six conditions in a ‘present’ or ‘absent’ form. With the exception of the two supernatural conditions, each condition was mentioned—as present or absent—in each scenario. For example, using the scenarios presented in Table 1, Phoebe was not morally blameworthy when she simply forgets she picks up Jane’s hat, or as the result of a genuine accident, drops the deadly pathogen. Phoebe was not motivated by malice when she is not pleased to have made Jane upset, or took no pleasure in the thought of killing the village’s inhabitants. When the condition of demonic control was present the scenario included the claim that ‘Phoebe becomes possessed by a demonic force. As a result of this, she is able to control the weather, the movement of objects, and take over people’s minds and make them do what she wants’. In the demonic powers condition, the scenario also included the claim that ‘Phoebe controls her demonic powers’. To view in full all the scenarios used in this study please refer to our project page on the Open Science Framework https://osf.io/s98re/.

Here we were interested in participants’ forced choice survey responses to the question: ‘Does Phoebe perform an evil action?’. To test the proportion of participants who judged a secular condition (or combination of secular conditions) to be *sufficient* for an action to count as evil, we looked to whichever of our secular scenarios (i.e. those who did not include either dP or dC) *including* that condition (or combination of conditions) had the *fewest* participants judge that Phoebe’s action is evil. That’s because, if a participant judges a condition (or combination of conditions) to be a sufficient condition for an action to be evil, they will always judge Phoebe’s action to be evil when that condition (or combination of conditions) is present. The idea is that if there is a scenario including the condition under investigation where only 10% of participants judge Phoebe’s action to be evil, no more than 10% of participants judge that condition to be sufficient. Thus the figures we report are to be read as the maximum value; in fact the value could be anything less than, and up to, 10%, so we express these percentages as ≤10%.

To ascertain what proportion of participants held that a secular condition (or combination of secular conditions) was *necessary*, we looked to whichever of the secular scenarios *lacking* that condition or combination of conditions had the *highest* proportion of participants judge that Phoebe had performed an evil action. Once again, we report these results as maximum values.[[10]](#footnote-10)

While all possible combinations of the four secular conditions were tested, table 2 reports results for secular theoreticians’ conceptions. Table 3 reports results for non-secular theoreticians’ conceptions. It is clear that these theorists take it to be *necessary* for evil action that the supernatural play some role in the generation of said actions. Hence the first half of table 3 reports participants’ judgements regarding the necessity of either dP or the combination of dC & dP (as dC entails dP).

It is, however, less clear what other conditions these theorists take to be necessary, and which they take to be jointly sufficient. To test the impact of dP, or dC & dP on judgements of sufficiency, we compared participants’ responses to two of the combinations of conditions in table 2—those which drew the highest agreement amongst participants regarding their sufficiency, since these seemed to be plausible candidates to be sufficient alongside the supernatural conditions—with their responses to those same conditions, plus the presence of either dP or dC & dP (see Table 3).[[11]](#footnote-11)

Subsequently, where appropriate, we report other combinations—those that represent engineered conceptions.

Table 2. *Summary of results for planned analyses of conditions and condition combinations in contemporary secular literature for sufficiency and necessity.*

|  |  |  |
| --- | --- | --- |
| **Factors** | **Sufficiency**  (yes≤) | **Necessity**  (yes ≤)[[12]](#footnote-12) |
| mW[[13]](#footnote-13) (*n*=94) | ≤5.3% | ≤10.8% |
| mB[[14]](#footnote-14) (*n*=93) | ≤63.8% | ≤25.8% |
| xW (& mW)[[15]](#footnote-15) (*n*=94) | ≤22.3% | ≤10.8% |
| mV[[16]](#footnote-16) (*n*=93) | ≤52.1% | ≤15.1% |
| mW & mB & xW[[17]](#footnote-17) (*n*=93) | ≤84.9% | ≤10.8% |
| mW & mB & xW & mV[[18]](#footnote-18) (*n*=93) | ≤96.8% | ≤10.8% |

To test the proportion of participants who judged a non-secular condition combination of conditions to be *sufficient* for an action to count as evil, we looked to whichever of our scenarios *including* that condition (or combination of conditions) had the *fewest* participants judge that Phoebe’s action is evil.

Table 3. *Summary of results for planned analyses of the sufficiency of non-secular condition combinations and the necessity of non-secular conditions.*

|  |  |
| --- | --- |
| **Factors**  dP[[19]](#footnote-19)  dC (& dP)[[20]](#footnote-20) | **Necessity**  (yes≤)  ≤3.2%  ≤3.2% |
|  | **Sufficiency**  (yes≤) |
| mW & mB & xW + [dP] (*n=*93) | ≤82.8% |
| mW & mB & xW + [dP & dC] (*n=*93) | ≤81.7% |
| mW & mB & xW & mV + [dP] (*n=*94) | ≤91.4% |
| mW & mB & xW & mV + [dP & dC] [[21]](#footnote-21) (*n=*94) | ≤88.2% |

Notably, we find that very low percentages of participants held that dP or dC & dP are necessary for an action to count as evil. Then when we compare condition combinations (mW & mB & xW) and (mW & mB & xW & mV) to the same conditions plus the demonic powers (i.e. (mW & mB & xW & dP) and (mW & mB & xW & mV & dP)) we find the result is non-significant (*p*>.125). This suggests that the presence of the dP condition makes no difference to participants’ judgements regarding whether the scenario is one in which there is an evil action present.

The same non-significant results were found when we tested the mW & mB & xW combination of secular factors with, and without dC & dP (*p*=.581). By contrast, the test involving mW & mB & xW & mV and dC & dP was significant (*p*=.021), but not in the direction one might have predicted. Instead, we found that participants were *more* likely to judge that an evil action had occurred when other conditions were present in the *absence* of the demonic control condition.

1. **The folk conceptions of evil action**

We took participants’ surveyed responses to vignettes to be defeasible evidence regarding the content of the conceptions they deploy (after all, we ask them to *use* the conception, and hence this is evidence of deployment, not possession). It’s defeasible evidence since participants’ actual responses may not perfectly track appropriately post-reflection responses.

We take the data to suggest three interesting things about the folk conceptions. Henceforth, we talk of the folk conceptions of evil action since the data strongly suggests that there is more than one folk conception deployed.

First, participants were no more likely to judge an action to be evil when the supernatural conditions were present, than when they were absent. This strongly suggests that there is *at least* one secular conception of evil action in the population, and is evidence that the presence of non-secular conditions in addition to any secular conditions does not make the scenario a better deserver to satisfy *any* folk conception of evil action.[[22]](#footnote-22) Moreover, given the extremely low percentage of participants who found dP or dP & dC to be necessary for an action to count as evil, this suggests a tiny percentage of the population deploy a non-secular conception of evil action.

The second notable result is the very low percentage of subjects who agree that any condition, or combination of conditions, is necessary for an action to count as evil. We noted previously that our study is intended to adjudicate the issue of whether there is a single shared conception of evil action in the population, rather than whether, if there are multiple conceptions, there might nevertheless be a single shared indeterminate *concept*. We think that the very low percentages of subjects who agree as to these necessary conditions is evidence against the thesis that there is a single, shared, indeterminate concept of which the various conceptions are determinate extensions. For were there such a shared concept we would expect there to be at least some set of necessary conditions, shared amongst all the competing conceptions. That is not what we find.

The highest result we get is ≤25.8% of participants holding that moral blameworthiness is necessary for an action to count as evil. Beyond that, almost all the other conditions and sets thereof hover around ≤11% of participants supposing them to be necessary. This shows there to be remarkably little convergence among the folk regarding which conditions, if *any*, are necessary for an action to count as evil. Obviously, this finding also tells against supposing there is any single shared *conception,* since if there is no shared *indeterminate* *concept*, there cannot be any shared *(determinate) conception*, which runs counter to the assumption made by theorists.

The third notable result is that there is substantial agreement (≤96.8%) that the combination of *all* of the secular conditions (mW & mB & xW & mV)—call this the *maximal* set of conditions—is sufficient for an action to count as evil. Call the conception according to which the maximal set of conditions is jointly sufficient for an action to count as evil, the *paradigmatic conception of evil action*. Then we see a very high degree convergence amongst the folk regarding this paradigmatic conception.

Nevertheless, it is also the case that significant percentages of participants hold that something less than this maximal set of conditions is sufficient for their conception of evil action to be satisfied.[[23]](#footnote-23) Since by definition conceptions are determinate (ruling out

that that there is a single shared conception that is exceedingly vague around the edges so that although participants agree about paradigmatic instances of evil, they disagree about non-paradigmatic instances) this evidence indicates that there are a number of overlapping, but distinct, folk conceptions of evil action being deployed by different participants, and that these distinct conceptions have associated with them different necessary and sufficient conditions.

On the assumption that there are multiple folk conceptions of evil action deployed, our data suggests that some of the theoreticians’ conceptions pick out folk conceptions that are deployed by *higher* percentages of the folk, than do competing theoreticians’ conceptions. We don’t wish to underplay that discovery: we take it to be genuinely interesting. Nevertheless, we argue that our results, alongside very wide disagreement amongst theorists of evil, strongly suggests that there is no single, best, philosophically interesting, answer to the question: what is the descriptive content of the folk conception of evil action? And that is because the question rests on the false methodological assumption that there is a single, shared, folk conception.

So far we have shown that thee are multiple conceptions present in the population. But that is not enough to show that one of these ought not be preferred (descriptively speaking) to any of the others. For perhaps, for instance one conception is considerably more prevalent in the population than others. In what follows we argue that there are no non ad hoc grounds to chose one of the competing conceptions as being *the* folk conception.

We call the first argument to this conclusion the *gradual decline argument*, and the second the *pair-wise comparison* *argument*. The gradual decline argument is as follows: as we decrease the number of conditions present in a scenario, in general the percentage of participants who think the action satisfies their conception of evil action declines. While this decline is not absolutely linear, it is such that there is no non *ad hoc* place to draw the line between some combinations of conditions that could count as articulating the sufficient conditions of *the* conception of evil action, and all the others that do not (see table 4).

The pair-wise comparison argument proceeds by noting that for any combination of conditions, there is *some* other combination such that the percentage of participants who take the former to be sufficient for an action to be evil will be very similar to the percentage of participants who take the latter to be sufficient for an action to be evil. Indeed, in almost all cases there is no statistically significant difference between the sufficiency results for the pair-wise compared combinations (see table 4). Thus, for any such pair there is no non *ad hoc* reason to judge that one, instead of the other, spells out the sufficient conditions associated with one correct conception of evil action.

Table 4. *Comparison of theoreticians’ and engineered conceptions of evil action. Percentages are for sufficiency.*

|  |  |  |
| --- | --- | --- |
| **Factors** | **Theoretician’s Conception**  **(Closest)** | **Engineered Conception**  **(Closest)** |
| mB  (≤63.8%) | mV  (≤52.1%) | mB & mV  (≤63.8%) |
| mV  (≤52.1%) | mB  (≤63.8%) | mW & mV  (≤52.1%) |
| mW & mB & xW  (≤84.9%) | mW & mB & xW & mV (≤96.8%)\* | mW & mB & mV  (≤75.3%) |
| mW & mB & xW & mV (≤96.8%) | mW & mB & xW  (≤84.9%)\* | mW & mB & mV  (≤75.3%)\* |

N.B. \*Judgements differ significantly different between paired conceptions (McNemar’s test: *p≤*.001). [[24]](#footnote-24)

Consider, first, the graduate decline argument. Looking at tables 2 and 4 combined, we see a drop from ≤96.8% of participants assessing the maximal set of conditions to be jointly sufficient for an action to count as evil, to between ≤75.3% (mW & xW & mV) and ≤84.9% (mW & mB & xW) of participants holding that various combinations of three conditions are sufficient for an action to count as evil. When we consider pairs of conditions (each of which corresponds to an engineered concept) we see another drop in the percentage of participants who take the combination to be sufficient for an action to count as evil: all combinations of pairs of conditions cluster between ≤52.1% (mW & mV) and ≤63.8% (mB & mV) of participants taking those conditions to be jointly sufficient for an action to count as evil.[[25]](#footnote-25) Finally, we see a further drop-off when we consider single conditions, where we find between ≤5.3% (mW)[[26]](#footnote-26) and 63.8% (mB) of participants take those conditions alone to be sufficient for an action to count as evil.[[27]](#footnote-27)

Substantial percentages of the population take something less than the maximal set of conditions to be sufficient for their conception of evil action to be satisfied. More importantly, there is no precipitous drop-off between the maximal set of conditions and the various sets of three conditions, or between sets of three conditions and sets of two conditions, in such a way that we could draw a principled, non *ad hoc* line around a *particular* set of conditions and suppose that this set best specifies the sufficiency conditions for *the* folk conception of evil action.

Now consider the pair-wise comparison argument. To see that there are such pairs look to table 4, which lists each of the theoreticians’ conceptions, on the left, and shows both the closest theoreticians’ conception and the closest engineered conception. The closest conceptions are those for which the *most* similar percentage of participants suppose that combination of factors to be sufficient for their conception to be satisfied.

Are the gradual decline and pair-wise comparison arguments good? There are two kinds of objection one might offer at this point. According to one objection, the lack of convergence on a *single* set of conditions amongst folk in our study is irrelevant, since what we care about is convergence after appropriate reflection and equilibration. If after such equilibration there is convergence on a single set of conditions, then this gives us the non *ad hoc* reason we need to accept that set of conditions as spelling out the content of the single shared folk conception. According to the second kind of objection, we don’t need there to be any such convergence on a single set of conditions, since there can be grounds other than such convergence to prefer one conception to another as being the right descriptive account of the folk conception. Let’s consider each of these in turn.

If theorists are attempting to articulate the content of an *appropriately post-reflection* conception then perhaps we cannot draw *any* conclusions about the descriptive adequacy of these theories from data about the way folk deploy their conception absent such reflection and equilibration. This response succeeds only if we have reason to think that after due reflection the folk would converge on one and *only* one set of conditions. That, however, seems very unlikely given the very wide disagreement between theorists who *have* engaged in careful equilibration and reflection. Since if anything, theorists disagree to an even greater extent than do the folk, there are no grounds for such optimism.

Perhaps, then, there are reasons to prefer one conception to another that do not stem from facts about the percentages of folk that deploy different conceptions. Perhaps, given the sorts of social roles the various conceptions of evil action *in fact* play, we have reason to prefer one over the others. For instance, maybe one conception plays more of these social roles than do the others, and on these grounds we prefer that conception.

We think this is by far the most promising way to resist our arguments. Importantly, though, to show that this one would need to begin by articulating the various social roles the conceptions of evil action play. That requires empirical research rather than *a priori* probing of theoreticians’ conceptions. What the folk are inclined to say regarding whether a scenario counts as one in which there is an evil action will partly constrain, and inform, what social role those conceptions are in fact playing. Moreover, it’s not clear whether this really counts as a descriptive project any longer, or a project in conceptual ethics: after all, what descriptive fact would make it better that a conception plays more, rather than fewer, of these social roles? Perhaps there are normative reasons to prefer that conception, but there are not obviously descriptive reasons.

To be clear, then, we don’t take ourselves to have offered unassailable reasons to think that further *a priori* engagement in the descriptive project is not worthwhile; but we do take ourselves to have provided reasons to focus on providing and evaluating *moral* or *prudential,* rather than descriptive, reasons why we ought to deploy one, rather than another conception.. While we cannot hope to decisively identify those reasons here, in what follows we aim to do something much more modest. We articulate a few candidate social roles that are *prima facie* desirable roles for a conception of evil action to play. We then say something about which of the conceptions so far articulated would be plausible candidates to play those roles both given the nature of the role, the content of the conception, *and* empirical considerations regarding the ways in which the population currently deploys the various conceptions. All of this leaves it open that the conceptual ethicist might recommend that we eliminate the concept of evil action altogether because there are no socially beneficial roles for any conception to play, or, at least, no socially beneficial roles that it is feasible that we could bring any conception to play.

1. **Conceptually engineering conceptions of evil action**

Recall that engineering conceptions is the process of designing new conceptions. Conceptual ethics is the project of normatively evaluating these conceptions. To that end, we begin by articulating six different social roles that one might want a conception of evil action to play.

*The Explanatory Role*: explaining why certain kinds of actions occur.

*The Predictive Role:* enabling the prediction of certain kinds of actions.

*The Preventative Role:* enabling the prevention of certain kinds of actions.

*The Retributive Role:* legitimating punishments that are either different in kind, or quantity, for certain actions.

*The Therapeutic Role:* facilitating therapeutic relief as a result of victims coming to understand, gain insight into, or otherwise explain, certain kinds of actions.

*The Expressive Role:* providing vocabulary for those wronged by certain kinds of actions to express to both themselves and to others, the way in which they have been wronged.

These social roles are connected in important ways. A conception that plays the explanatory role might also play the predictive and preventative roles. That will not always be so: there are explanations that would not afford any predictive or preventative power. But at least some kinds of explanations do, and a conception that plays all three roles would, other things being equal, be preferable to one that plays only the explanatory role. Bearing that in mind, let’s call those three roles, jointly, the EPP roles.

It is also clear that the remaining three roles are connected. The expressive and therapeutic roles are connected, insofar as a conception that plays the expressive role can be expected to go some way to towards playing the therapeutic role. There are also potential connections between the retributive and therapeutic roles. It could be that having something that plays the retributive role is part of what it takes to have something to play the therapeutic role. However, given empirical evidence of this matter we will be careful not to suppose this is so.[[28]](#footnote-28) Ideally, however, one might hope that a conception that plays the therapeutic role will also play the expressive role, and so we call these roles the TE roles.

**4.1 The EPP roles**

We think that none of the theoreticians’ conceptions are well placed to play the EPP roles. There seems little reason to think that there is anything *explanatorily* unified about the set of actions that fall under any of the theoreticians’ secular conceptions beyond, perhaps, their moral significance. Although there is something explanatorily *unified* about a non-secular conception of evil action—i.e. that it is the product of demonic possession/control—that conception is explanatory only insofar as there are indeed demonic possessions. Moreover, it is usefully predictive only if (i) there are demonic possessions, and (ii) it is predictable who will be possessed, and what possessed individuals will do. Since we think neither of these is the case, we think this conception is not well placed to play any of the EPP roles.

Nevertheless, this leaves open that we might engineer an explanatory conception of evil action. There are two options available. If evil actions are actions of a certain kind which issue from evil persons, and if we can explain the behaviour of evil persons in virtue of something they have in common, we might be able to both predict and prevent such behaviour. Hence we might engineer a *psychologically medicalised* conception of evil person: a conception which categorises people as evil just in case they share some kind of underlying psychological mechanisms which generate certain suites of behaviour. Such mechanisms would provide good explanations for such behaviour as well as, potentially, the power both to predict and prevent such behaviour. If evil actions simply are those actions that issue from those mechanisms then we can explain and predict evil action.

A second option would be to hold that evil actions are actions with certain features (such as moral wrongness etc.) that are the result of the presence of certain social forces. Then we would explain evil actions as the product of certain social mechanisms.

We could call any conception of this kind a *socially medicalised* conception. Such an explanation would also afford some predictive and preventative power, at least on the assumption that we can detect the relevant social mechanisms and have some opportunity to intervene upon them. We take Arendt’s (1951) view of evil to be one example of something approaching a socially medicalised account of evil action, though we take it that a more nuanced account of the relevant social mechanisms would be necessary in spelling out a socially medicalised conception that is best able to play the EPP roles. Again, though, it is worth noting that any socially medicalised conception of evil action would be significantly different from the conceptions that the folk in fact deploy, since the folk are inclined to call actions evil in the absence of *any* particular social forces being at work. It is, therefore, an open question whether the engineered conception could feasibly be introduced into the population to play the desired role.

**4.2 The Retributive role**

Let’s now consider the retributive role. Could the psychologically or socially medicalised conceptions of evil action play the retributive role? Yes. At first blush one might worry that the sorts of psychological or social mechanisms to which these conceptions appeal will be precisely the sorts of mechanisms that undermine the blameworthiness of the person whose actions are in question, and hence which undermine any grounds for retribution against that person. That need not be so. One can understand that behaviour issues from a certain psychological or social mechanism without concluding that the agent is blameless for that behaviour. Nevertheless, it is certainly true that these candidate conceptions are more closely tied to therapeutic intervention rather than retribution.

The sorts of conceptions that better lend themselves to playing the retributive role are those that include, amongst their necessary conditions, the moral blameworthiness of the agent, alongside the moral wrongness of the action. Any such conception can license retribution, at least insofar as retribution is ever morally licensed. It is worth noting, however, that a high percentage of our participants deployed a conception of evil action (74.2%) even where the individual was not morally blameworthy (the action was an accident). This suggests that a conception that plays the retributive role (if having such a role played is desirable) would be a significant departure from the conceptions deployed amongst significant proportions of the folk.

**4.3 The TE roles**

That brings us to the therapeutic and expressive roles. Could the psychologically or socially medicalised conceptions of evil action play the TE roles? Yes. Either could provide the sort of insight that victims can use to come to process and understand why they were the victim of certain actions, and to provide a vocabulary to express both their feelings towards, and their understanding of, their situation. But these conceptions are not the only ones that could play the TE roles. Let’s consider some other candidates.

Consider an engineered conception of evil action according to which it is necessary and sufficient for an action’s being evil that it is an extreme wrong. While this conception might play the expressive role, it’s not clear to what extent it plays the therapeutic role: it’s not clear to what extent merely understanding that one has been the victim of an extreme moral wrong aids in one understanding and coming to term with that wrong. It is also notable that this conception is a significant departure from the folk conceptions of evil action. Only ≤22.3% of our participants held that an action’s being an extreme wrong was sufficient for it to count as evil, and still fewer (≤10.8%) think it is necessary. This raises the question of whether it would be feasible to bring the population to deploy that conception.

So consider an engineered conception according to which being the product of malice is necessary and sufficient for an action to count as evil. This conception of evil action arguably fails to play either the expressive or the therapeutic roles. For some actions will count as evil whether they are extreme wrongs or not, so long as they issue from malice. Hence this conception elides the distinction between victims of extreme wrongs and victims of minor wrongs—insofar as both turn out to be victims of evil action—and this would seem to deny the former certain expressive power. Similarly, it is not clear that any real therapeutic benefit would be gained from deploying such a conception.

Of the various candidate conceptions, those that are most plausibly able to play the TE roles are those that combine at least the conditions of extreme moral wrongness, malice, and possibly also blameworthiness. It is noteworthy that our participants were very sensitive to the motivations of agents when choosing whether to describe their actions as evil. 89.2% of participants held that an action (can) count as evil even when it is not morally wrong.This suggests that any attempt to engineer a conception to play the TE roles would need to be sensitive to the fact that large percentages of the population currently deploy a conception of evil action in a much broader context than would be desirable if one wants that conception to play the TE roles.

In sum, it appears that entirely new conceptions might need to be engineered if we want a conception of evil action to play as many of those social roles as possible. It is plausible that some kind of psychologically or socially medicalised conception might do best here, since they can play both the EPP and the TE roles, though there is an open question how well they play the retributive role. More work needs to be done in establishing whether or not newly engineered conceptions could come to be deployed by the expression ‘evil action’ given the way folk actually deploy the conceptions associated with that expression. Or perhaps, since few of the social roles considered above can be played by either the theoreticians’ or the folk’s conceptions of evil action, we have reason to eliminate the concept of evil action—and associated expressions—from our conceptual repertoire altogether. But these are questions for another day.

1. **Conclusion**

We have used new empirical research into the content of the folk conceptions of evil action to argue that there is no single, shared folk conception of evil action, and, further, that it is likely that there is also no shared, indeterminate, folk concept of evil action. Nevertheless, we think, there are important questions about the conception of evil action that theoreticians have articulated: questions about which of these conceptions, if any, we have prudential or moral reason to deploy, and, about whether we have prudential or moral reason to deploy some yet to be engineered conception. There are also questions about whether any of the conceptions which we have prudential or moral reason to deploy are ones that can feasibly be deployed using the term ‘evil action’, given the empirical realities: given the content of the conceptions the folk deploy using this expression. This paper has not attempted to determine which conception we have most reason to deploy, nor whether it is feasible to so do. Instead, we hope to have articulated some considerations—both normative and empirical—that are relevant to answering these questions, and to have provided some groundwork for a more thorough investigation of these issues.

**Mailing Address:**

School of Philosophy

University of Sydney

Quadrangle A14

Sydney NSW 2006

Australia

**Email Address:**

[kristie\_miller@yahoo.com](mailto:kristie_miller@yahoo.com)

**References**

Arendt, H., 1951 [1985], *The Origins of Totalitarianism*, San Diego: A Harvest Book, Harcourt, Inc.

Barry, P.B., 2009, “Moral Saints, Moral Monsters, and the Mirror Thesis,” *American Philosophical Quarterly*, 46 (2): 163–176.

Burgess, A. & Plunkett, D., 2013a, ’Conceptual Ethics I’ *Philosophy Compass* 8(12) 1091–1101.

–––, 2013b, ‘Conceptual Ethics II’ *Philosophy Compass* 8(12) 1102–1110.

Calder, T., 2002, “Towards a Theory of Evil: A Critique of Laurence Thomas's Theory of Evil Acts,” in *Earth's Abominations: Philosophical Studies of Evil*, D. M. Haybron (ed.), New York: Rodopi, pp. 51–61.

–––, 2003, “The Apparent Banality of Evil,” *Journal of Social Philosophy*, 34 (3): 364–376.

–––, 2009, “The Prevalence of Evil,” in *Evil, Political Violence and Forgiveness,* A. Veltman and K. Norlock (eds.), Lanham, Maryland: Lexington Books, pp. 13–34.

–––, 2013, “Is Evil Just Very Wrong?” *Philosophical Studies,* 163 (1): 177–196.

Cappelen, H., (forthcoming) *Fixing Language: An Essay on Conceptual Engineering* Oxford University Press

Card, C., 2002, *The Atrocity Paradigm: A Theory of Evil*, Oxford: Oxford University Press.

–––, 2010, *Confronting Evils: Terrorism, Torture, Genocide.* Cambridge: Cambridge University Press.

Carlsmith, K. M., Wilson, T. D., & Gilbert, D. T. (2008). “The Paradoxical Consequences of Revenge.” *Journal of personality and social psychology*, *95*(6), 1316.

Clendinnen, I., 1999, *Reading the Holocaust*, Cambridge: Cambridge University Press.

Cole, P., 2006, *The Myth of Evil: Demonizing the Enemy*, Westport, Connecticut: Praeger.

Dworkin, R, 1978, *Taking Rights Seriously: New Impression with a Reply to Critics*. Oxford: Duckworth.

Floridi, L., 2011, “A Defence of Constructionism: Philosophy as Conceptual Engineering.” *Metaphilosophy* 42 (3):282-304.

Formosa, P., 2008, “A Conception of Evil,” *Journal of Value Inquiry*, 42 (2): 217–239.

Gallie, W.B., 1956, “Essentially Contested Concepts,” *Proceedings of the Aristotelian Society*, 56:167-198.

Garrard, E., 1998, “The Nature of Evil,” *Philosophical Explorations: An International Journal for the Philosophy of Mind and* Action, 1 (1): 43–60.

–––, 2002, “Evil as an Explanatory Concept,” *The Monist*, 85 (2): 320–336.

Graham, G., 2003, *Evil and Christian Ethics*. Cambridge: Cambridge University Press.

Haybron, D.M., 1999, “Evil Characters,” *American Philosophical Quarterly*, 36 (2): 131–148.

–––, 2002a, “Consistency of Character and the Character of Evil,” in *Earth's Abominations: Philosophical Studies of Evil,* D.M. Haybron (ed.), New York: Rodopi, pp. 63–78.

–––, 2002b, “Moral Monsters and Saints,” *The Monist,* 85 (2): 260–284.

Kekes, J., 1988, “Understanding Evil,” *American Philosophical Quarterly*, 25 (1): 13–23.

–––, 1990, *Facing Evil,* Princeton: Princeton University Press.

–––, 2005, *The Roots of Evil,* Ithaca: Cornell University Press.

Lukes, S., 1974, *Power: A Radical View*. London, Macmillan.

McGinn, C., 1997, *Ethics, Evil, and Fiction,* Oxford: Clarendon Press.

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

Russell, L., 2006, “Evil-Revivalism Versus Evil-Skepticism,” *Journal of Value Inquiry*, 40: 89–105.

–––, 2007, “Is Evil Action Qualitatively Distinct From Ordinary Wrongdoing?” *Australasian Journal of Philosophy*, 85 (4): 659–677.

–––, 2010, “Dispositional Accounts of Evil Personhood,” *Philosophical Studies*, 149 (2): 231–250.

–––, 2014, *Evil: A Philosophical Investigation*. Oxford: Oxford University Press.

1. Philosophical theorising focuses on the ‘narrow’ or ‘moral’ concept of evil (as opposed to a broader concept that picks out anything sufficiently bad, whether morally bad or not. [↑](#footnote-ref-1)
2. For uses of the concept/conception distinction see Rawls (1971), Dworkin, (1978) and Lukes (1974). [↑](#footnote-ref-2)
3. As first described by Gallie (1956: 167-198.) [↑](#footnote-ref-3)
4. We reserve ‘conceptual engineering’ for the process of developing new concepts, or conceptions and ‘conceptual ethics’ for normatively evaluating the deployment of both new, and engineered, concepts and conceptions. See Cappelen (forthcoming), Floridi (2011) Burgess and Plunkett (2013a, 2013b). [↑](#footnote-ref-4)
5. Deploying a concept/conception entails possessing that concept/conception. Possessing a concept/conception does not entail deploying it. For instance, all of the authors of this paper possess the concept *ginger* (a derogatory concept that applies to those with red hair), insofar as each of us can tell you roughly what the content of that concept is. But we do not deploy it, insofar as none of us uses the concept in our thought or speech, nor do we conceptually carve up the world in terms of those who are ginger, and those who are not. [↑](#footnote-ref-5)
6. For example, while Garrard (1998:fn. 2) claims that “[t]hose who are attracted to a particularist form of moral realism will not be surprised at the failure of any given natural property to provide necessary and sufficient conditions for the presence of a moral property such as evil”, she later notes that “Objections might be raised here to the search for necessary and sufficient conditions for a phenomenon such as evil, on the grounds that such a search implies an unwarranted essentialism, or alternatively that for other phenomena such searches have not had a successful track record. I cannot address this methodological issue adequately here, but would like to note that philosophical practice, in the face of a proposed account of a phenomenon, usually includes consideration of whether there are uncontentious cases of the phenomenon which do not fit the account, and on the other hand, whether there are cases which do fit the account but aren't examples of the phenomenon in question. It's hard to see how this differs from considering whether the account provides necessary and sufficient conditions for the phenomenon in question.” (2002:fn. 15). [↑](#footnote-ref-6)
7. Notice that on some moral theories not only can there be wrong actions for which an agent is not morally blameworthy, but there can also be actions for which the agent is morally blameworthy, but which are not wrong. [↑](#footnote-ref-7)
8. Participants were recruited through advertisement on Amazon’s Mechanical Turk. Data from 18 participants was excluded for failing to correctly answer an attentional check question. This gave us a final sample size of 94. There were a statistically balanced number of males and females, as well as those who possessed a religious belief or not. [↑](#footnote-ref-8)
9. Ethical approval for this study was obtained from the University of Sydney Human Research Ethics committee. Consent was obtained from participants on return of a completed survey. All participants were naïve to the study’s hypotheses. [↑](#footnote-ref-9)
10. One might worry that if participants would never respond that an action is evil under any of the conditions tested, then all of the conditions will come out as necessary given this methodology. In fact, however, only one participant failed to find *any* of the conditions presented to be ones in which there was an evil action. [↑](#footnote-ref-10)
11. The influence of demonic powers and demonic control on people’s forced choice responses was assessed using McNemar’s test. A significant test statistic indicates that the presence of demonic powers or control had an impact on the distribution people’s judgments. [↑](#footnote-ref-11)
12. Figures report the percentage of participants *no greater than which* took the condition to be necessary. Some of the single case conditions have the same percentage reported as those for condition combinations including that condition. These results are consistent with its being the case that the percentage of participants who in fact take the single condition to be necessary is higher than the percentage of participants who in fact take the combination to be necessary. All we are offering is an upper bound. [↑](#footnote-ref-12)
13. Russell (2014), Kekes (1988), and Garrard (1998) defend the necessity of mW. [↑](#footnote-ref-13)
14. Russell (2014), Kekes (1988), Garrard (1998) and Arendt (1951) defend the necessity of mB. [↑](#footnote-ref-14)
15. Russell (2014) and Arendt (1951) defend the necessity of xW. As xW entails mW, the scenario we used to test the sufficiency of xW was naturally one in which the action was also mW—the combination xW & ~mW is contradictory. [↑](#footnote-ref-15)
16. Kekes (1988) defends the necessity of mV. Calder (2013) requires that an agent must intend her victim to suffer significant harm in order for her action to be evil, but it is not clear whether this requirement is captured by mV. [↑](#footnote-ref-16)
17. Russell (2014) defends the sufficiency and necessity of this combination. [↑](#footnote-ref-17)
18. Russell (2014) and Kekes (1988) deem this combination sufficient, though for each it bundles in more than is required. [↑](#footnote-ref-18)
19. Clendinnen (1999:79-113) suggests that the presence of supernatural forces or creatures is necessary. It is unclear whether this is a defence of the necessity of dP or dC. Likewise, Graham (2003) quoting St Paul in Ephesians, seems to think that the presence of supernatural forces is necessary for the presence of evil action (2003:154). Again, it is unclear whether he has in mind dP or dC. Finally, Cole (2006:5-6) thinks that appealing to evil can only be explanatory “if we suppose there is some other force at work other than the woman in question, either some kind of force that chooses to work through her, or some kind of narrative force, a story unfolding in which she is simply a character playing a specific and prescribed role”. Here, he seems to be suggesting that either dP or dC might be necessary for this concept of evil action (i.e. one he calls mythological) to be satisfied. [↑](#footnote-ref-19)
20. Recall that dC entails dP. [↑](#footnote-ref-20)
21. We think this comes closest to Cole’s (2006) mythological conception of the sufficiency conditions for evil action. [↑](#footnote-ref-21)
22. We did not test *every* secular theoretician’s conception plus dP or dC & dP. But since we want to know whether dP or dC & dP are some necessary part of a set of jointly sufficient conditions, it seems most reasonable to test their conjunction alongside those other non-secular conditions which drew the highest levels of agreement amongst participants that the actions were evil. [↑](#footnote-ref-22)
23. Only one participant judged that no non-maximal combination was sufficient for an action to be evil—and that participant also judged that the maximal combination of conditions was not sufficient. [↑](#footnote-ref-23)
24. While the difference here is statistically significant, it’s still pretty close: it’s sufficiently close that it’s not clear that there is a good basis to choose between the two concepts. [↑](#footnote-ref-24)
25. With the exception of the combination of mW & xW—which is really not a combination at all, since xW entails mW—where that drops to only ≤22.3%. [↑](#footnote-ref-25)
26. ≤3.2% if we include the supernatural condition of dP. [↑](#footnote-ref-26)
27. We don’t mean to suggest that the decline is gradual in that for any single condition, a lower percentage of participants will find that condition sufficient, than they will any pair of conditions (and likewise for any trio of conditions). There are stand-out cases (such as mV) where a single condition attracts high convergence amongst participants (≤52.1%), higher than do some pairs of conditions. Nevertheless, there is generally a (relatively) gradual decline as the number of conditions present declines. [↑](#footnote-ref-27)
28. Evidence suggests that victims in general feel worse when they seek retribution, than when they do not (which is of course not the same as feeling worse when a perpetrator is punished, than when they are not). See Carlsmith et al. (2008). [↑](#footnote-ref-28)