

## Negation and Abstention: Two Theories of Allowing

### By Jonathan Bennett

#### 1. Introduction

The tree fell, in consequence of your behavior: whether it fell or not depended on how you acted. There are two ways for that to be so. You could have felled the tree, made it fall, caused it to fall; or you could have allowed it to fall, not prevented or saved it from falling. This difference between making and allowing looms large in moral philosophy, with many urgent problems depending upon it. We need a clear understanding of what distinction it is. Here are some of the locutions through which we express it:

- |  |  |
|--|--|
| •she fells the tree                                      | •she lets the tree fall                                |
| •she causes it to fall, or makes it fall                 | •she allows it to be the case that the tree falls      |
| •it falls because of something she does                  | •it falls because of something she doesn't do          |
| •it falls because she intervenes in the course of nature | •it falls because she allows nature to take its course |

Those will do to go on with; there are others. The consequential states of affairs are endless: she cures the patient (lets him recover), she spoils the cheese (lets it deteriorate), she compacts the earth (lets it settle), and so on. On the allowing side of the line the consequential state of affairs is always explicitly declared—either in a whole nested sentence ('allows it to be the case that the tree falls') or in a nouninfinitive transform of that ('allows the tree to fall' or 'lets the tree

fall'). In reports on makings, we can also explicitly state the consequential state of affairs, but in practice we usually do not. By far the most common form of 'making' report is with an active verb phrase that has the consequential state of affairs buried in its meaning: she fells the tree, they rescued him, you ruined me, we amused them, and so on.

My guess is that these two lists of expressions are separated by a single distinction which I call 'making/allowing'. I do not mean to give a privilege to the words 'make' and 'allow': we shan't get to the bottom of this distinction by tracing out the precise meanings of any pair of verbs. In trying to analyze the making/allowing distinction I shall be looking for the longest clean, hard thread running through my two lists of locutions.

The important question for moral theory is whether the distinction between making and allowing has basic moral significance. That is to ask:

If someone's behavior has a bad state of affairs as a consequence is the morality of his conduct affected by whether he made the consequence obtain or only allowed it to do so?

The answer 'no' means that the difference between making and allowing is morally neutral; I shall call this 'the neutrality thesis'. The question arises in first- and second-order morality (I here borrow Donagan's terminology); it concerns which sort of conduct is worse and which justifies the greater moral indignation.

Some moral philosophers have argued for the neutrality thesis, and others have argued against it, by comparing pairs of cases and inviting us to agree with their moral intuitions

about them. There are reasons for doubting whether that approach will get us far,<sup>1</sup> and this paper will concern a different one, namely, trying to get clear about the moral significance of making/allowing by first getting deeply clear about what distinction it is.

Discontent with an analysis which I offered in 1966 triggered a spate of attempts to do better.<sup>2</sup> Two analyses, in my opinion, show more promise than all the rest. Indeed, the others look best when they are seen as attempts to formulate one of those two. One is my old one, which I do not concede is discredited by the single argument that is always brought against it; the other is an analysis presented in Alan Donagan's *The Theory of Morality*. I shall present and compare Donagan's and my attempts to analyze the making/allowing distinction.

## 2. Positive/negative

I need these two short stories, in each of which a vehicle stands on ground that slopes down to a cliff top:

**Push:** The vehicle stands, unbraked, on the slope; Agent pushes it, and it rolls to its destruction.

**Stayback:** The vehicle is already rolling; Agent could but does not interpose a rock which could stop it, and the vehicle rolls to its destruction.

In each case we have a person, a time, and a vehicle's smashing at the foot of a cliff. It is uncontroversial that in Push the relevance of conduct to upshot belongs on the left (positive, active-verb, 'making') side of the line, and that in Stayback it belongs on the right (negative, passive, 'allowing')

side. An analysis of making/allowing should at least get these two right and make clear how it does so.

My analysis identifies making/allowing with something I call the positive/negative distinction. This does not distinguish two kinds of action: there are no negative actions. I use 'negative' primarily in the phrase 'negative proposition' and thus also in 'negative fact' (facts are true propositions). In Stayback, for instance, Agent moves his fingers, he makes a daisy chain, he smiles; also, he does not resign his job, he does not stand on his head, he does not interpose the rock; three positive facts about how he behaved and three negative ones.

Given 'positive' and 'negative' as predicates of facts or propositions, we can define the relations 'is positively relevant to' and 'is negatively relevant to' between behavior and upshots. These definitions are the first of the three stages in my analysis. In giving it, I use 'T<sub>1</sub>' to name the moment when Agent pushes the vehicle in Push, and the last moment when he could usefully have interposed the rock in Stayback.

(1) Suppose we want a full T<sub>1</sub>-dated explanation of the vehicle's fate in Stayback—that is, facts about how the world is at T<sub>1</sub> from which it will follow causally that the vehicle is destroyed at T<sub>2</sub>. Let E be the environmental portion of this material: it is the proposition about Agent's environment that is needed for the causal explanation that we seek. What must we add to E to complete the explanation? That is, what is the weakest proposition A about Agent's conduct at T<sub>1</sub> such that (E & A) causally imply that the vehicle is destroyed at T<sub>2</sub>? The answer is that we need A = the negative fact that Agent does not interpose the rock. There are positive facts

<sup>1</sup> 1. For discussion, see Jonathan Bennett, 'Positive and Negative Relevance', *American Philosophical Quarterly* 20 (1983): 185–94.

<sup>2</sup> Jonathan Bennett, 'Whatever the Consequences', *Analysis* 26 (1965–66): 83–102. A greatly improved version of the analysis is presented in Jonathan Bennett, 'Killing and Letting Die', lecture 1 (pp. 47–72) in 'Morality and Consequences', in *The Tanner Lectures on Human Values*, ed. S. McMurrin (Salt Lake City: University of Utah Press, 1981), vol. 2, pp. 47–116.

about how he behaves at  $T_1$ , but none fits the description ‘weakest fact that, when added to E, yields a complete causal explanation of the disaster’. The fact that Agent smiled at  $T_1$  does not yield a complete causal explanation; the fact that he turned and walked away does yield one, but it is not the weakest fact that does so. The negative fact that Agent does not interpose the rock is exactly what is needed: it is just strong enough to complete the explanation.

Now look for a  $T_1$ -dated explanation for the disaster in Push. Here there are countless negative facts about how Agent behaves at  $T_1$ , but none is strong enough to complete the explanation. For that we need the positive fact that he pushes the vehicle hard enough to start it moving.

In short: Agent’s behavior in Push is positively relevant to the destruction of the vehicle. That is, the weakest fact about his conduct that suffices to complete a causal explanation of the vehicle’s being destroyed is positive. Agent’s conduct in Stayback is negatively relevant to the disaster. That is, the weakest fact about the conduct that suffices to complete a causal explanation of the vehicle’s being destroyed is negative.

(2) What is it for a fact or proposition to be negative? We can call a *sentence* ‘negative’ if it contains an odd number of negating expressions, but that is not useful. The ‘negative proposition’ problem is hard, and Frege suspected that it cannot be solved.<sup>3</sup> I know of only two lines of solution that have been tried: one is due to Chisholm, and I don’t accept it because it assumes that a negative proposition, like a negative sentence, contains a negating part and a negated part, and I don’t believe this.<sup>4</sup> The other has occurred to philosophers as disparate as Berkeley, Kant, and Ayer; and I

shall work with it.

I shall present it through Venn diagrams. Let all the ways Agent could move at time  $T_1$ —including staying still—be represented by a square. Each point on the square represents a proposition attributing to him some absolutely specific way of moving. A region of the square represents the disjunction of the propositions represented by the points in the region. Start, for example, with the proposition that *He walks fairly slowly northwards*. Take every point proposition—every absolutely specific way of moving—that would make it true that Agent walks fairly slowly northwards; identify the points that represent those; then the region of the square that represents our original proposition is the region that contains just exactly those points.

A line across this square represents a pair of propositions which are complementary within the square. They are not strictly contradictories because each entails the existence of Agent at that time. I offer to define propositional negativeness only with respect to that framework.

I propose this: A proposition about how Agent moves at  $T_1$  is negative if it is fit to be represented by a region that covers nearly the whole of the space of possibilities for him at that time. That is, a proposition and its complement are positive and negative, respectively, if they divide the relevant space extremely unevenly; the highly informative one is positive, and its almost empty complement is negative. This is the idea that is at work in Kant’s account of why the judgment ‘The soul is non-mortal’, which has the positive form ‘S is P’, is negative in content. Its predicate ‘non-mortal’ is negative, Kant says, because it picks out only ‘the unlimited sphere of non-mortal beings’:

<sup>3</sup> Gottlob Frege, ‘Negation’, in *The Philosophical Writings of Gottlob Frege*, ed. P. Geach and M. Black (Oxford: Blackwell, 1952), pp. 117–35, at pp. 125–26.

<sup>4</sup> Roderick M. Chisholm, *The First Person* (Minneapolis: University of Minnesota Press, 1981), pp. 124–25.

'Nothing more is said by [this] proposition than that the soul is one of the infinitely many things which remain when I take away all that is mortal. The infinite sphere of all that is possible is thereby limited only to the extent that the mortal is excluded from it, and that the soul is located in the remaining part of its space. But, even allowing for such exclusion, this space still remains infinite, and several more parts of it may be taken away without the concept of the soul being thereby... determined in an affirmative manner.'<sup>5</sup>

This explains the familiar fact that negative commands are in general easier to obey than positive ones, negative plans easier to execute, and negative hopes easier to realize. To obey a positive command or execute a positive plan (become a teacher, go to Sri Lanka), you have to actualize some one of a relatively small range of possibilities, and it may be that each is in some way costly—or that most are and it would be hard to identify any that are not. On the other hand, to execute a negative plan (don't become a teacher, keep out of Sri Lanka), one need only behave in a way that actualizes one of an enormous range of possibilities, a range so large that the odds are that there are easily findable costless ones among them.

### 3. A metric for the possibility space

We need a metric for the relevant space of possibilities—a basis for saying that one proposition is consistent with more possible states of affairs than another. This is easy when one entails the other, but otherwise it is hard. I need a basis for saying, of two complementary propositions, that

they divide up their total space of possibilities unevenly, and for saying which corresponds to the larger subspace. We cannot do this by counting the number of points in each, for there will usually be infinitely many points on each side of the line—for example, infinitely many ways of walking fairly slowly northwards and infinitely many ways of not doing so. Or so it seems. One might stave off this threat of infinity, but I prefer to acknowledge the threat and steer around it.

Here is how. Two propositions about how Agent moves at a particular time are to be accorded the same amount of the possibility space if they are equally specific. This will not work with every determinable, for example, with colors, because for them we have no agreed objective measure of specificity; but we have such measures for space and time, and thus for movement and for specificity of propositions about movement. This puts us in a position to say that a given complementary pair divides the possibilities very unevenly, so that one is positive and the other negative. For example, *He walks northwards* and *He does not walk northwards* are positive and negative, respectively, because the former is much more informative, much more specific, than the latter. The fact that one is and the other is not expressed with help from the word 'not' has nothing to do with it.

I do not claim that any two propositions about how Agent moves at a time can be compared for specificity. If P1 describes with great precision the trajectory of his right hand, while P2 gives a somewhat vague account of how he moves his whole body, there is no determinate answer to the question of which is the more specific. However, as I explain how the parts of my analysis fit together, it will be seen that this does not matter.

<sup>5</sup> Immanuel Kant. *Critique of Pure Reason* (London: Macmillan, 1929). A 72f = B 97f. See also A. J. Ayer, 'Negation', in his *Philosophical Essays* (London: Macmillan, 1954) pages 36–65.

If someone's behavior is (in my sense) relevant to upshot U, the space of his possible ways of moving can be divided into those that satisfy:

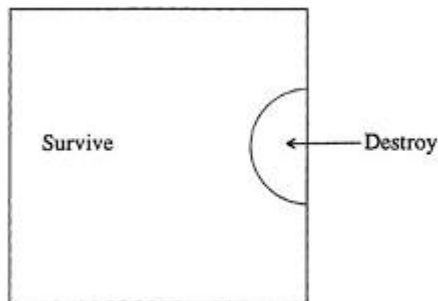
- If he moved like that, U would obtain,

and those that satisfy:

- If he moved like that, U would not obtain.

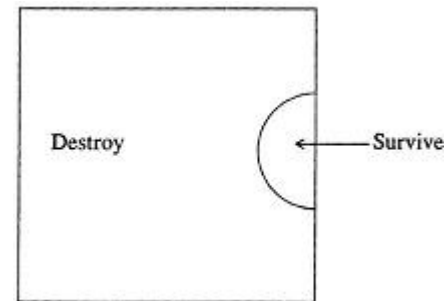
The line between these is what I call 'the U-line'. For example, the disaster line through Agent's space of possible conduct has all the vehicle-smashes movements on one side and all the vehicle-survives ones on the other. (For simplicity's sake, I am pretending determinism, so as to get rid of probabilities between 0 and 1.) Now, what makes Agent's conduct positively relevant to the disaster in Push is the fact that the disaster line separating his vehicle-is-destroyed options from his vehicle-survives ones is like that in figure 1.

Figure 1:



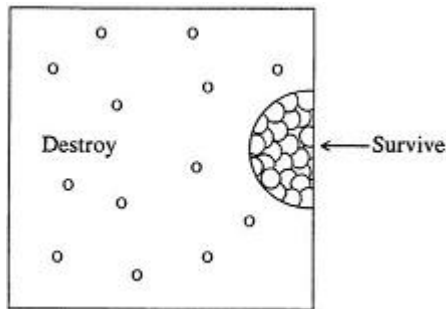
That is, of all the ways in which he could have moved, only a tiny proportion were such as to lead to the vehicle's destruction; virtually all would have had its survival as a consequence. In Stayback, on the other hand, Agent's conduct is relevant to the disaster in a negative way, represented by figure 2.

Figure 2:



To see why this is right about Stayback, consider the proposition that Agent interposes the rock, and think about the different physical ways he could have done this: a few dozen pairwise contrary propositions would pretty well cover the possibilities, each identifying one fairly specific sort of movement which would get the rock into the vehicle's path. Thus, Survive in figure 2 can be divided up into a few dozen smaller regions, each representing some kind of push or kick or the like. Now, each of those can be paired off with an 'echo' of it in Destroy—that is, with a proposition which has the same amount of content as it and is indeed very like it except that its truth would not rescue the vehicle. For instance, if Survive contains a proposition attributing to Agent a certain kind of movement with his left foot, let its 'echo' attribute to him a similar movement of that foot but with the direction differing so that the foot misses the rock. In general, for each little proposition in Survive, let its echo be one whose truth would make it look as though Agent were trying but failing to interpose the rock. (The 'echo' propositions must be pairwise contraries so that their regions don't overlap.) The general idea is given by figure 3.

Figure 3:



The pockmarks in Destroy represent the echoes. My ‘degree of specificity’ criterion secures that their combined area is the same as that of Survive; and clearly they take up only a tiny proportion of Destroy. I do not base that on how the circles are drawn, but on considerations of specificity. Each of the little propositions in Survive has many echoes in Destroy; we assign just one echo to each, and all the remaining echoes take up further space in Destroy. That region also contains countless propositions that do not echo anything in Survive—that is, countless ways in which Agent could move without looking as though he was trying to interpose the rock—and millions of those will also be comparable for specificity with the echoes. The result is that, according to my criterion, the space of vehicle-rescuing movements that Agent might have made in Stayback is enormously much smaller than the space of vehicle-is-destroyed movements that he could have made; which is to say that figure 2 is correct for Stayback except that it understates the disparity in size between the two regions. A similar line of thought, *mutatis mutandis*, can be applied to Push and figure 1.

That is why Agent’s conduct in Push is positively relevant to the vehicle’s being destroyed, and why in Stayback it is negatively relevant to this. In Stayback, the truth of any proposition in Destroy would suffice to complete the causal explanation for the disaster, but relevance is defined in terms of the weakest proposition that would suffice for this; that is the disjunction of all the propositions that would suffice, that is, the proposition represented by the whole of Destroy. I hope it is clear that occurrences of the word ‘not’ play no part in this line of thought.

#### 4. The immobility objection

My account implies that *He moves* is negative and that *He does not move* is positive. In the latter claim I have Locke on my side. He was opposed in this by Leibniz but only with a bad argument.<sup>6</sup> The question does not matter enough to discuss here. I call the distinction that I am presenting ‘positive/negative’, I freely use those words in expounding it, and I think this is reasonable. If not, and that is the only trouble I am in, I shall drop those two words and retain the distinction. What really matters is whether the line I have drawn is plausible as an analysis of the making/allowing distinction which is common property.

The status of *He moves* comes into that question too, but not through a sterile debate about whether it is a negative proposition. To see how, consider the following story:

Henry is in a sealed room where there is fine metallic dust suspended in the air. If he keeps stock still for two minutes, some dust will settle in such a way as to close a tiny electric circuit which will lead to some notable upshot U. Thus, any movement from Henry,

<sup>6</sup> John Locke, *An Essay concerning Human Understanding* II.viii.6; G. W. Leibniz, *New Essays on Human Understanding*, trans. and ed. P. Remnant and J. Bennett (Cambridge: Cambridge University Press, 1981), p. 130.

and U will not obtain; perfect immobility, and we shall get U.

My analysis says that if Henry keeps still he makes U obtain, whereas if he moves he allows U not to obtain. This story is my version of a kind of example that has repeatedly been brought against my analysis, by people who are sure that in this situation if Henry keeps still the relation of his conduct to the upshot U is of the allowing rather than the making kind—belongs on the right rather than the left of the line we are inquiring into.

I do not dispute that if Henry keeps still he allows the dust to fall; but that is an example of how the detailed meaning of the word ‘allow’ is a poor guide in our present problem area. It would lead us astray about Henry as it would over the following variant on my vehicle stories:

**Kick:** In the path of the moving vehicle there is a rock which could bring it to a halt; Agent kicks the rock away, and the vehicle rolls to its destruction.

In Kick the relation of conduct to upshot obviously belongs on the left of the line: this is a positive, active, making. Yet it is proper and normal to say that by kicking the rock away Agent *allows* the vehicle to roll on down the slope. In our present context, the niceties of the meaning of ‘allow’ are an *ignis fatuus*.

If that is not what lies behind the immobility objection, then what does? What else would make people so sure that if Henry keeps still his conduct relates to upshot U in a right-of-the-line or allowing way? Here is a guess about that. In most actual situations, relative to most values of U, immobility would belong on the roomy side of the U line. When in Stayback Agent does not interpose the rock, he may do something else instead, but his staying still would have had the same effect on the vehicle’s fate. It takes work to rig up a situation where stillness is almost the only route to

some interesting upshot. Perhaps that is why, when people confront a result that is produced by Agent’s immobility, they immediately and invalidly infer that this is a case of allowing.

To evoke intuitions on my side, suppose that it is almost impossible for Henry to be still for long, and he sweats and strains to do so because he wants U to obtain. If he succeeds, doesn’t this feel like a ‘making’? If he fails—gives up trying because the effort is too great, relaxes, and lets his body shift a little—doesn’t that feel like an ‘allowing’?

‘What would it take for you to accept defeat?’ someone asked me once. He could have continued like this: ‘Hardly anybody agrees with you that if Henry keeps still he makes U obtain; you can try to talk them around, but even if you succeed that will be by changing how people think about making and allowing. Considered as an analysis of the making/allowing distinction that we actually have, yours is wrong. It would be more graceful if you were to admit it.’ This deserves an answer.

How people use the words ‘make’ and ‘allow’ is a pointer to my concerns, but it does not define them. Let us distinguish three ways things might stand. **(1)** The ordinary uses of ‘make’ and ‘allow’, and of the other expressions in the two lists at the start of this paper, reflect an underlying jumble with no systematic core. **(2)** People use those locutions to express a single clean, systematic distinction, and I have failed to describe it. **(3)** In their uses of the locutions in question, people are guided by a clean, deep concept, but only imperfectly, because they sometimes drift away from it and use the terminology of making/allowing in ways that have no solid conceptual support.

If **(1)** is true, then there is no such thing as ‘the making/allowing distinction’, and my project is doomed. I cannot prove that this is wrong, but it is implausible, and I have seen

no strong case for it. The friends of the immobility objection do not say ‘You purport to describe a distinction that doesn’t exist’; rather, they hold that the distinction exists and that I have misdescribed it. That is to assert **(2)**. I cannot prove that it is wrong either, but I have seen no convincing evidence that it is right: an intuition that goes against my analysis is not the same as a rival to it. Of the rivals that have been produced, by far the most considerable is Donagan’s, which I shall discuss shortly. It conspicuously does not make a big issue of immobility as such, so that it sides with me rather than with those who bring the immobility objection. I conclude that the facts fit best with **(3)**: our thought and talk about how conduct relates to its upshots do reflect a decently grounded making/allowing distinction, from which we sometimes drift away. It is in that spirit that I stand my ground in face of dissent about the case of Henry.

## 5. Doubts about the metric

If my analysis is wrong in some matter of detail, I think I know what that must be. The use of ‘negative fact’ to define ‘negatively relevant’ seems secure, as does the explanation of negativeness of facts or propositions in terms of an uneven division of a possibility space. My way of measuring that space is vulnerable, however: I could be using the wrong ruler.

Of two ways of criticizing it, the more radical says that ‘degree of specificity’ has no place in this story. Alternatives to it suggest themselves. One is that the amount of space a proposition occupies is inversely proportional to its probability; another—better looking—would make it inversely proportional to how difficult it would be to make that proposition true. I have not been able to make either of those work out satisfactorily.

The truth may involve a mixture of metrics. When we classify facts about behavior as positive or negative, we may be steering by something that involves specificity and probability and difficulty, and who knows what else. Or perhaps no one metric is right, and we carve up the space varyingly according to context. If that is so, it might be that the making/allowing distinction has moral significance sometimes but not always. All of this is idle speculation at this point; I cannot find any way of making it real.

The less radical criticism allows that specificity is crucial but rejects the framework within which I use it. It may seem perverse to define the metric for a space of possible ways of behaving in terms of propositions about ways of moving. That is indeed a defect, because an item of behavior might basically or intrinsically consist not in moving but rather in directing one’s thought in a certain way. In practice this matters little, because such nonmovement behavior plays such a small part in our lives; but this gap in the account is theoretically bad; and I do not know how to fill it.

My metric does cover all the rest of human behavior, but not in the terms that interest us: it attends to hand gestures, arm swings, foot shoves, and vibrations of the larynx rather than to kinds of behavior that make up the stuff of the moral life: giving up, betraying him, keeping the faith with her, sharing, hoarding, and so on. These, I repeat, are all covered. To say that he kept the faith with her is to attribute a certain relational property to how he moved; similarly with ‘He gave up the project’ and ‘She shared her food with them’ and all the rest. Still, it’s a suspicious fact about the metric that it does not use any of the relational concepts that give importance to human conduct.

There may be a good deal hanging on this. The positive/negative distinction that I have defined obviously has no basic moral significance: if someone moves in a way



that causes or makes probable some bad upshot, the moral status of his conduct obviously cannot depend on how many other movements by him would have done the same. The obviousness of that, though, depends on the metric I have chosen; if we switched to a different way of measuring the possibility space, the claim might be less compelling.

There are various loose, intuitive ways of revising the metric. Consider these three stories:

**Suit:** An African village is in need. I launch a lawsuit that deprives them of a thousand dollars they would otherwise have had.

**Cancel:** Same village, same everything, but this time I learn that my accountant thinks he is supposed to sign away a thousand of my dollars to the village, and I tell him not to.

**Nohelp:** Same village, etc. This time I could but do not give the villagers a thousand dollars.

My positive/negative line has Suit and Cancel on the left and Nohelp on the right. One might hope for a distinction that would put Suit on the left and bracket Cancel with Nohelp on the right: Cancel and Nohelp, one might think, are just two ways of not giving money to the village, and it does not matter that one does and the other does not require a fairly specific kind of movement. This presupposes a possibility space that represents not ways I could move but rather things I could do with my money, with these (perhaps) dividing into such equal-sized kinds as investing it in bonds, giving it to my children, spending it on a swimming pool for myself, and giving it to that African village. That sounds natural, but it is tailored to fit this trio of cases; it is no use until we can generalize it, and I cannot see how to do that.

That brings me to the end of what I have to say about the positive/negative distinction, except in comparing it with Donagan's distinction, to which I now turn.

## 6. Agency and 'the course of nature'

A person's agency can relate to a state of affairs in any of three ways. *Quiescence:* Agency is not involved. Her head moved because a brick hit it. *Intervention:* The person exercises her agency so as to make a difference to what happens. Her head moved because she nodded. *Abstention:* The person exercises her agency so as to make no difference to what happens. Her head moved because, feeling the onset of a suppressible sneeze, she decided to let it happen.

These can be described using Donagan's phrase 'the course of nature'. In the first there is no agency, so the course of nature is followed. In the second, agency makes a difference, so the course of nature is not followed. In the third, the person exercises her agency in such a way that the course of nature is followed. Underlying all this is the principle:

A train of events occurs in the course of nature if (and only if) it would have occurred if human agency had not been at work.

So the course of nature may be followed because of the person's quiescence or her abstention (from intervening). Here is part of what Donagan says about all this:

Should she be deprived of all power of action, the situation, including her own bodily and mental states, would change according to the laws of nature. Her deeds as an agent are either interventions in that natural process or abstentions from interventions. When she intervenes, she [causes] whatever would not have occurred had she abstained; and when she abstains, [she allows] to happen whatever would not have happened had she intervened. Hence, from the point of view of action, the situation is conceived as passive, and the agent, qua agent, as external to it.

She is like a *deus ex machina* whose interventions make a difference to what otherwise would naturally come about without them.<sup>7</sup>

There are some signs that Donagan regarded human agency as unnatural, for example, as able to defeat the laws of nature. Few of us would accept that, and perhaps Donagan didn't either: I choose to interpret him not as advancing any substantive thesis about some kind of antithesis between agency and nature but, rather, as defining 'the course of nature' or 'what naturally happens' as something like 'what happens independently of human agency'.

Independently of all human agency? If we say that what happens in the course of nature is what comes about without any human agency anywhere in its causal history, we shall have a basis for a possible moral idea, namely, piety about what is 'natural' in this strong sense—the world untouched by man. We would have to soften it, perhaps turning it into a matter of degree, because most significant aspects of our planet already bear the marks of human intervention. I shan't follow this up here; it is not a promising approach.

We'll do better if we attend to human agency during a certain interval, and it is easy to see what one it should be. We are relating behavior to an upshot, so we should attend to human agency during the interval bounded by that conduct and that upshot. So the 'course of nature', relative to a given conduct-upshot pair, is what comes about unaffected by agency during that interval.

Whose agency? That of the person whose conduct is in question, of course; but what about others? Suppose that you start a fire and that I could have stopped you. Clearly I did not make the fire start. Donagan says that I did not allow it to start either, because it didn't come about

in the course of nature (pp. 50–51). For him, the course of nature is what is unaffected by any human agency during the conduct-to-upshot interval.

For a reason that I do not have space for here, it is not absurd or arbitrary to handle 'allowing' in this way; but I shan't follow Donagan in it, because it runs together two questions that are better taken separately:

- (i) U comes about through my intervening, or U would have happened even if I had abstained. Does it morally matter which?
- (ii) When U comes about because I do not intervene and stop it, it does so either through someone else's intervention or through processes not involving agency. Does it morally matter which?

I am inclined to answer no and no; Donagan, I think, would answer yes and yes. The questions are different, though, and it is better not to run them together as Donagan does. His answer to (i) is 'Yes: if U came about because you intervened, it is not something you allowed to happen.' His answer to (ii) is 'Yes: if U came about through the agency of someone else, it is not something you allowed to happen.' It is better to address these questions separately, not letting 'allow' suggest that they have something real in common.

I therefore choose to understand Donagan's distinction like this: What happens in the course of nature relative to Agent is what would have happened if he had not intervened. When I speak of 'Donagan's distinction', I always mean this one that I have developed out of his materials.

Donagan sometimes seems to imply that what you allow to happen is what happens because you abstain from intervening in the course of nature (p. 50). That cannot be right, because we allow countless things to happen

<sup>7</sup> Alan Donagan, *The Theory of Morality* (Chicago: University of Chicago Press, pages 42–43, lightly edited).

whereas abstaining from intervening is a strange and rare performance: exercising one's agency so that the rest of the world runs just as it would have done if one had been asleep or on autopilot. In Stayback, for example, where Agent allows the vehicle to go over the cliff, he is probably shuffling his feet or waving his arms or making a daisy chain or somehow intervening in the course of nature.

Donagan presumably made that mistake because he slid into thinking that the crucial question is a monadic one about the nature of the person's conduct: Did he abstain from intervening in the course of nature? Really, it is a dyadic one about how the conduct relates to a specific upshot: Would U have come about if he had abstained from intervening in the course of nature? That points us to the best way to understand the analysis. A person allows U to obtain if (i) the person could have behaved so that U didn't obtain and (ii) U would have obtained if the person had abstained, which will be true just in case U would have obtained if the person's agency had not been exercised at all. The analysis, in short, divides the consequences of a person's behavior according to whether they came about because of some difference that the person's agency made to the course of events. This involves counterfactuals about what would have happened if the person's agency had not been at work, and these are sometimes problematic: some problems beset virtually all counterfactuals, and others come from our imperfect grasp of the concept of agency. Still, we can apply Donagan's active/passive distinction (as I call it) in enough cases to be able to assess it and compare it with my positive/negative one.

## 7. How far do the two distinctions coincide?

Construct again the possibility space for Agent at T, and draw the U line through it, dividing it into the U region, each of whose points satisfies this:

- If he had behaved in that way, U *would* have ensued, and the non-U region, each of whose points satisfies this:
- If he had behaved in that way, U *would not* have ensued.

My positive/negative distinction asks: Is the U region vastly larger than the non-U one? Donagan's asks instead: Does the U region contain the abstention point? That is to ask: Would U have ensued even if Agent had abstained from intervening in the course of nature?

So there is a single structure, with one question being asked about it by the positive/negative distinction and another by Donagan's active/passive distinction. Different as they are, however, the questions usually have the same answer because it is usually the case that if U would have occurred even if the person's agency had not been exercised, then also whatever the person did (with a few exceptions), U would have occurred, and vice versa. Still, it isn't hard to devise cases where the two come apart, and I shall present some shortly, starting with cases that involve the vexed matter of immobility.

Donagan's active/passive distinction makes no more of immobility or unchangingness than mine does. In most possibility spaces, relative to most upshots, the immobility point will be in the region that contains the abstention point, and thus also in the larger of the two regions; which is 'why, in most cases where U would have ensued if the person had remained still, his conduct's relevance to U will be on the right of Donagan's line and of mine. But neither line employs the concept of stasis: neither asks what would

have happened if the person had remained still. Donagan is explicit about this. In the natural course of events, he says in the passage I quoted early in section 6 above, the person's 'bodily and mental states change according to the laws of nature'—and the changes could include movements. Thus, two considerable analytic treatments of our topic are silent about immobility. If they are both wrong, and immobility deserves a special emphasis, I would like to hear the reasons for this.

We encountered the immobility objection in a case where U will occur if Henry moves and not otherwise. Now consider two ways the story could unfold:

- (i) Henry sits completely still, *slack and comfortable*; U ensues and would not have done so if Henry had moved.
- (ii) Henry sits completely still, *heroically enduring a terrible itch*; U ensues and would not have done so if Henry had moved.

The positive/negative distinction ignores the emphasized phrases and classifies Henry's conduct as positively relevant to U both times. But if Henry's itch means that in the course of nature he would move, so that his stillness is an intervention, then Donagan's analysis counts Henry's behavior as relevant passively in (i) and actively in (ii). Case (ii) is the one I appealed to when I first introduced Henry: I was trying to shift your intuitions about immobility my way, by bringing them up against a case where both analyses were against them. Mine without Donagan's is less powerful than are both together.

This situation is symmetrical, however: against the immobility intuition, Donagan's approach has less power on its own than when associated with mine. To see this, consider the following version of the Stayback scenario: At the time when he could be interposing the rock, Agent remains mo-

tionless; to do this, he has to suppress a muscular spasm which, if unchecked, would make his leg move in such a way as to shove the rock into the vehicle's path. This is still Stayback, so the relation of conduct to upshot is on the right or allowing side of my line; but it is on the left of Donagan's—for him this is a case of making the vehicle go to its destruction—because if Agent had not exercised his agency at all his leg would have jerked out, the rock would have been interposed, and the vehicle would have been safe. In this case, I think, those who are impressed with immobility will side with me rather than Donagan. So, as I said, the two analyses together exert more force against the immobility intuition than does either on its own.

This does not matter much, because the immobility intuition is significant only as a matter of epidemiology: There's a lot of it about. There are people out there who have strong, confident ideas about the import of someone's keeping still; but I have offered two ways of explaining these ideas away, and nobody has said anything in their defense. Until that situation changes, I don't think the immobility intuition deserves much respect.

## 8. The two distinctions as collaborators

Still, the symmetry that I have just uncovered goes further. Right across the board, it seems to me, one is happier applying the making/ allowing distinction to cases on which active/passive and positive/negative give the same answer. For a case where they do not, and where immobility is not involved, try a variant on Push. I described that as a case where a fairly specific kind of movement by Agent—call it a 'shove'—is needed to get the vehicle moving. Now add this detail: Agent feels the onset of an involuntary muscular spasm which, if not checked, will result in his body's making

a shove. He could quell this spasm, but he chooses not to and his body does produce the shove which starts the vehicle rolling. For Donagan this is a case of allowing the vehicle to roll: the relevance of conduct to upshot is on the right of his line. It remains on the left of mine. If you are perfectly content to classify this as a case of allowing the vehicle to be destroyed, then for you Donagan's distinction is correct. I suspect, however, that few people will be comfortable describing the case in that way; and that, I believe, is part of a general pattern—we confidently apply the notion of allowing only when the negativeness thought and the abstention thought both favor our doing so.

It is a trivial exercise to modify Stayback in a similar way, so that Donagan puts it on the left while I still have it on the right. I have indeed given one such modification, but now I am talking about one that does not involve immobility. Here again, I think that most people's intuitions about the case will be unsteady.

(Trying to resolve a crucial problem about how we understand ourselves, I have ended up talking about twitches and spasms! This may seem grotesque; but I had to, if I was to consider cases where the two distinctions diverge. If a particular conduct-upshot relation is to fall on the left of my line it must involve some fairly specific kind of movement; for it to fall on the right of Donagan's line, the movement must be one which would occur in the course of nature. Put the two together and you get a fairly specific kind of movement which would occur in the course of nature; which is why I bring in twitches and spasms. And since the behavior in question must involve agency, the twitches and spasms must be suppressible ones. Of course the 'kind of movement' could include complete stillness, but I have explained why we ought not to conduct the entire discussion in terms of that.)

I conclude that Donagan and I are both right. When people use the locutions that express the making/allowing distinction, they may be guided by the abstention thought or by the negative ness thought or by both at once, with neither being always uppermost. In the form in which most people have it, the making/allowing distinction does not equip them to deal with the odd cases where only one of those thoughts is available. In short, the two analyses somehow share the truth between them.

## 9. Formal contrasts between the two

This is an odder result than it might at first seem. The two distinctions are so unlike in their formal aspects that one would hardly expect them to collaborate, or to operate interchangeably, in our thinking. I shall sketch that contrast, which is (to me, anyway) too interesting to pass up. Assume throughout that a person behaves in a certain way which is relevant to an upshot U. We have the person's possibility space, with the U line drawn through it.

**(1)** Active/passive identifies one particular point on the person's possibility square, namely, the abstention point; positive/negative does not in that way pick out anyone absolutely specific way of behaving.

**(2)** Active/passive identifies that particular way of behaving by means of a monadic predicate: *... is an abstention from intervening in the course of nature*. All the other possible ways of behaving fall under another monadic description: *... is an intervention in the course of nature*. Positive/negative uses no monadic descriptions of items of behavior.

**(3)** Active/passive's central question about the U line is, On which side of it does the abstention point lie? Active/passive does not ask about the relative sizes of the two

regions, and the question it does ask must have an answer (vagueness apart). Positive/negative's question about the regions that are marked off by the U line is, Which, if either, is by far the larger? Positive/negative does not ask what movements each region contains, only how many. Also, if neither region is much larger than the other, positive/negative's question has no answer.

(4) Given that the person's behavior was relevant to U, active/passive can classify the relevance on the basis of information about either one of the regions that are separated by the U line. If active/passive is told that the U region contains the abstention point, or that it does not, or that the non-U region contains it, or that it does not, it can draw its conclusion. Not so with positive/negative. Even when it knows the whole truth about one of the regions, it cannot definitely classify the relevance until it looks at the other. One region, taken on its own, might be so limited as to suggest that it is the smaller of the two; but the person might be physically disabled so that the movements represented in that small region are nearly his whole repertoire. Or one region might be large enough to represent virtually the whole repertoire of a normal person, strongly suggesting that it is the larger of the two; but it is theoretically possible that the person's extraordinary powers give him a repertoire of which the possible behaviors we have considered are only a small fraction.

(5) I explained in section 2 above why it is apt to be easier to pursue a plan or obey a command if it is negative than if it is positive. That explanation depended on the size difference between two subregions of the possibility space, so it was special to positive/negative and could not work for active/passive. One might think: 'Active/passive has another route to the same conclusion. Whatever its size, the subregion in which the person allows U to happen

always contains the abstention point, which represents a costless way of behaving. It can't be much trouble to the person to slump into inactivity for a while, thus refraining from intervening in the course of nature in any way.' This implies that the two distinctions share a certain feature but for different reasons. That would be enjoyable, but it isn't right. Granted that it is physically easy to slump, abstaining altogether from intervening in nature's course, an abstention might be costly in other ways: painful, morally repugnant, mentally demanding, and—perhaps above all—a damaging departure from whatever project the person is currently engaged in. It is true that active/passive tends to resemble positive/negative in respect of cost, but the only reason I can find for that is that the two coincide along most of their length. That is, for most significant upshots U, U is reachable through abstention only if it is reachable through most things that the agent might do; so usually the negative (= larger) region marked off by the U line is also the one containing the abstention point; so usually the two distinctions coincide. To the extent that they do, active/passive can borrow a cost difference from positive/negative.

## 10. Other attempted analyses

In working through other attempts to analyze the making/allowing distinction, I have been struck by how many of them, when pressed for clarity and depth, do best by turning themselves into versions of Donagan's active/passive distinction. I shall give three examples; there are others.

It has been proposed that you allow P to obtain if **(i)** you could have acted so that P did not obtain and **(ii)** P would have obtained even if you had been absent from the scene. This does not draw the line in the right place, because

sometimes the person's being absent would have altered the whole structure of the situation. For example, it might be that the fool who started the car rolling in Stayback did so only because he saw Agent standing there; if Agent had been absent, the car would not have rolled to its destruction. Still, we can rescue this account by making it speak of what would have happened if the person had been absent *qua* agent, that is, if he had not exercised his agency; and that brings in Donagan's active/passive distinction.

The same idea may lie some distance behind Bentham's definition of 'negative acts' as 'such as consist in forbearing to move or exert oneself'.<sup>8</sup> The immobility aspect of this is indefensible, I have argued, but 'forbearing to exert oneself' is better. It might be a gesture toward this: negatively relevant conduct is behavior that bears on the given upshot in the same way as would the person's not exerting himself, that is, his agency's not being exercised.

Again, when philosophers offer to explain our distinction—as the late Warren Quinn did—in terms of 'the distinction between action and inaction', they have clearly gone wrong.<sup>9</sup> A person who allows something to happen need not be in any reasonable sense 'inactive' at the time. Perhaps these philosophers meant to be talking about what would have happened if the person had been inactive; in which case we again have rescue through reinterpretation and Donagan.

Some of the informal locutions that express making/allowing—and some of the analytic assaults on it—use the language of the contrast between positive and negative or between 'did' and 'didn't'. One of these, too, occurs in Quinn's paper where he writes that 'harmful negative agency' involves 'harm occurring because of. . . the noninstantiation of some kind of action that [the person]

might have performed'. Quinn used 'positive' and 'negative' as I do 'making' and 'allowing', merely to label the analysandum; evidently he thought he did not need these terms in his analysans and so did not need to get them clear. He did, though, as part of an account of what kinds of actions there are. If he set no limits to those, we could say that not pushing a vehicle is a kind of action; then in Push the disaster occurs because of the noninstantiation of that kind of action, which puts the Push scenario, absurdly, on the allowing side of the line. Of course Quinn would reject this, declaring that not pushing the vehicle is not a kind of action. Then he should explain why; and that, I believe, would force him to confront the notion of negativeness.

In several places, Quinn comes dangerously close—as have others—to the attractively simple idea that when someone allows U to come about, it comes about 'because of something he doesn't do'. It's simple all right, and I slipped it past you in my introductory lists of locutions; but taken literally it is wrong. The disaster in Stayback occurred because Agent did not interpose the rock; if it occurred because of something he didn't do, what is that something? In fact there is nothing such that he didn't do it and the disaster occurred because of it. Interposing the rock? He didn't do that, but the disaster didn't occur because of his interposing the rock. Well, then, not interposing the rock? It's true that the disaster occurred because of his not interposing the rock, but that is something he did, not something he didn't do. Of course there is nothing wrong with 'It happened because of something he didn't do' when this is taken as an idiom. If we want clarity, however, we should replace it by a literal statement of its meaning, which is provided by 'It happened because of a negative fact about

<sup>8</sup> Jeremy Bentham, *An Introduction to the Principles of Morals and Legislation* (1780), chap. 8, sec. 8.

<sup>9</sup> Warren S. Quinn, 'Actions, Intentions, and Consequences: The Doctrine of Doing and Allowing', *Philosophical Review* 98 (1989): 287–312.

his behavior’.

It is interesting to see what lengths philosophers will go to in avoiding the concept of a negative fact. Donald Davidson remarks, ‘We often seem to count among the things an agent does things that he does not do: his refrainings, omissions, avoidances.’<sup>10</sup> It must be a mistake to count things I don’t do among the things I do! It isn’t a mistake to include my omissions, but they are not things I don’t do. What Davidson should have said is that in describing a person’s behavior we often include negative facts about it.

### 11. The morality of the two distinctions

I have remarked that the positive/negative distinction obviously lacks basic moral significance. That was based on one metric for the possibility space, but I have not been able to find any plausible alternative that does better; so the result seems to stand. Donagan’s active/passive distinction, on the other hand, is not so obviously devoid of moral significance. Still, it is not obvious, either, that it does have any, so there is something to be thought about here.

If the active/passive distinction carries basic moral weight, then something like this is right:

If someone’s  $\phi$ ing would have a bad consequence U, that fact counts morally against his  $\phi$ ing. Its weight as a reason against his  $\phi$ ing is lessened if the following is true: If in this situation the person had not exercised his agency at all, U would have come about.

I can find no reason to accept this. When some bad state of affairs U comes about, this implies nothing about the

morality of my behavior if my agency was not involved: there is an unbreakable link between

•My agency was not involved in the causation of U, and ‘I did not act wrongly’. The moral force of that, however, does not carry over to

•U would have obtained even if my agency had not been involved.

Nothing but muddle could lead anyone to think that ‘I could have prevented it, but I didn’t’ is significantly like ‘I had nothing to do with it’. With that muddle set aside, I see no way of finding moral significance in the active/passive distinction. This is not to say that it obviously hasn’t any, as I do say about positive/negative when that is based on my specificity metric; but I do contend that those who think that it is morally significant owe us reasons.

There is evidence that Alan Donagan himself had no great faith in the power of his analysis to invest making/allowing with moral significance. Later on in *The Theory of Morality*, when he is defending moral absolutism against an attack of mine (pp. 157–58) he does not point out that my attack avowedly depends on the positive/negative analysis of making/allowing and therefore collapses if that is rejected. Given the weakness of what he does say in defense of absolutism, it is striking that he does not counterattack using the weapon he has forged earlier in the book—his active/passive analysis of the making/allowing distinction. Indeed, he seems to have underrated that analysis in every way. The pages of his book in which it is presented are nowhere referred to in the index; and in correspondence with him I found him to be less interested in it than I was.

<sup>10</sup> Donald Davidson, ‘Reply to Bruce Vermazen’, in *Essays on Davidson: Actions and Events*, ed. Bruce Vermazen and Merrill Hintikka (Oxford: Clarendon Press, 1985).