Philosophers' Imprint

VOLUME 14, NO. 11
MAY 2014

# Moral Coercion

Saba Bazargan

*University of California, San Diego*



## 1. Introduction[1]

In this paper I will analyze a type of conduct exemplified by the following two cases:

> HOSTAGE
>
> A villain credibly threatens to kill fifty innocent hostages she has taken unless you kill the villain's enemy — an innocent against whom the hostage-taker holds an irrational grudge. If you do not kill the villain's innocent enemy, she will remain unharmed by the villain, but the hostages will die. If you *do* kill the villain's innocent enemy, then the hostages will be released, unharmed.

The villain has put you in a situation where you have a moral reason to commit a pro tanto wrong — namely, killing an innocent. The villain does this by leveling a credible conditional threat: if you do not accede to her demand that you kill one innocent, she will commit a morally worse harm. Compare this case to the following:

> SHIELD
>
> A villain wishes to kill an innocent enemy of hers. She knows that if she tries to do so, you will shoot her. So she grabs three children and uses them as human shields; the only way for you to stop the villain from killing the innocent is by shooting through the children. If you do not shoot the villain, her innocent enemy will die, but the three children will be allowed to go free. If you shoot the villain, her innocent enemy will remain unharmed by the villain.

1. I thank Craig Agule and Sam Rickless for invaluable criticisms of an earlier draft. I also received helpful feedback in presenting a much-abbreviated version of this paper at the annual meeting of the Society for Applied Philosophy in 2013.

In this case, the villain has put you in a situation where you have a moral reason to allow her to commit a wrong — namely, killing an innocent.

Though in both examples you face a dilemma, there are differences in how the dilemma is imposed. In HOSTAGE, the villain is imposing a conditional threat. In SHIELD, the villain is not (even implicitly) imposing a conditional threat. Rather, the villain intentionally makes it pragmatically impossible for you to stop her without making things worse.

Despite these differences, there is a common element that allows a unified moral analysis of these examples. In such cases a wrongdoer (C1) intentionally denies an agent (C2) the option of preventing both of two distinct sets of harms ($\phi$ and $\psi$) from befalling others; C1 does this in order to provide C2 with an incentive to commit or allow the lesser of the two harms ($\phi$), thereby achieving C1's goal. I call this "moral coercion".

It is clear moral coercion is worthy of consideration in its own right, considering its prevalence, especially in war. Hostage-taking, certain forms of terrorism, and the use of human shields are just a few examples of morally coercive tactics commonly used in warfare. Yet moral coercion remains under-theorized in normative ethics. I will develop an analysis of moral coercion by addressing two sets of questions that will help resolve how we ought to respond to moral coercion:

1. *The Liability Question*: To what extent, if at all, is C2 morally liable for the harms she fosters when she responds to moral coercion by committing or enabling the lesser evil? More specifically, by acceding to C1's wishes and thereby occasioning harm to third-party innocents, is C2 morally liable to defensive and compensatory harms? I will argue that C2 can indeed be morally liable for the harms resulting from doing as C1 wishes, even if C2 is morally obligated to accede. The fact that C2 is wrongfully coerced does not itself mitigate her responsibility for what she does, and is thus not itself a basis for diminished liability.

2. *The Badness Question*: Intuitively, acceding to moral coercion allows "evil to succeed". How do we make sense of this intuition? I will claim that we can make sense of this intuition by arguing that C1's malign intentions are manifest in C2's actions when the latter accedes to C1's wishes, thereby affecting the weight that the resultant harms ought to receive in our deliberations.

I will begin by developing a prescriptive account of moral coercion. I will then discuss when and why moral coercion is wrongful, before turning to the Liability and Badness Questions. But before beginning, I will make several preliminary points:

i. I will assume that there are three ways in which an agent *A* can foster some event *e*. She can commit *e*, she can enable *e*, or she can allow *e*. I will leave 'commit' unanalyzed, except to say that if *A* commits *e*, then there is no other agent who, subsequent to *A*'s act, causally contributes to *e* in a way necessary or sufficient for *e*'s occurrence. If *A* enables *e*, then she provides another agent with the means to commit *e*, who does so. If *A* allows *e*, she has the power to prevent another agent from committing *e* but refrains from exercising this power. These are not full-fledged analyses of 'commit', 'enable', and 'allow'; they are instead simplifying assumptions sufficient for the analysis of moral coercion I develop here.

ii. I will assume that the intention/foresight distinction is morally relevant in that, all things being equal, it is morally worse to commit a harm intentionally than it is to do so collaterally (*i.e.*, foreseeably but non-intentionally). I also assume that the commission/omission distinction is

morally relevant in that, all things being equal, it is morally worse to commit a harm than it is to allow it to occur through inaction.

iii. I will address the Liability and Badness Questions in broadly consequentialist language. This is not because I think that some version of consequentialism provides the ultimate grounds for what is morally right and wrong, but rather because consequentialism has expressive power that makes speaking in its terms especially convenient; any moral features of actions canonically emphasized in deontological accounts of morality (such as the intrinsic value of the act committed, the relevance of the commission/omission and the intention/foresight distinctions, the intrinsic value of rights, agent-centered restrictions and permissions, etc.) are *expressible* in consequentialist terms, even if they cannot be *grounded* in a consequentialist theory.[2] Accordingly, in determining whether one ought to accede to moral coercion, I will say that we ought to do a "proportionality calculation" in which we weigh the moral benefits against the moral costs, where features morally relevant from the personal standpoint are included in this proportionality calculation.

## 2. Varieties of Moral Coercion

Here I explain in greater detail what moral coercion is. In doing so, I distinguish various types of moral coercion. The resulting taxonomy is not comprehensive. I will limit myself to drawing those distinctions that will ultimately reveal how moral coercion functions at the most fundamental level, which helps answer the Liability and Badness Questions. I will not consider "non-central" types of moral coercion, exemplified at the end of this section.

2. For a defense of this view, see (Portmore, 2007). For more on the possibility of "consequentializing" non-consequentialist moral theories, see especially (Dreier, 1993) and (Louise, 2004), but also (Schroeder, 2007).



C1 commits an act of moral coercion against C2 if and only if the following conditions hold: C1 intentionally puts C2 in a position in which C2 must choose exclusively between one of two harms, $\phi$ and $\psi$, where to "choose" a harm is to commit, enable, or allow that harm. If C2 chooses $\phi$, a harm will befall a third party, P1. If C2 chooses $\psi$, a harm will befall a distinct third party, P2. From an impartial standpoint, $\psi$ is morally worse than $\phi$. C1 forecloses the conjunctive option of $\sim\phi$ and $\sim\psi$, in order to motivate C2 into choosing the lesser harm, viz., $\phi$, which C1 knows C2 will have a (perceived) moral reason to do.

A difference between HOSTAGE and SHIELD is that the former is an example of "active" moral coercion whereas the latter is an example of "passive" moral coercion. In cases of active moral coercion, C1 puts C2 in a position where she must choose between $\phi$ and $\psi$, by threatening to commit $\psi$ unless C2 chooses $\phi$. This is what happens in HOSTAGE: C1 credibly threatens to commit $\psi$—*i. e.*, to kill the hostages—unless C2 commits $\phi$—*i. e.*, kills C1's innocent enemy. In cases of *passive* moral coercion, C1 puts C2 in a position where she must choose between $\phi$ and $\psi$—but C1 does this not by leveling a conditional threat, but by making it pragmatically impossible for C2 to choose neither. This is what happens in SHIELD: C1 intentionally puts C2 in a position in which the only way to prevent C1 from killing an innocent is for C2 to kill the innocents that C1 is using as a human shield. Here is another example of passive moral coercion:

> ALLEY
>
> Thirty children, separated from their class during a field trip, find themselves in an alley, at a dead end. C2, who has also accidentally turned into the alley, is several meters from the children. And several meters from C2 is an innocent whom C1 has been attempting to kill for some time. C1 is on top of the buildings, above the alley, and can see everyone below. She has a bomb, but does not have a clear shot at her enemy. However, she reasonably guesses that if she throws it toward the children, C2 will

> run toward it, grab the bomb, and throw it in the only direction available to save the children: towards the villain's innocent enemy. This will achieve the villain's end of killing her target.

As in SHIELD, C1 is making no demand of C2 — not even implicitly. She does not conditionally threaten to kill the children as a means of motivating C2 into killing C1's enemy. And C1's actions, subsequent to throwing the bomb, do not conditionally depend on what C2 chooses to do. Instead, C1 ensures that $\psi$ will occur unless C2 chooses $\phi$; and C1 does this as a means of motivating C2 into committing the act that achieves C1's aims.

Part of what characterizes all cases of moral coercion, both active and passive, is that C1 intentionally denies to C2 the conjunctive option of choosing both $\sim\psi$ and $\sim\phi$ — C2 can only choose one. The only difference between active and passive moral coercion is *how* C1 forecloses the relevant conjunctive option. The difference between active and passive moral coercion is a difference in C1's tactics.

We can further categorize moral coercion by distinguishing cases in which C1 aims to manipulate C2 into *committing an act* from those in which C1 aims to manipulate C2 into *refraining from preventing an act*. These cases differ with respect to who commits the lesser harm, *i.e.*, $\phi$. Likewise, we can distinguish between cases that differ with respect to whether it is C1 or C2 who would commit the greater harm, *i.e.*, $\psi$, should C2 refuse to comply with C1's wishes. When C1 commits a harm, it is *coercer-enacted*. When C2 commits a harm, it is *coercee-enacted*.

For example, in ALLEY, if C2 saves the lives of the children by intercepting and throwing the bomb, *she* will be the one who does the killing — that is, she will be the one who commits $\phi$. If C2 chooses *not* to intercept and throw the bomb, she will have refrained from preventing C1 from committing $\psi$. Accordingly, ALLEY is an example of moral coercion in which, if $\phi$ occurs, it is coercee-enacted, and if $\psi$ occurs, it is coercer-enacted; mutatis mutandis for HOSTAGE. Compare those with the following case of moral coercion:

> HOSTAGE 2
>
> C1 is attempting to rob a bank. She takes a dozen innocent hostages and credibly threatens to kill all of them should C2 interfere with her plans to rob the bank.

In this example, the lives of the innocents are being used to morally coerce C2 into *permitting* a wrongful act. Since C1 is coercing C2 into refraining from preventing a harmful act (rather than into committing a harmful act), it is an example in which $\phi$ is coercer-enacted rather than coercee-enacted. C1 motivates the desired conduct — the omission — by threatening to make things go morally worse if C2 refuses to comply. This is in contrast to HOSTAGE, in which C2 is coerced into doing C1's dirty work by actually killing C1's innocent enemy. But whether the conduct incentivized is a commission or an omission makes no difference to whether moral coercion has occurred.

Compare HOSTAGE 2 with SHIELD. Because in SHIELD C1 is coercing C2 into refraining from preventing the murders of her innocent enemy, $\phi$ is coercer-enacted, as in HOSTAGE 2. But unlike in HOSTAGE 2, $\psi$ would be coercee-enacted in SHIELD should $\psi$ occur; preventing the murder of the innocent enemy requires actually *killing* the human shield, rather than merely refraining from preventing their deaths. Also, unlike HOSTAGE 2, SHIELD is an example of passive coercion: C1 denies C2 the conjunctive option of preventing the deaths of the innocent's enemy *without* killing the human shield — and C1 does this without conditionalizing her conduct to C2's response.[3]

3. Note that coercee-enacted deterrents are not limited to passive moral coercion. The following example demonstrates this:

   > BOMB
   > C2, a bomb-maker, has negligently lost a bomb. C1 knows its location. She proves that it is in a populated area, and that when it goes off, many will die. She claims that she will not reveal the location of the bomb to C2 unless C2 kills C1's enemy first.

   Here the lesser harm, $\phi$, is the death of C1's innocent enemy. The greater harm, $\psi$, is the death of the many innocents resulting from the bomb's

Differentiating types of moral coercion along these two dimensions — active vs. passive and coercer-enacted vs. coercee-enacted — allows us to appreciate precisely how the use of involuntary human shields and the use of hostages compare: the former is passive and the latter is active, and the deterring act in the former is coercee-enacted whereas that of the latter is coercer-enacted. Though both count as instances of moral coercion in that C1 is foreclosing the possibility of both $\sim\phi$ and $\sim\psi$ as a means of morally motivating C2 into choosing $\phi$ — which is C1's goal — the differences between them are important; I will explore them in sections 4 and 5.

As previously noted, I am setting aside non-central cases of moral coercion. For example, I am setting aside instances in which C1 *deceives* C2 into justifiably but mistakenly believing that a greater harm will occur unless C2 commits a lesser harm. I set aside such cases, not because I do not think that they should be analyzed under the aegis of moral coercion, but because I think that addressing instances of sincere moral coercion is a prerequisite for an analysis of insincere moral coercion.

I am also setting aside "partial" moral coercion. These are cases where either $\phi$ or $\psi$ does *not* wrong a third party. Suppose C1 promises to provide supplies for famine relief efforts on the condition that C2 harm C1's innocent enemy. This counts as partial moral coercion (assuming that refraining from giving to charity does not wrong those who are in need). Likewise, suppose C1 threatens to kill an innocent unless C2 gives the villain a thousand dollars. Acceding to C1's demand does not wrong a third party; accordingly, this is also an example of partial moral coercion. This is in contrast to HOSTAGE and SHIELD, in which both options available to C2 impose a wrong on a third party. Though I will not consider cases of partial moral coercion explicitly, the account I develop will have implications for it.

Before discussing what grounds the wrongfulness of moral coercion, one issue remains. What partly characterizes *moral coercion* is the type of reason operative in C2's deliberation between $\phi$ and $\psi$. The operative reasons are moral, as opposed to pragmatic; this is what distinguishes moral coercion from non-moral coercion. But what if C2's pragmatic and moral reasons converge on the same option? Suppose C1 kidnaps C2's child, whom C1 threatens to kill unless C2 robs a bank and gives the money to C1. Does this count as *moral* coercion? Even if C2 is morally obligated to accede to C1's demands, it is doubtful — presuming she is a psychologically typical parent — that she is motivated by *moral* reasons. Rather, she is motivated out of what C.D. Broad famously called a "self-referential altruistic" concern.[4] Such a concern, despite that it is not overtly moral, is morally appropriate given her relation to her child. This is to make the oft-noted point that special relations grounding *moral* reasons need not function as the agent's *motivating* reasons.[5] In this case, does C1's coercive threat count as moral or non-moral, given that C2 has moral reasons to accede but is motivated by reasons of practical rationality? The account of moral coercion I have outlined provides no definitive answer. But this is, I believe, a strength of the account: the line between moral and non-moral coercion *ought* to be blurry precisely because the line between moral reasons and reasons of practical rationality is blurry.

### 3. The Wrongfulness of Moral Coercion

Now that I have outlined what moral coercion is, we can ask: What makes wrongful instances of moral coercion wrongful? In answering this question, it is helpful to consider *non-moral* coercion.[6] In cases

explosion. This is an example of active moral coercion in which $\psi$ would be coercee-enacted rather than coercer-enacted.

4. (Broad, 1930, pp. 54–55)

5. See especially (Railton, 1984).

6. There are at least three types of (non-moral) coercion, broadly conceived. In "act-negating" coercion, C1 physically forces C2 to commit $\phi$. In such a case C2 literally has no choice but to be an instrument in the commission of $\phi$. In "autonomy-negating" coercion, C1's threat puts C2 under duress, of the sort that makes it psychologically impossible (in a modally weak sense) for her to refuse to acquiesce. See (Frankfurt, 1973, p. 78). In "compossibility-negating" coercion, C1 provides an incentive for C2 (who is not under duress) to commit $\phi$ by ensuring that some *other* event, $\psi$, will occur if C2 refuses to commit

of both moral and non-moral coercion, C1 disincentivizes non-compliance with her wishes by attaching a cost to C2's non-compliance. The difference between moral and non-moral coercion is where those costs lie. In cases of non-moral coercion, C1 forces C2 to choose between perceived costs to *her own* interests, whereas in cases of moral coercion, C2 has to choose between perceived costs to the interests of third-party innocents.[7] Determining what the wrong-making features of wrongful non-moral coercion are can help determine what makes moral coercion wrongful. (I will refer to cases of non-moral coercion simply as "coercion".)

I do not here attempt to provide necessary and sufficient conditions for when coercion is wrongful. Instead, I present a particular and important necessary condition of wrongful coercion which explains the sense in which such coercion *uses* its victim. On the view I defend, C1 uses C2 by turning the teleological structure of C2's goals on its head. C1 puts the victim in a position where C2's goals become self-undermining in that *she would better-achieve her own goals if she didn't have them*. That is, by having such goals, C2 worse-achieves them, as a result of C1's influence. A coercive act, then, uses its victim by putting her in a position where her commitments further the achievement of the opposite ends to which those commitments are teleologically directed.

Consider the example of blackmailing an adulterer. If the adulterer didn't have the aim of keeping his adulterous conduct private, then the blackmailer's attempt at blackmail would find no purchase, since the adulterer would have no incentive to accede to the blackmailer's demand. It is the adulterer's own goal of keeping his adultery secret that makes him worse off.

I will call the tactic of intentionally putting someone in a position where her goals become self-undermining "hacking" her aims.[8] Hacking an agent's aims does not necessarily wrong her, in that we do not have a fundamental right against aim-hacking. But treating someone in this way becomes a distinct wrong — not just instrumentally, but in itself — when it is used in furtherance of an end that wrongs the agent. In general, using an agent in furtherance of wronging her treats her wrongly over and above the wrong furthered.[9] And using an agent's *own aims* as a means of wronging her is an especially iniquitous way of treating her.

Extant accounts of coercion are incomplete in that they miss the necessary role that aim-hacking plays in explaining how wrongful coercion wrongs its victim. For instance, some argue that wrongful coercion consists in impermissibly constraining C2's deliberative options,[10] or denying her the standing to legitimately demand that the coercer abandon her intention,[11] or impermissibly attaching a cost to an option she already had and to which she has a right.[12] But these accounts underdescribe how C2 is wronged. Treating C2 in these ways is wrong also because it involves hacking her aims. Adopting this tactic in furtherance of wronging C2 *itself* maltreats her, thereby compounding an already existing wrong. An account must appeal to both sources of wrongs to explain fully what makes wrongful instances of coercion wrongful.

The account of wrongful coercion I have presented is, then, open-ended. It doesn't provide necessary and sufficient conditions for when

$\phi$, thereby denying her the option of bringing about both $\sim\phi$ and $\sim\psi$. This is the type of coercion I am concerned with here.

7. "Mixed" cases are also possible, in which C1 threatens to harm a third party unless C2 complies with demands detrimental to C2's interests. Though I will focus on "pure" cases, where the coercion is wholly prudential or wholly moral, the account I develop will be applicable to mixed cases as well.

8. To count as having hacked C2's aims, C1 need not wish that C2 actually accede to C1's threats. It is enough if C1 intends for C2 to be in a position where her self-interested commitments make her worse off than she would be if she didn't have those commitments.

9. This general idea is found elsewhere, such as in Thomas Nagel's account of the wrongness of using chemical, biological, and incendiary weapons, and David Sussman's account of what makes torture wrongful (see (Nagel, 1972) and (Sussman, 2005)).

10. See (Shaw, 2012).

11. See (Pallikkathayil, 2011).

12. See (Wellman, 2005, pp. 132–138).

coercion wrongs its victim. Rather, it says that whatever account of wrongful coercion we adopt, it must include the claim that part of what it is to coerce someone is to engage in aim-hacking — and that treating someone in this way becomes wrongful in itself when it is used as a tactic in furtherance of wronging her in some other way, as specified by a candidate theory of wrongful coercion. This is an important necessary condition of coercion in this respect: a desideratum of any account of wrongful coercion is that it identifies its wrong-making features at least partly in the manner in which the coercer uses her victim. And I claim that wrongful coercion does so in part by putting C2 in a position so that her own goals become self-undermining as a means of wronging C2. I will call this the "Aim-Hacking Condition" of wrongful coercion.[13]

The Aim-Hacking Condition of wrongful coercion generalizes to cases of passive coercion as well. Consider this case:

> BOAT
>
> C1 and C2 are on a boat. A gust of wind has blown overboard C1's only life-jacket. She is a cautious person and does not want to continue without it. But neither does she want to get wet, which is necessary to retrieve the life-jacket. So she pushes C2 into the choppy water, knowing that C2 cannot swim and that C2 will consequently be motivated, on pain of drowning, to grasp and don the life-jacket. Once she climbs back on board, C1 will have her life-jacket.

In this example, C1 does not conditionally threaten C2 in any way. There is no coercive proposal she puts to C2, even implicitly.[14] And C1's actions, subsequent to pushing C2, are not sensitive to how C2 chooses to respond. C1 has set in motion a series of events over which she has no further control. Of course, C2 might choose to drown. But C1 knows that this is a very unwelcome prospect for her; this is why she pushes her in the first place — to force her to choose between the exclusive options of retrieving the life-jacket and drowning.

Passive coercion is importantly similar to active coercion, which involves the pronouncement of conditional threats. They both impose a forced choice on C2, in which both options worse-achieve C2's aims relative to the relevant alternative (which is usually the status quo ante).[15] The choice contrary to C2's preferred outcome imposes a greater cost on C2 than the alternative, thereby disincentivizing non-compliance with C1's preferred outcome. Part of what characterizes both active and passive coercion is that C1 denies to C2 the conjunctive option of $\sim\psi$

13. The Aim-Hacking Condition helps dissolve the paradox of blackmail. According to the paradox, it is wrong for C1 to threaten to reveal C2's infidelity as a means of coercing money from her, even though C1 is permitted to reveal the infidelity without attaching a cost to her silence. So it seems (contrary to a basic account of coercion (see, *e.g.*, (Haksar, 1976)) that the wrongfulness of acting on the threat cannot explain the wrongfulness of the threat itself (see (Lindgren, 1984) and (Berman, 2011)). James Shaw defends the basic account of coercion by arguing that the cost that C1 attaches to C2's noncompliance constitutes an impermissible sanction since it manifests a morally problematic "disregard" for C2, in that the harm C1 threatens to cause is not offset by the value of the ends that C1 furthers in so doing (Shaw, 2012). But Shaw admits that if revealing the infidelity absent the demand would be morally discretionary, then blackmailing C2 would not be a wrongful instance of coercion (185). He instead appeals to C1's problematic disregard for C2 to explain how C1 wrongs C2 in such cases. But grounding the way that the blackmailer wrongs C2 by appealing to a kind of impermissible influence *other* than coercion leaves out a morally relevant aspect of the relationship between them. The account I present fills the gap in Shaw's account: hacking C2's aims in furtherance of an end that wrongs her impermissibly *coerces* her.

14. For these reasons, one might protest that passive coercion is not a species of coercion at all. Some might want to restrict the concept 'coercion' to acts that involve pronouncing conditional threats of a certain sort. I am happy with such a view, so long as its proponents recognize a fundamentally unifying feature between the two types of acts: they both involve using the agent by hacking her aims.

15. The problem of determining what this relevant alternative is — *i.e.*, what the "baseline" is against which we determine whether C2's options make her worse off — has persisted since Nozick discussed it (Coercion, 1969). One option (which Alan Wertheimer takes (Wertheimer, 1987)) is to abandon a non-moral baseline in favor of a moralized one, according to which both $\phi$ and $\psi$ make C2 worse off than she *ought* to be. The Aim-Hacking Condition is also buck-passing insofar as I claim that aim-hacking wrongs its victim (not just instrumentally but in itself) only in furtherance of an end that wrongs her.

and ~ϕ — C2 can choose only one. The only difference between active and passive coercion is *how* C1 denies this conjunctive option.

Part of what explains what makes particular instances of passive coercion wrongful is that they involve intentionally putting someone in a position where she worse-achieves her own aims as a result of having those aims — and she is put in this position in order to achieve aims that wrong her. In BOAT, for example, pushing C2 overboard would not serve C1's aim if C2 didn't have the goal of staying alive. And the act of retrieving the life-jacket by pushing C2 overboard itself wrongs C2. So C1 wrongs C2 twice over: by foreclosing options to which C2 has a right, and by hacking her aims in furtherance of that end.

So the Aim-Hacking Condition helps explain how both the victim of wrongful active coercion and the victim of wrongful passive coercion are mistreated; the account, then, treats fundamentally alike cases alike. This is not to say that there are no important moral differences between active and passive coercion.[16] Rather, the point is that they share an important similarity which the Aim-Hacking Condition helps reveal.[17]

We are now in a position to better see what makes wrongful instances of *moral* coercion wrongful. In general, moral coercion is wrongful when it either a) wrongs C2 or b) wrongs the third party whose well-being is used as leverage to coerce C2. Of course, the substantive task is to explain *how* wrongful instances of moral coercion wrong C2 or third-party innocents. I will start with the latter, since its explanation is simpler.

Consider cases such as SHIELD or HOSTAGE. In these cases, C1 endangers third-party innocents. But this alone is not necessarily wrongful. It can be permissible to endanger those who have a right not to be harmed if doing so has a sufficiently high probability of yielding a sufficiently important moral good, and if the probability that they will actually be harmed is sufficiently low. But in SHIELD and HOSTAGE, C1 endangers third-party innocents in furtherance of achieving a wrongful end. No matter how low the probability is that the innocents will actually be harmed, endangering the innocents wrongs them if it is in furtherance of a wrongful end. So there are *two* factors undergirding the impermissibility of endangering the innocents: 1) it is impermissible qua means to the achievement of some further wrongful end, and 2) it is impermissible insofar as it endangers the innocents. It is due to 1 that 2 violates the rights of the innocents.

That the coercion in HOSTAGE and SHIELD wrongs third-party innocents does not explain, however, how or why *C2* is wronged. After all, threatening to harm innocents does not itself violate the rights of C2 (unless, of course, she has agent-relative interests in the welfare of those innocents). We can appeal, then, to extant accounts of wrongful coercion to explain how C1 wrongs C2. A theory (such as Wellman's or Shaw's) might tell us that in HOSTAGE C2 has a right to refrain from killing C1's innocent enemy without sanctions attached to that option. Likewise, in SHIELD, such theories say that C2 has a right to disable C1 without thereby killing innocent bystanders. But this under-describes how C1 wrongs C2 in these cases. As Nancy Davis puts it, we "think of ourselves as instruments of evil when we are 'blackmailed' to inflict pain or cause deaths at the bidding of evil men. And we find the role 'instrument of evil' an especially repugnant one, an assault on our dignity, and a threat to our status as autonomous moral agents".[18] Terrence McConnell makes a similar point:

16. Benjamin Sachs, for instance emphasizes a morally important aspect of *pronouncing* conditional threats — namely that doing so motivates the threatener to enforce the threat (Sachs, 2013).

17. The Aim-Hacking Condition explains how C2 is used only given independent grounds for thinking that C2 is wronged. So the Aim-Hacking Condition will not explain how C2 is used in cases of passive coercion if it is combined with an account such as Japa Pallikkathayil's, since her account focuses on the wrongfulness of coercive speech-acts, which are absent in cases of passive coercion. This is not a failure of Pallikkathayil's account — her aim was to provide an account of active coercion. But it does show that for the Aim-Hacking Condition to have the advantage of explaining how C2 is *used* in both active *and* passive coercion, it must be combined with an account of wrongful coercion that provides some independent grounds for how C1 wrongs C2 in such cases.

18. (Davis, 1980, p. 202)

> For reasons that are not always easy to explain, we are especially repulsed by the idea of one moral agent manipulating the other. This, of course, is what happens in cases of moral blackmail. The blackmailer attempts to get a person to do certain acts by threatening to do something much worse. To the extent that the person complies with these demands, he is surrendering his moral autonomy. He is, in a sense, a puppet in the blackmailer's hands.[19]

Part of the challenge is to make more precise how we are being used when we are subjected to moral coercion, and how this use violates us. The Aim-Hacking Condition does this. Moral coercion wrongs C2 in that her aim — specifically, her commitment to morality — is being leveraged to serve as a means in furtherance of an unjust end. After all, if C2 were not a moral person, C1's efforts at coercion wouldn't work. It is precisely because C2 has the moral commitment of preventing the worse outcomes in SHIELD and HOSTAGE that C1 is in a position to morally coerce C2.

In cases of non-moral coercion, the *self-interested* motivations C2 has are used against her; if she lacked a motivation to prevent perceived harm to herself, then paradigm examples of coercion would fail to incentivize compliance. In cases of moral coercion, C1's *moral* motivations are used as a means of undermining their own purpose. At one remove, then, wrongful moral coercion wrongs C2 in the same way that wrongful non-moral coercion does — both involve using C2 as a means to the achievement of C1's ends by intentionally putting C2 in a position where she worse-achieves her own legitimate aims as a result of having those aims. So we can appeal to the Aim-Hacking Condition to explain a manner in which wrongful moral coercion wrongfully uses C2: she is wronged not only in the way specified by standard theories of coercion but moreover by having her aims hacked in furtherance of those wrongs. Doing so uses her own aims as a means to wronging her, which is wrongful in its own right. The application of the Aim-Hacking Condition to cases of *moral* coercion helps reveal two additional ways in which C1 wrongs C2 by morally coercing her.

First, C1 relies on C2's compliance with the very norms that C1 wrongly flouts (or threatens to flout) as a means of achieving her own ends. Unilaterally flouting norms is unfair to those who comply with the norms, when their compliance benefits the flouter (as in cases of free-riding). By counting on C2's compliance with the very norms that C1 wrongly flouts, C1 treats C2 unfairly.

Second, when C1 morally coerces C2, the aims that are hacked are not merely aims that C2 is *entitled* to have — *i.e.*, prudential commitments — but aims that both C1 and C2 are *obligated* to have — *i.e.*, moral commitments. In cases of non-moral coercion, it is typically procrustean to ask whether we are permitted to refuse to accede to the coercer's demands, since the sorts of norms operative in such a case are norms of practical rationality, rather than norms of morality. Even if there are decisive reasons for C2 to accede to non-moral coercion, she is still morally permitted, all things considered, to refrain from doing so. But if there are decisive reasons for C2 to accede to moral coercion, then by definition she is not morally permitted to refrain. Violating prudential requirements is discretionary in a way that violating moral requirement is not. So even if the psychological pressure associated with both sorts of coercion are equal in severity, moral coercion can trap its victim in a way that non-moral coercion cannot, by foreclosing any moral permission to do other than what C1 wants. Morally (rather than merely prudentially) foreclosing an option to which C2 should be entitled wrongs C2 in a way over and above the way she is wronged when she is non-morally coerced. Consequently, whereas there is a sense in which non-moral coercion is morally "escapable" (albeit at potentially significant cost to C2), instances of moral coercion might not be similarly escapable.

In summary, moral coercion, whether active or passive, wrongs C2 in three ways. First, C1 uses C2 in furtherance of aims that wrong her by hacking her aims. Second, C1 treats C2 unfairly by relying on her

19. (McConnell, 1981, p. 562)

compliance with the very norms that C1 wrongfully flouts. And finally, insofar as these are commitments that C2 is not merely entitled but obligated to have, moral coercion can wrong C2 by illicitly foreclosing any moral permission to do other than what C1 wants.

One might argue that the account I have presented over-generalizes, in that moral coercion does not always involve a pro tanto wrong. Consider the following case of what might be called "reverse moral coercion".

> RACIST
>
> A flash flood is endangering an innocent. C1 is not in a position to save her, but C2 can do so, at little cost to herself. However, C2 is a racist who believes that morality requires the extermination of the innocent's race. She consequently believes that the innocent should not be saved. Knowing all this, C1 conditionally threatens to devote himself to a lifetime of charity directed to members of the innocent's race, unless C2 saves that innocent.

Here, C1 is coercing C2 into committing a *good* act which C2 mistakenly believes to be wrongful by threatening to do something which C2 mistakenly believes is even more wrongful. One might argue that C1's act is not even pro tanto wrongful. But the account of moral coercion that I have presented is consistent with this result, in that none of the three explanations of moral coercion's wrongfulness apply in RACIST.

First, hacking C2's aims in furtherance of C1's goals wrongs C2 only if C1's goals wrong C2. In RACIST, C2 has an enforceable positive duty to save the innocent — consequently, it does not wrong her to use her mistaken commitments in furtherance of enforcing that duty. So, though C1 uses C2, this does not wrong C2.

Second, C1's conduct does not treat C2 unfairly. Though C1 is relying on C2's compliance with racist norms that C1 is herself flouting as a means of incentivizing the desired act, these are not norms that C1 is obligated or even permitted to abide by. Thus C1 does not treat C2 unfairly by refraining from abiding by them herself.

Third, though C2 is morally obligated to do what C1 wants — *i.e.*, to save the innocent — this is not because C1 has foreclosed an option to do otherwise. Rather (as I noted), C2 was antecedently required to save the innocent (though C2 mistakenly thinks otherwise).

One might raise another challenge. I have claimed that moral coercion wrongs C2 by putting her in a position where she worse-achieves her moral aims as a result of having those aims. At the time that C2 is coerced, the world would be better off if C2 didn't have her moral commitment — which is precisely the opposite effect that her moral commitments are supposed to have. One might argue that this analysis problematically presumes that C2 must be a consequentialist. But consider a committed deontologist who believes, for instance, that the Doctrine of Doing and Allowing provides absolutist agent-centered constraints against committing harms. Now suppose that this deontologist is C2 in ALLEY, or SHIELD, or HOSTAGE. How would C2 respond? She certainly wouldn't comply with C1's wishes — even though doing so would make the world better. This is because compliance would require violating absolutist agent-centered constraints against committing harms. This suggests that hacking moral aims works as a tactic only against those whose moral reasoning is at least partly teleological. This shouldn't be a surprise, since moral coercion functions precisely by threatening to make things go worse. Such a threat will have little purchase on an absolutist deontologist. The upshot is that the sort of consequentialist reasoning I tacitly impute to C2 is appropriate. Otherwise, C2 couldn't be coerced.[20]

20. But suppose C1 says to C2, an absolute deontologist, "Commit harm x now, or I will put you in a position where *you* will be forced to commit ten such harms in the future." Whether C2 should comply depends on whether agent-centered constraints against doing harm are not only agent-relative, but *time*-relative as well. A doubly relative deontologist wouldn't commit x, even if it meant that she would have to commit worse harms in the future. But a *time*-neutral deontologist would indeed be coercible. This is precisely because such a deontologist is committed to *promoting* her own non-violation of agent-relative constraints over time. So it is because there is a residual

Now that we have a better picture of what makes moral coercion wrongful — and specifically of how it involves using C2 — we can turn to the Liability and Badness Questions.

## 4. The Liability Question

Suppose C2 is morally obligated to accede to C1's coercive demand. That is, suppose C2 is morally obligated to choose $\phi$ over $\psi$. Even in such a case the third-party innocent victim of $\phi$ might still be morally entitled to defend herself against the harm that C2 is committing, or to compensation. That is, C2 is morally liable to defensive or compensatory harm (where a person is liable to be harmed just in case she has done something to forfeit her right not to be harmed in that way). But how can it be that C2's victims are entitled to prevent (or seek compensation for) what C2 is morally obligated to do?

What justifies choosing $\phi$ is that it is the lesser evil; this means that the occurrence of $\phi$ will still infringe (though not violate) the rights of an innocent. Since this victim has done nothing to lose her right not to be harmed, to harm her wrongs her, even if wronging her is the right thing to do, all things considered.[21] This is why the victim (or her estate) can be owed compensation for the harm she suffers, even if C2 permissibly imposes this harm.[22] And the fact that the innocent is wronged can also ground an agent-relative permission for the innocent to engage in proportionate defensive violence necessary to prevent the harm she is threatened with — even though C2 is obligated, from an agent-neutral standpoint, to impose that harm.

So, at least in principle, the third-party victim can be entitled to impose defensive and compensatory harms. But *on whom* is she morally permitted to impose such harms? At first, it seems unfair to say that she can impose them on C2. This seems unfair not because C2 was morally obligated to commit $\phi$ — one can have agent-relative reasons to prevent what another person has an agent-neutral obligation to do. Rather, imposing the harms on C2 might seem unfair because C2 is, after all, a victim of coercion. In cases of moral coercion, there is a harm which must fall somewhere, and C2 is in a position to choose where the harm will fall among a limited menu of options — but she is not responsible *for the fact* that the harm must fall somewhere. Rather, the agent who is morally coercing her is the one responsible for the predicament. It seems unfair to hold C2 responsible for the harm she commits, when C1 is responsible for putting C2 in the predicament in the first place. Though C2 is the proximate cause of the harm, and C1 is causally "upstream", C1 bears more responsibility for that harm, because the degree of responsibility depends on the options available — and C1 has the option of choosing both $\sim\phi$ *and* $\sim\psi$, whereas C2 has only the option of choosing $\sim\phi$ or $\sim\psi$, exclusively.

Because C2 is less responsible than C1, the victim of $\phi$ (or her estate) should seek redress from C1 rather than from C2, given that she can choose only one or the other. Likewise, if the potential victim of $\phi$ can prevent her own death either by killing C1 or by killing C2, it is clear that C1 is the more morally appropriate target.

The claim that C1 is more responsible than C2 for $\phi$ is, however, compatible with the claim that C2 is no less responsible than she would be if she committed the harm in response to a functionally equivalent adventitiously imposed dilemma (*i.e.*, a dilemma imposed by happenstance), rather than in response to C1's coercion. This point can be put differently. C2's diminished responsibility for $\phi$ is grounded solely in the fact that she has a limited menu of options from which to choose. That her menu of options is limited *as a result of being coerced by C1* does not itself diminish her responsibility over and above the degree to which it is mitigated as a result of having her menu of options reduced. To see this, consider the following modification of ALLEY:

teleology to C2's non-consequentialism — *i.e.*, its time-neutrality — that she is coercible. So imputing some sort of teleological thinking to C2 is appropriate if we are to assume that she is coercible. (For more on the distinction between time-relative and time-neutral views, see (Louise, 2004)).

21. For more on the distinction between infringing and violating rights, see (McMahan, 2009, p. 10).

22. For more on this, see (Rodin, 2012).

ALLEY 2

> Thirty children, separated from their class during a field trip, find themselves in an alley, at a dead end. C2, who has also accidentally turned into the alley, is several meters from the children. And several meters from C2 is another innocent. At that moment a bomb, left over from the previous war, happens to roll off the roof of the tall building above C2. As it bounces down toward the children, it arms. The only way for C2 to save the lives of the children is to intercept the bomb and throw it in the only direction available to her, which is down the alley, toward the innocent several meters away, foreseeably killing her. C2 does so.

The only difference between ALLEY and ALLEY 2 is that in the former case an agent intentionally forecloses options that would otherwise be open to C2, whereas in the latter case the options are foreclosed as a matter of happenstance. We can even imagine that both cases are identical from C2's standpoint: from her perspective in both cases the bomb simply falls from the sky. It would be strange to think that C2 is less responsible for the pro tanto harm she causes in ALLEY that in ALLEY 2, even though she is wronged in the former case but not the latter. This suggests that coercion *per se* does not diminish C2's responsibility for what she does. But the fact that her options are delimited, in combination with the fact that she does not intend the death of the innocent, makes her less responsible than C1 for $\phi$.

Though C2's responsibility for what she does in ALLEY and ALLEY 2 is the same, the degree of liability she bears can differ between the two cases. This is because, in ALLEY but not in ALLEY 2, there is someone else who is *more* responsible than C2. So if the innocent could save her own life by killing C2 in ALLEY 2, she would be permitted to do so. But whether she is so permitted in ALLEY — that is, whether C2 is liable to be defensively killed — depends on whether the victim is able to target C1 instead. If she can, then C2 is not liable to be killed.[23]

The upshot is this: C2's liability to defensive and compensatory harms depends on a) the degree of responsibility she bears for the harm she imposes on the victim, b) whether there is anyone else who is more responsible for that harm, and c) whether C2's victims are able to impose defensive and compensatory harms on a more responsible party (viz., C1). What is important to recognize here is that being *wronged* by being coerced does not itself affect C2's liability for what she does in response to being coerced.

Note that everything that has been said also applies in cases where $\phi$ is coercer-enacted. In cases such as SHIELD or HOSTAGE 2, C2 must decide between allowing a lesser harm and committing a greater harm. Suppose that she chooses the former; she is clearly less responsible than C1 for the harms she allows, both because C1 is responsible for putting C2 in this predicament *and* because C1 is the one who actually commits the harm. But she is no less responsible for the harms she allows than she would be if she allowed the same harms in response to facing a functionally equivalent adventitiously imposed dilemma. Again, whether C2 is liable to defensive or compensatory harms depends on whether such harms can be imposed on C1 instead.

Though being wronged does not diminish C2's responsibility for what she does, it can (as we shall see) serve as a pro tanto reason for refusing to accede to moral coercion.

23. It might seem strange that one's liability is contingent in this way — that it can appear and disappear depending on whether the more responsible party can be targeted. Liability is instrumental in this way due to the role that negative rights play in our moral economy. Negative rights protect us from being used without our consent as a means to the achievement of another's ends. When a person infringes another's right not to be used as a means, the infringer herself forfeits her right not to be used as a means to preventing or rectifying the harms which she was threatening or imposing. A rights-infringer can become liable to the means required to prevent or rectify that rights-infringement. Thus liability is necessarily instrumental, because its function is to prevent a rights-infringement or to restore a right.

## 5. The Badness Questions

So far I have argued that being morally coerced does not itself affect C2's liability for the harms she fosters when she chooses $\phi$. But how do we determine *whether* C2 ought to accede to C1's wishes, when C2 is being morally coerced? I obviously cannot address all the factors relevant to determining whether C2 ought to accede; but I will address one factor that plays a special role in cases of moral coercion — and that is the role of intention.

In addressing this issue, I will argue as follows:

1. Moral coercion can efface the relevance of the intention/foresight distinction, in that when C2 is weighing $\phi$ against $\psi$, the former should be weighed as heavily as it would be if it were committed *intentionally*, even if C2 is actually committing it collaterally (*i. e.*, foreseeably but non-intentionally).

2. In defending 1, I will argue that we need not think that intention has only *first-personal* (and not *third-personal*) relevance in our moral deliberations. Even if what grounds the moral relevance of the intention/foresight distinction is wholly first-personal, it still has third-personal relevance.

3. A consequence of 1 is that we have a basis for thinking that acceding to moral coercion is worse than acceding to a functionally equivalent adventitious moral dilemma.

4. C2 has a (defeasible) agent-centered prerogative against being morally coerced, which means that in weighing the disvalue of acceding to C1's moral coercion against the disvalue of resisting it, C2 can augment the latter.

To determine whether C2 is morally permitted to accede to moral coercion, we have to weigh the morally relevant costs of doing so against the morally relevant benefits relative to the relevant alternative — which is to refuse to accede. That is, C2 has to choose among evils, and in doing so, C2 will be committing, enabling, or allowing a harm.

One constraint upon imposing a harm is the constraint of proportionality, according to which the aversion of the relevant evils must be worth the severity of the pro tanto wrongful harms imposed. And one factor relevant to determining whether a harm satisfies the proportionality constraint is whether that harm is committed intentionally. Committing a harm intentionally rather than collaterally is relevant in that the former receives greater negative weight in the calculation of proportionality than the latter.[24] It might seem, then, that instances of coercion in which C1 coerces C2 into *intentionally* committing a harm (such as HOSTAGE) are worse than instances of coercion in which C1 coerces C2 into committing a harm merely *foreseeably* (such as in ALLEY). But this is an overly simplistic picture of how intention functions in cases of coercion. In all such cases C2 is fulfilling C1's aims. By foreclosing the conjunctive option composed of $\sim\psi$ and $\sim\phi$, C1 intentionally incentivizes the commission of $\phi$, thereby effectively enlisting C2's assistance — albeit without her consent — in furtherance of C1's aims. Under these conditions, the moral measure of C1's action depends, *inter alia*, on the intentional status of C1's action.

To see this, consider again ALLEY. Suppose C2 throws the bomb away from the children whom it would otherwise kill. In doing so, she foreseeable kills an innocent bystander — which is precisely the outcome at which C1 was aiming. C1's role affects the morality of what C2 does in the following way: though C2 did not intend the innocent's death, it should be weighed as heavily as an intentional killing, because a) C1 aimed at the death of the innocent, and b) C1 furthered that aim by contributing substantially to that death. Indeed, it is not infelicitous to describe the innocent's death as intentional, even though C2 did not kill her intentionally.

This suggests that, even presuming the moral relevance of the intention/foresight distinction, the pro tanto wrong which C2 commits in HOSTAGE is no worse than the pro tanto wrong which C2 commits in ALLEY — *even though in the latter case C2 kills merely foreseeably*

24. See (McMahan, 2009), Ch. 1.

*rather than intentionally*. Accordingly, when we do the calculation of proportionality determining whether to commit $\phi$, that act should be weighed as heavily as an *intentional* harm in the calculation. In this respect, moral coercion can efface the relevance of the intention/foresight distinction.

One might argue in response that intention has only *first-personal* and not *third-personal* relevance in the calculation of proportionality. It is morally worse for me to intentionally commit a wrongful harm that it is to do so collaterally, in that such a harm should receive greater disvalue in the proportionality calculation determining what I should do. But the fact that a harm I bring about was intentionally sought by someone else — who coerced me into committing it — does *not* entail that it should similarly receive greater disvalue in the calculation of proportionality. On this view, the constraint against *intentionally* killing innocents is agent-relative. I will call this "the agent-relative view".

On the agent-relative view, the augmented disvalue that a harm receives for being intentional is agent-relative in that it appears only in the intender's calculation determining the permissibility of committing the act. It does not appear in anyone else's, including causally "upstream" agents who enable the intender to commit the harm as well as causally "downstream" agents who commit the intended harm.[25] In doing the calculation of proportionality, the causally downstream agents should not augment the disvalue of the harm they are committing, even though it was intended — since it was not *they* who did the intending. This is in contrast to a view according to which the moral relevance of the intention/foresight distinction is grounded in agent-neutral reasons to prefer the latter to the former. I will call this the "agent-neutral view".

The agent-relative view has been explicitly endorsed by some, including, notably, Thomas Nagel. He says, in "The View From Nowhere", that even though it's morally more objectionable for me to harm someone intentionally than collaterally, it's not true that if I can either stop one person from being intentionally murdered or another from being killed collaterally, I have a stronger reason to do the former.[26] On this view, the agent-relative reason to abide by the constraint against intentionally committing pro tanto wrongs is not grounded in the promotion of any *value* that aggregates across instances of compliance with the constraint. A world in which innocents are merely foreseeably killed rather than intentionally killed is indeed a better world, but this is not *why* there is a stronger agent-relative duty against killing. Rather, a world in which innocents are merely collaterally rather than intentionally killed is a better world because such a world is one in which people are abiding by the agent-relative constraint against intentionally killing. This reflects a traditional picture of deontology, in which the right precedes the good.

But even if we accept the agent-relative view, there are, I believe, still agent-neutral reasons to see to it that pro tanto wrongs are not committed intentionally. On my view, all agent-relative constraints generate a corresponding agent-neutral reason to promote compliance with the agent-relative constraint.[27] That is, the agent-relative view entails the agent-neutral view. Accordingly, a world in which innocents are not intentionally killed is a better world — but this is not what *grounds* the agent-relative constraint. Rather, such a world is better precisely because agent-relative constraints are being met. And though this does not ground the agent-relative reason to refrain from killing intentionally, it *does* generate an agent-neutral reason to see to it that innocents are not killed intentionally.

Against this, one might deny not only that making the world go better is what grounds agent-relative constraints against intentional harming, but also that abiding by agent-relative constraints has agent-neutral value. But such a view comes at a significant cost. Compare two worlds which are the same except that in the first everyone violates agent-relative constraints, and in the second everyone abides

25. An exception is if they are intentionally cooperating in furtherance of a joint goal. I address this in (Bazargan, 2013).

26. (Nagel, 1986, p. 178)

27. The view closest to this picture belongs to (McNaughton & Rawling, 1995).

by them — and we can choose between these two worlds. If abiding by agent-relative constraints has no agent-neutral value, then the fact that in the first world everyone violates agent-relative constraints would itself provide literally no reason to prefer the alternative world. And this seems like a big bullet to bite.

Of course, the claim that agent-relative constraints generate corresponding agent-neutral values leaves open the possibility of conflict between an agent-relative reason to refrain from intentionally killing an innocent, and an agent-neutral reason to see to it that intentional killings of innocents are minimized. How we resolve this conflict depends on how we weigh these competing reasons — which is beyond the scope of this paper.[28]

So the agent-relative reason to prefer committing collateral over intentional harms generates an agent-neutral reason to prefer that others commit collateral over intentional harms. Where does this leave us? I argued that even if C2 does not intend $\phi$, it should be weighed as heavily as an intentional harm, because a) C1 aimed at $\phi$'s occurrence, and b) C1 furthered that aim by contributing substantially to $\phi$. One might attempt to forestall this argument by claiming that *someone else's* intentions regarding what *I* do generally cannot affect the morality of what I do. But I responded to this by arguing that the intention/foresight distinction can have third-personal relevance: by committing $\phi$, C2 brings to fruition C1's wrongful intentions, which is why it should receive as much weight in C2's deliberations as an *intentionally* committed harm. Hence moral coercion can efface the relevance of the intention/foresight distinction, *even though* the moral relevance of the intention/foresight distinction is grounded in agent-relative reasons not to intentionally commit pro tanto harms.

The fact that C1's intentions are relevant to the moral measure of what C2 does when C2 accedes does not mean that C2 ought never to accede when doing so requires enacting $\phi$. But the fact that $\phi$ receives augmented disvalue in the calculation of proportionality makes it less likely that the proportionality calculation will work out in favor of acceding under these circumstances. In addition, the fact that C1's intentions affects the morality of what C2 does means that we have grounds for thinking that it is worse for C2 to accede by enacting $\phi$ than it would be if she responded in an analogous fashion to a morally adventitious dilemma (such as ALLEY 2). We have, then, grounds for the intuition that by acceding to moral coercion we allow "evil to succeed": the manipulator's evil intentions are manifest in our actions when we enact her aims; choosing $\phi$ ought to be weighed accordingly.

Does the same argument apply to cases of moral coercion where $\phi$ is *coercer*-enacted, as in SHIELD or HOSTAGE 2? In these cases, C2 is instructed to do nothing — to refrain from preventing C1 from achieving her wrongful aims (which C2 would otherwise prevent). Intentions can be relevant to the moral assessment of omissions. But C2's omission cannot properly be said to be a *cause* of $\phi$; as a result, we lack a necessary basis for thinking that C1's intentions affect the moral measure of C2's omission. For this reason, I do not think that the agent-neutral view provides a basis for augmenting the disvalue of choosing $\phi$ when $\phi$ is coercer-enacted. A consequence is that it should be more difficult to justify acceding to moral coercion in cases where doing so involves actually committing $\phi$, not merely because of the moral relevance of the difference between doing and allowing, but because $\phi$ is a wrong intended by C1. That is, the action/omission distinction combines with the intention/foresight distinction to explain how and why acceding to coercee-enacted moral coercion is especially egregious when $\phi$ is a morally wrongful aim.

So far I have discussed the relevance of the intention/foresight distinction as it applies to determining when and whether we ought to

28. One might note that if the moral relevance of the doing/allowing distinction is grounded in an agent-relative reason against *committing* harms, and if the agent-relative view entails the agent-neutral view, then we have an agent-neutral reason to choose allowings over committings in general (for harms of equal severity). It might seem that choosing in this way is not possible, since allowing an act entails that *someone else* commits that act. But this overlooks the fact that *not all harms are the result of acts*. There is, then, a straightforward way to characterize the agent-neutral counterpart of the agent-relative grounds for the relevance of the doing/allowing distinction: it is a reason to choose non-committed harms over committed harms (of equal severity).

accede to moral coercion. There is another characteristic endemic to moral coercion and relevant to the proportionality calculation that I will briefly mention.

As I argued in section 3, C1 uses C2 as a mere means in furtherance of C1's evil plans by hacking C2's aims. C2 has a (defeasible) agent-centered prerogative against being used as a mere means in general, and being morally coerced specifically. Having this prerogative means that, in weighing the disvalue of acceding to C1's moral coercion against the disvalue of resisting it, C2 can augment the latter — within limits. If the disvalue of the two is sufficiently close, the agent-centered prerogative can tip the balance, thereby permitting C2 to refrain from doing what would make things go impersonally best.

Note that if an individual is faced with a morally adventitious dilemma, as in ALLEY 2, there would be no similar agent-centered prerogative to refrain from committing the lesser harm, since in doing so the agent is not being morally coerced.

## 6. Conclusions

There are three morals I wish to draw from my discussion of the wrongness of moral coercion, the Liability Question, and the Badness Question.

A) If my characterization of moral coercion is correct, then a terrorist who uses hostages as coin for concessions is engaged in fundamentally the same activity as a war criminal who straps innocents to the side of his tank. They are both engaged in moral coercion. We are repulsed by the way our moral commitments are plied in these sorts of cases. The account I have laid out provides a basis for this intuition. C1 hacks C2's aims — she intentionally puts C2 in a position where her moral commitments become self-undermining, in that C2 worse-achieves the object of those commitments *because* she has those commitments. Putting C2 in a position where her moral commitments are used in the service of the very ends that they are tasked with avoiding impermissibly uses the agent when it is done in furtherance of ends that wrong the agent.

B) Should C2 accede to C1's demands, she can be liable to defensive or compensatory harms, even if C2 was morally obligated to accede. Whether she is liable depends on whether the defensive or compensatory harms can be imposed on C1 instead, since C1, unlike C2, is responsible for imposing the dilemma in the first place. Still, for C1 to wrong C2 by morally coercing her does not in itself diminish the degree of responsibility that C2 bears for the harm in which $\phi$ consists. Though being morally coerced can serve as grounds for an agent-centered prerogative to refrain from acceding, C2 is no less responsible than she would be if she responded analogously to a functionally equivalent adventitiously imposed dilemma. Still, we need to account for the robust intuition that a harm resulting from a morally coercive dilemma is *worse* than a harm resulting from an analogous morally adventitious dilemma. In cases of moral coercion, C1's intentions are manifest in C2's actions, even though the latter is not intentionally cooperating with the former. The result is that the harm ought to be weighed as heavily as an intentional harm in C2's proportionality calculation, even if she committed the harm collaterally rather than intentionally.

C) One of the goals of this paper is to articulate the intuition that by acceding to moral manipulation we "allow evil to succeed". If what I have adumbrated in A and B is correct, there are two ways acceding allows evil to succeed. For C2 to accede to moral coercion allows her to be used in the way I have described in section 3. But this cannot exhaust the sense in which compliance with C1's wishes allows evil to succeed. This is because the duty we have not to be used

as a mere means is fundamentally a *self-regarding* duty. C1, though, wrongs not just C2, but the third-party innocents as well. Any characterization of how compliance allows evil to succeed should capture this fact as well. Suppose that an uninvolved third party confronts C2 and reminds her that if she does as C1 wishes, C2 will have thereby allowed evil to succeed. And suppose C2 responds by saying, "Don't worry — I don't mind being so used." Such a response intuitively misses an additional sense in which compliance allows evil to succeed. And the argument I presented in section 5 explains this: compliance allows evil to succeed not only in that it wrongfully uses C2, but also in that it allows C1's wrongful intentions to come to fruition — viz., the goal of harming the third-party innocents. So there are two (instrumentally related) evils here that we allow to succeed when we comply with C1's wishes: one evil is self-directed, and the other is third-party directed.

This account is but a first step in the moral analysis of moral coercion. I hope here to have provided a useful foundation for further discussion.

**Works Cited**


Bazargan, S. (2013). Complicitous Liability in War. *Philosophical Studies, 165* (1), 177–195.

Berman, M. N. (2011). Blackmail. In J. Deigh & D. Dolinko (Eds.), *The Oxford Handbook of Philosophy of Criminal Law* (pp. 37—106). New York: Oxford University Press.

Broad, C. D. (1930). *The Philosophy of C.D. Broad*. Open Court Publishing Co.

Davis, N. (1980). The Priority of Avoiding Harm. In B. Steinbock (Ed.), *Killing and Letting Die* (pp. 172—214). Englewood Cliffs, New Jersey: Prentice Hall, Inc.





Dreier, J. (1993). Structures of Normative Theories. *The Monist, 76* (1), 22—40.

Epstein, R. A. (1983). Blackmail, Inc. *The University of Chicago Law Review,* 50, 553–566.

Feinberg, J. (1986). *Harm to Self*. New York: Oxford University Press.

Frankfurt, H. (1973). Coercion and Moral Responsibility. In T. Honderich (Ed.), *Essays on Freedom of Action* (pp. 65—86). London: Routledge & Kegan Paul.

Haksar, V. (1976). Coercive Proposals. *Political Theory, 4* (1), 65—79.

Lindgren, J. (1984). Unraveling the Paradox of Blackmail. *Columbia Law Review, 84* (3), 670—717.

Louise, J. (2004). Relativity of Value and the Consequentialist Umbrella. *The Philosophical Quarterly, 54* (217), 518—536.

McConnell, T. C. (1981). Moral Blackmail. *Ethics, 91* (4), 544—567.

McMahan, J. (2009). *Killing in War*. New York: Oxford University Press.

McNaughton, D., & Rawling, P. (1995). Value and Agent-Relative Reasons. *Utilitas, 7* (1), 31—47.

Nagel, T. (1986). *The View From Nowhere.* New York: Oxford University Press.

Nagel, T. (1972). War and Massacre. *Philosophy & Public Affairs, 1* (2), 123—144.

Nozick, R. (1974). *Anarchy, State, and Utopia.* New York: Basic Books.

Nozick, R. (1969). Coercion. In S. Morgenbesser, P. Suppes, & M. White (Eds.), *Philosophy, Science, and Method: Essays in Honor of Ernest Nagel* (pp. 440—472). New York: St. Martin's Press.

Pallikkathayil, J. (2011). The Possibility of Choice: Three Accounts of the Problem with Coercion. *Philosophers' Imprint, 11* (16), 1—20.

Portmore, D. W. (2007). Consequentializing Moral Theories. *Pacific Philosophical Quarterly, 88* (1), 39–73.

Railton, P. (1984). Alienation, Consequentialism, and the Demands of Morality. *Philosophy & Public Affairs, 13* (2), 134–171.

Rodin, D. (2012). Justifying Harm. *Ethics, 122* (1), 74–110.

Sachs, B. (2013). Why Coercion is Wrong When it's Wrong. *Australasian Journal of Philosophy, 91* (1), 63–82.

Schopp, R. F. (1998). *Justification Defenses and Just Convictions.* Cambridge: Cambridge University Press.

Schroeder, M. (2007). Teleology, Agent-Relative Value, and 'Good'. *Ethics, 117* (2), 265–295.

Shaw, J. R. (2012). The Morality of Blackmail. *Philosophy & Public Affairs, 40* (3), 165–196.

Sussman, D. (2005). What's Wrong with Torture? *Philosophy & Public Affairs, 33* (1), 1–33.

Wellman, C. H. (2005). *A Theory of Secession: The Case for Political Self-Determination.* Cambridge: Cambridge University Press.

Wertheimer, A. (1987). *Coercion*. Princeton, New Jersey: Princeton University Press.