

Weyma Lübke, Regensburg University

Action Theory and Decision Theory on Attributing Outcomes

(may be quoted as: Unpublished paper, Workshop “Causes, Norms, and Decisions”, Philosophy Department, University of Hannover, August 2018)

1 Ascriptivism

2 Anscombe on Sidgwick on “intended” outcomes

3 Modelling acts by their expected outcomes

4 Are “outcomes” caused?

Before deciding whether it might be useful that I visit this workshop, Enno and I emailed a bit about our respective research interests. Mine presently concerns conceptual gaps between juridic and economic ways of describing and judging actions and omissions. For this occasion here, I then chose to speak about problems that arise if one tries to represent in decision theoretical terms certain normative elements which are inbuilt in ordinary and legal talk about actions and omissions. As far as I can see, these problems have some relevance for conceptualizing the *causal* element in the term “outcome” as well – the idea that an outcome, as opposed to just any state of affairs, is a state of affairs which is caused by an agent’s choice. My examples will have nothing to do with Newcomb-style cases, nor with cases of overdetermination and the like which bother counterfactual analysis in particular. I will compare the views of decision theory in its standard, Savage-like form on the one hand and of “ascriptivism” on the other hand – a tradition within philosophical action theory that was inspired by philosophy of law. I start by giving you some information about what that tradition is.

1 Ascriptivism

Ascriptivism is a term used by philosophers for a special position in action theory. The term was originally coined by Peter Geach to label claims made by the legal philosopher Herbert L. A. Hart in an article published in 1948.

Folie 3 mit 2

Herbert L. A. Hart (1948/1949), The Ascription of Responsibility and Rights.
Proceedings of the Aristotelian Society 49, 171 ff

Peter Geach (1960), Ascriptivism. *The Philosophical Review* 69, 221 ff

Hart's thesis was that ordinary as well as legal talk of actions, much like talk of rights, is of an "ascriptive" as opposed to descriptive nature. Primarily, if not generally, it is used to attribute responsibility, like in "He did it". Hart's article was harshly criticized in the early days, by Geach and by others, and was then as good as forgotten within analytic action theory. The position has, however, seen a philosophical revival in the last years. I give you an idea of what the debate is about by discussing one of the early critiques, George Pitcher's:

Folie 4 mit 2 und 3

George Pitcher (1960), Hart on Action and Responsibility.
The Philosophical Review 69, 226 ff

Pitcher opposes, among else, Hart's talk of "ascribing responsibility for actions" as such. He gives an example: The kindergarten teacher looks at her broken vase and asks "Who did it?". If little Jane answers "Johnny did it", it is fine to say that Jane ascribes responsibility to Johnny. However, Pitcher says, she does not ascribe to him responsibility for an action. By the word "it", both Jenny and the teacher refer to the broken vase, not to Johnny's breaking the vase. So it is not an action, but a state of affairs, which is a result or a consequence of an action, that Johnny is claimed to be responsible for.

I think we understand the point: The teacher is not pointing towards an action when she asks “Who did it?”, but to the broken vase. She could have asked as well “Who is responsible for this?” (that is, for the state of the vase), Pitcher says. And then he adds: “That is to say”,

Folie 5 mit 2

“she simply wants to know whose act it was that resulted in the smashing of her vase.”
Pitcher (1960)

I do not agree with this last analysis. The teacher asked “Who did this”, which, in the context, means “Who broke the vase?”. Why should we read this as saying “Who did something that resulted in the breaking of the vase”? Pitcher analyses an ordinary-language action sentence (“Johnny broke the vase”) in a consequentialist manner, that is, as a judgment that describes a consequence of some *previous* action of Johnny. Why?

The answer is this: Pitcher wants to separate normative talk from descriptive talk. Let us assume that Johnny shot a rubber ball through the room. Pitcher would classify this information as descriptive – just like his own example sentence (“He sat down to dinner”), which, in Pitcher’s view, describes a bodily movement. Talk on what happens when the ball hits the vase will also be descriptive. But if we go on and judge that Johnny is to blame for what happened, we are no more describing things. We are then talking normatively, since blaming implies the judgment that Johnny *should* have behaved otherwise. So this is the only sensible meaning Pitcher can detect in Hart’s claim that action talk is ascriptive: Hart must be referring to what we do when we blame people for certain acts. When Hart claims that we literally *ascribe acts* instead of describing them, however, he seems to stipulate that blaming someone for an act is somehow constitutive of the act – which, in Pitcher’s view, must surely be nonsense.

Again, I think, we can see what Pitcher means. If the teacher blames Johnny for having thrown the ball, she does not thereby constitute his throwing the ball. That happened before, after all. By analogy, it seems, if she blames him for having broken the vase, she does not thereby constitute his breaking the vase. Still, I cannot fully agree with Pitcher again. What happened, happened and does not depend of whether and whom one is later on going to blame – so far I *do* agree (even if one should keep in mind that what happened will often depend on an agent's expectations of eventually being blamed later on). But in any case: in which words – and that includes: with which verbs – a person refers to what happened must not therefore be independent of whether and whom she is going to blame as well. If the teacher thinks that Johnny should not have thrown the ball, she will say “Johnny has broken the vase”. If she thinks that Johnny was perfectly justified to play ball in the children's playing room, and that she obviously wasn't in her right mind to decorate that room with a vase, she might be very far from saying or even from thinking that Johnny has broken the vase. Instead, she might later on say to her husband “Sorry, darling, I ruined that vase”.

So far, this shows two things (which are, I think, quite trivial at least to legal philosophers):

Folie 6 mit 2

1. Action talk itself, not only responsibility talk or blaming, varies with norms about how one should behave.
2. Noting that an event is among the consequences of a certain act does not as such amount to, nor does it suffice for, ascribing responsibility for it to the actor (unless causal talk itself is used in the ascriptive way, which does happen).

Pitcher seems to be aware of the second point, since he claims later on that a person's being responsible for a state of affairs *x* involves two things:

Folie 7 mit 2

- (a) x is a consequence (or result or upshot) of something he has done
- (b) if x is unfortunate (felicitous), then, *under certain conditions*, he is deserving of censure (praise)

Pitcher (1960), *emphasis added*

So there are further conditions that must be fulfilled to ascribe responsibility, or better (since this was the ordinary and legal language observation that motivated Hart's article): that must be fulfilled to use a phrase that classifies a consequence of something a person has done itself as that person's act – like in “you broke the vase”. Pitcher, for one, seems somewhat helpless about these additional conditions – he does not elaborate on the clause and in the vase example he simply ignores the point.

The most widely discussed concept in philosophical action theory to somehow grasp these further conditions has been “intention”. In my second part, I present a classical debate on this notion, and I will present it in a way that brings us closer to decision theory's idea of characterizing acts by their expected consequences or outcomes (remember Savage:)

Folie 8 mit 2

“[...] an **act** may therefore be identified with its possible consequences.”
L. Savage (²1972, ¹1954)

Folie 9

2 Anscombe and Sidgwick on intentions

Elizabeth Anscombe's article „Modern moral philosophy”,

Folie 10 mit 9

G. E. M. Anscombe (1958), Modern Moral Philosophy. *Philosophy* 33, 1 ff

– the article in which she coined the noun “consequentialism” as it is since used in moral philosophy – is among the most read philosophical papers in the last century. Anscombe there refers to a proposal for using the concept of intention, made by the utilitarian philosopher Henry Sidgwick.

Folie 11 mit 9 und 10

Henry Sidgwick (⁷1907/¹1874), *The Methods of Ethics*

According to Sidgwick, one intends all foreseen consequences of one’s voluntary actions. Anscombe identifies this proposal as the origin of the „very bad degenerations of thought“ that she attributes in her article to her own time’s moral philosophy. Sidgwick uses his definition to put forward an ethical thesis which, as Anscombe says, is now accepted by many people. This thesis is that

Folie 12 mit 9

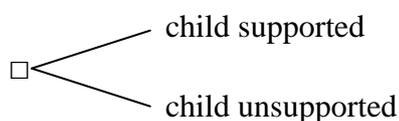
„[...] it does not make any difference to a man’s responsibility for something that he foresaw, that he felt no desire for it, either as an end or as a means to an end.”

Anscombe (1958)

So the debate is about the respective roles of foresight and desire for ascribing responsibility, that is, for classifying an event or state of affairs as somebody’s doing. In order to explain why Sidgwick’s position is untenable, Anscombe discusses an example. There is a man who is responsible for the maintenance of a child. Withdrawing support from the child just because he does not want to maintain it any longer would be a bad thing to do, Anscombe says.

Folie 13 mit 9

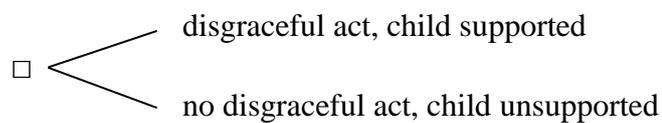
Case 1



(As you see, I now describe choices with decision trees, whose outcomes are to be brought into a moral preference ordering and thus describe purportedly morally valuable or disvaluable events or states of affairs.) Anscombe now asks us to imagine the man to be in a different situation of choice. Either he does „something disgraceful“, as Anscombe puts it, or else he will go to prison, which means that there will be no further support for the child.

Folie 14 mit 9 und 13

Case 2

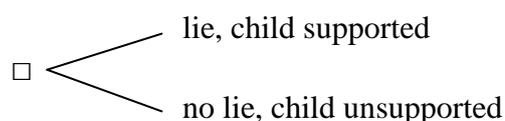


Anscombe comments that Sidgwick's position will lead him to say that in this situation the man must weigh up the badness of the disgraceful act and the badness of intentionally withdrawing support from the child. And if the disgraceful thing is (quote Anscombe) "in fact a less vicious action than intentionally withdrawing support from the child would be", then Sidgwick would have to recommend that the man should do the disgraceful thing.

You will probably agree that this is not as such a result so untenable that we must feel deterred from Sidgwick's position. It rather seems that Anscombe, catholic as she was, voices the old doctrine that certain things must never be done, whatever the consequences (perhaps the disgraceful thing, which Anscombe herself does not name, is a lie).

Folie 15 mit 9

Case 2

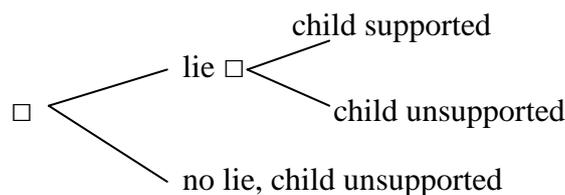


In that direction, at least, goes the dominant reading of her text. But I am not convinced that this is Anscombe's point, or better: I believe that there is a point lying behind her opposition against Sidgwick which is independent of her deontological convictions. Let me explain.

Anscombe opposes, as we heard, that one should weigh the moral badness of the disgraceful act with the moral badness of *intentionally* withdrawing support from the child. According to Sidgwick, all outcome parts are indeed intended because they are all foreseen. They are thus all part of the act that must be morally judged. However, there is a difference between the upper arm and the lower arm which I have not modelled so far.

Folie 16 mit 9

Case 2



In the upper arm, we can note the two parts of the outcome as independent acts, since if the man lies, and thus does not have to go to prison, he is free to decide whether he further supports the child or not. In the lower arm, the parts cannot be separated in that way. If the man refuses to lie, the second part of the outcome will come about no matter what else he does or omits. Leaving the child unsupported is here no *separate* act of the man, and if it is not a separate act, how can it be ascribed a separate moral value?

Mind that I do not want to suggest (even if this is what Anscombe aimed at) that the man cannot be blamed for the child's being unsupported if he refuses to lie, because then he can no more avoid the child's being unsupported. If only he could have avoided it at all, he can in principle be blamed for it, I think. Imagine the lie to be a lie against an intelligence officer of a totalitarian regime.

Almost all people would classify this as a white lie, and possibly, in view of the man's responsibility for the child, as a mandatory act as long as it can safely be done. From the point of view of such non-absolute convictions about lying, the man can be blamed for abandoning his child if he refuses to lie. However (and this is the point I am heading at in this section): From this point of view, too, the normative judgment is not arrived at by weighing the badness of the lie against the badness of abandoning the child. It is rather that here the lie is not judged to have any badness. Nobody, in other words, first ascribes moral goodness to refusing the lie, then, second, ascribes moral badness to abandoning the child, and then, third, weighs these values, such that, fourth, the right choice is arrived at by maximizing the overall moral value of the outcome. There are no independent moral values when there are no independent acts.

So much for this debate. I meant this to establish the following point (and otherwise than my first two points, this one is not trivial, not even to legal philosophers):

Folie 17 mit 9

3. An actor's foresight may include outcome parts that she does not desire. Call them intended or not – you may only ascribe separate moral value to them if they are outcomes of separate choices.

In the next section, I explain how this claim threatens a standard assumption of decision theory, and then I ask whether it may even affect the purportedly causal notion of the outcome of a choice.

Folie 18

3 Modelling acts by their expected outcomes

Imagine a medical doctor. There are three of her patients in front of her, David, Peter, and Mary.

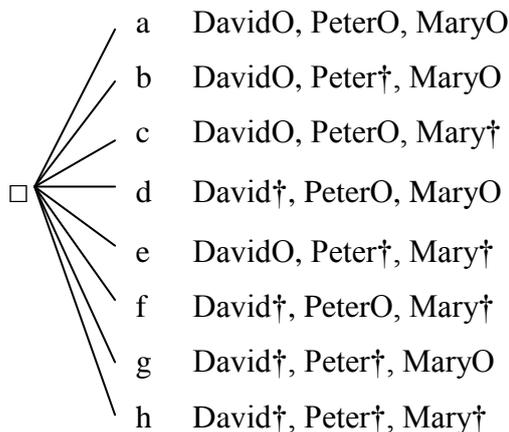
Folie 19 mit 18

David (10 mg), Peter (5 mg), Mary (5 mg)

They will all die if they do not get a certain amount of a drug. David needs 10 mg of the drug to survive, Peter and Mary each need 5 mg, and there is nobody else who is affected by the doctor's choice. Fortunately for the patients, she has 20 mg at hand. So here is the list of her options to act (or to omit) as described by the outcomes that may be expected to come about:

Folie 20 mit 18

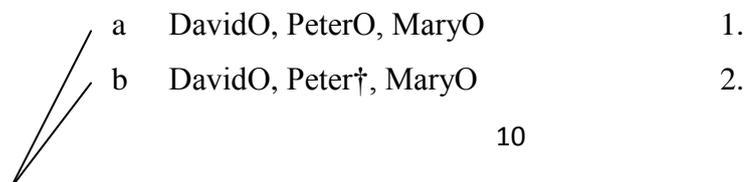
20 mg-case



Circles stand for welfare levels that are reached by survival, crosses for welfare levels that are reached by death. I assume that every person's welfare level in case of her death and in case of her survival equals that of any other person respectively. Perhaps you join the great majority of philosophers who say that the doctor's moral preference ordering over these outcomes should then simply depend on the number of people saved.

Folie 21 mit 18

20 mg-case



□	c	DavidO, PeterO, Mary†	2.
	d	David†, PeterO, MaryO	2.
	e	DavidO, Peter†, Mary†	3.
	f	David†, PeterO, Mary†	3.
	g	David†, Peter†, MaryO	3.
	h	David†, Peter†, Mary†	4.

(I do not belong to that majority, but never mind this just now.)

Let us now assume that the doctor is in the same situation, except that she has only 10 mg of the drug. That means she can save either Peter and Mary, or David, or less, but not more.

Folie 22 mit 18

10 mg-case

	a	DavidO, PeterO, MaryO	1.
	b	DavidO, Peter†, MaryO	2.
	c	DavidO, PeterO, Mary†	2.
□	d	David†, PeterO, MaryO	2.
	e	DavidO, Peter†, Mary†	3.
	f	David†, PeterO, Mary†	3.
	g	David†, Peter†, MaryO	3.
	h	David†, Peter†, Mary†	4.

Only the five options below are left. Now please look at options e and f in particular:

Folie 23 mit 18

10 mg-case

a	DavidO, PeterO, MaryO	1.
b	DavidO, Peter†, MaryO	2.

	c	DavidO, PeterO, Mary†	2.
□	d	David†, PeterO, MaryO	2.
	e	DavidO, Peter†, Mary†	3. ?
	f	David†, PeterO, Mary†	3. ?
	g	David†, Peter†, MaryO	3.
	h	David†, Peter†, Mary†	4.

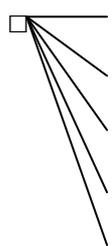
These two options are ranked equal. I claim, however, that their moral value is not equal, and what is more, I expect that you will agree when I explain why. You can grasp my reason best if you properly imagine what the doctor actually does when she chooses these options. If she chooses f, she saves Peter by giving him 5 mg of the drug, and then she watches Mary and David die while she could very well save Mary with the rest of the drug. She omits, in other words, to make a Pareto improvement. She does no such thing when she chooses option e: Here she saves David by giving him 10 mg of the drug (all she has), and then she watches Mary and Peter die because there is nothing left that she could do for them.

I hope you see how this point relates to the claim I made in my last section: An actor's foresight may include outcome parts that she does not desire. Call them intended or not – you may only ascribe separate moral values to them if they are outcomes of separate choices. The morally relevant difference between options e and f is

Folie 24 mit 18

10 mg-case

a	DavidO, PeterO, MaryO	1.
b	DavidO, Peter†, MaryO	2.
c	DavidO, PeterO, Mary†	2.

	d	David†, PeterO, MaryO	2.
	e	DavidO, Peter†, Mary†	3. ?
	f	David†, PeterO  Mary†	3. ?
	g	David†, Peter†, MaryO	3.
	h	David†, Peter†, Mary†	4.

that in option f Mary's death can be modelled as the outcome of a separate choice (the choice to make or not to make a Pareto improvement, as I said). In option e Mary's death cannot be modelled as the outcome of a separate choice. So we cannot ascribe a separate moral disvalue to Mary's death in that option. We must give it the same moral value which we give to the doctor's decision to save David, because David's survival and Mary's death are inseparable outcomes of one and the same choice (like, in Anscombes example, the refusal to lie and the end of support for the child).

I claimed that the point threatens a standard assumption of decision theory. Which one? Savage-style decision theory assumes that the preference ordering is complete (it reaches over the whole domain of possible alternatives) and that it may be generated by pairwise comparisons which satisfy the transitivity axiom. Such a preference ordering does not depend on the extent of the feasible set which happens to prevail in a concrete situation of choice. The inequality between options e and f, however, does depend on the extent of the feasible choice. The non-separability of David's and Mary's deaths in option e vanishes when the feasible set comprises the whole domain – that is, when we go back to the 20 mg-case.

Folie 25 mit 18

20 mg-case

	a	PeterO, MaryO, DavidO
	b	Peter†, MaryO, DavidO
	c	PeterO, Mary†, DavidO

- — d PeterO, MaryO, David†
- e Peter†, Mary†, DavidO
- f PeterO, Mary†, David†
- g Peter†, MaryO, David†
- h Peter†, Mary†, David†

If there is enough of the drug to save all three person, like here, every single death will be the outcome of a decision not to make a Pareto-improvement. No patient's fate is by scarcity linked to the fate of any other patient. There is then no reason left why options e and f should not be judged as morally equivalent.

I have no time here to elaborate on possible decision theoretical amendments – on proposals to redescribe the outcomes, on choice-functional approaches like Amartya Sen's, or what have you. For the present occasion, I want to draw your attention to another point: On the basis of examples like this, one may wonder whether – and if so, in what sense – the outcomes can taken to be “caused” by the respective choices. This point – a question, indeed, rather than a thesis – I put forward in my last section.

Folie 26

4 Are “outcomes” caused?

Look at the 10 mg-case again.

Folie 27 mit 26

10 mg-case

- a DavidO, PeterO, MaryO 1.
- b DavidO, Peter†, MaryO 2.
- c DavidO, PeterO, Mary† 2.
- — d David†, PeterO, MaryO 2.
- e DavidO, Peter†, Mary† 3. ?

- | | | |
|---|-----------------------|------|
| f | David†, PeterO, Mary† | 3. ? |
| g | David†, Peter†, MaryO | 3. |
| h | David†, Peter†, Mary† | 4. |

Can the doctor who chooses one of the five outcomes within the feasible set be said to have caused the respective outcomes? There are no classical overdetermination or preemption problems looming here. So let us perhaps use the counterfactual test. And please concentrate, in particular, on option h:

Folie 28 mit 26

10 mg-case

- | | | |
|-------------------------------------------------------------------------------------|---|-----------------------|
| | a | DavidO, PeterO, MaryO |
| | b | DavidO, Peter†, MaryO |
| | c | DavidO, PeterO, Mary† |
|  | d | David†, PeterO, MaryO |
| | e | DavidO, Peter†, Mary† |
| | f | David†, PeterO, Mary† |
| | g | David†, Peter†, MaryO |
| | h | David†, Peter†, Mary† |

Had the doctor chosen any other option out of the feasible set, there would have been a different outcome. So it seems that she caused the outcome. For the single deaths, the *parts* of the outcome h, slightly different counterfactual judgments hold: For any of the parts, it is true that there would have been a different result, had the doctor chosen one or several *particular* (not just any) other outcomes out of the feasible set. So it seems that the doctor caused each part. Still, the doctor cannot be said to have caused three deaths. One death at least would have happened whatever she does or omits. One death at least is thus not, I should say, a consequence of her choice. One might say that it should then be deleted from the outcome descriptions. However, one cannot delete “one death at least”

from these descriptions without deleting one or several particular deaths – and I see no factual basis to decide which one or which ones that should be in option h. So I wonder whether under the different approaches to causality which are studied in the present philosophical literature there is one, to your knowledge, which has an answer ready as to what exactly the doctor who chose option h has caused, and whether the answer is claimed to be given on a purely descriptive basis.

Thank you.