

## WHAT HAVE I DONE?

– Timothy Chappell –

**Abstract.** An externalist view of intention is developed on broadly Wittgensteinian grounds, and applied to show that the classic Thomist doctrine of double effect, though it has good uses in casuistry, has also been overused because of the internalism about intention that has generally been presupposed by its users. We need a good criterion of what counts as the content of our intentional actions; I argue, again on Wittgensteinian grounds, that the best criterion comes not from foresight, nor from foresight plus some degree of probability, nor from any metaphysics of “closeness”, but simply from our ordinary shared understanding of what counts as doing a given action, and what does not.

**Keywords:** intention, externalism, Wittgenstein, double effect, Thomas Aquinas.

Was ist der natürliche Ausdruck einer Absicht? –Sieh ein Katz an, wenn sie sich an einen Vogel heranschleicht; oder ein Tier, wenn es entfliehen will.<sup>1</sup>

Respondeo dicendum quod nihil prohibet unius actus esse duos effectus, quorum alter solum sit in intentione, alius vero sit praeter intentionem. Morales autem actus recipiunt speciem secundum id quod intenditur, non autem ab eo quod est praeter intentionem, cum sit per accidens, ut ex supradictis patet. Ex actu igitur alicuius seipsum defendentis duplex effectus sequi potest, unus quidem conservatio propriae vitae; alius autem occisio invadentis. Actus igitur huiusmodi ex hoc quod intenditur conservatio propriae vitae, non habet rationem illiciti, cum hoc sit cuilibet naturale quod se conservet in esse quantum potest. Potest tamen aliquis actus ex bona intentione proveniens illicitus reddi si non sit proportionatus fini. Et ideo si aliquis ad defendendum propriam vitam utatur maiori violentia

---

<sup>1</sup> Ludwig Wittgenstein, *Philosophical Investigations* I, 647. “A cat’s movements in stalking a bird are hardly to be called an expression of intention... Wittgenstein seems to me to have gone wrong in speaking of ‘the natural expression of an intention’” (Anscombe, *Intention*, section 2; thanks to Roger Teichmann for reminding me of the passage). Anscombe seems to me to be reading Wittgenstein rather uncharitably here. We can say both that my involuntary and non-convention-governed sigh is an expression of relief, and also that my voluntary and convention-governed remark “I’m relieved” is an expression of relief. The sigh *betrays* my relief, just as the cat’s movements betray its design on the bird. We might want a word to capture the distinctive way in which this sort of behaviour counts as expression. If we do, why not “natural”?

quam oporteat, erit illicitum. Si vero moderate violentiam repellat, erit licita defensio, nam secundum iura, vim vi repellere licet cum moderamine inculpatae tutelae. Nec est necessarium ad salutem ut homo actum moderatae tutelae praetermittat ad evitandum occisionem alterius, quia plus tenetur homo vitae suae providere quam vitae alienae. Sed quia occidere hominem non licet nisi publica auctoritate propter bonum commune, ut ex supradictis patet; illicitum est quod homo intendat occidere hominem ut seipsum defendat, nisi ei qui habet publicam auctoritatem, qui, intendens hominem occidere ad sui defensionem, refert hoc ad publicum bonum, ut patet in milite pugnante contra hostes, et in ministro iudicis pugnante contra latrones. Quamvis et isti etiam peccent si privata libidine moveantur.<sup>2</sup>

Good ethics presupposes good philosophical psychology, and in particular a good account of the voluntary and the intentional: so famously insisted Elizabeth Anscombe<sup>3</sup>. And while there may be a flourishing sub-culture of Wittgensteinianism in philosophy nowadays, that is no evidence that the general culture of contemporary philosophy is as influenced as it ought or needs to be by Wittgenstein: a need that Sir Anthony Kenny has reminded us of. My project in this paper, accordingly, is to bring in Wittgenstein to illuminate a familiar Thomist thesis.

I shall argue that there is a thesis in philosophical psychology which is true, which has real application in ethics, and which deserves to bear the name “the doctrine of double effect”, *duplex effectus* as Aquinas calls it in my second epigraph.<sup>4</sup> This doctrine does an important part of the work of showing us something ethically fundamental, namely the limits of our negative responsibility. We want to know, for instance, what makes the moral difference that we intuitively see there must be between the terrorist who kills the hostages because no ransom is paid, and the authorities that refuse to pay this ransom. Or we want to know why,

---

<sup>2</sup> Thomas Aquinas, *Summa theologiae*, 2a2ae.64.7.

<sup>3</sup> MMP, p. 169 in HLAE: “it is not profitable for us at present to do moral philosophy; that should be laid aside at any rate until we have an adequate philosophy of psychology, in which we are conspicuously lacking.” Cp. p.174.

<sup>4</sup> Here is the doctrine in Mangan 1949: 43’s classic statement: “A person may licitly perform an action that he foresees will produce a good effect and a bad effect provided that four conditions are [satisfied]: (1) that the action in itself from its very object be good or at least indifferent; (2) that the good effect and not the evil effect be intended; (3) that the good effect be not produced by means of the evil effect; (4) that there be a proportionately grave reason for permitting the evil effect.” Grisez (CMP 239-241), Boyle, *Toward understanding the principle of double effect*, “Ethics” (90) 1980, p. 528-530, and others give virtually identical statements of the doctrine. I don’t know whether any statement of it is in any sense authoritative for Catholics.

when Sir Thomas More resisted the government of Henry VIII over the divorcing of Catherine of Aragon, in the full knowledge that his so doing would very probably deprive his wife of a husband, his children of a father, and his estate of an inheritance, it was not thereby correct for his wife Alice to protest (as she does in Robert Bolt's play<sup>5</sup>) that *Sir Thomas More* was so depriving them, rather than Henry VIII and his minions. If the true structure of our intentional action was simply what pure consequentialists say it is – that we should make all and only those bodily movements which will bring about the best overall consequences in the long run – then Alice More's accusation would be right, and so would the terrorist who says that the authorities "left him with no alternative". Quite generally: if pure consequentialism were true, our negative responsibility would be unlimited; even if – as consequentialists find it comforting to add – we would not then do best to *believe* that our negative responsibility is unlimited. The doctrine of double effect wards off this prospect, by dealing with the question of philosophical psychology that lies at the heart of the issue, namely the question "What have I done?"<sup>6</sup> Is my intentional action always really just another instance of the single action-type that is all that the consequentialist fundamentally recognises, namely a better or worse pursuit of overall good? Or are there other, more limited and less profoundly counter-intuitive, answers that we can offer to the question what I, or anyone, has done?

The answers to these two questions are, respectively, No and Yes, and I shall show how the doctrine of double effect is part of what we need to spell out these answers. First, however, I shall also argue that, for a reason made clear by Wittgenstein in my first epigraph, the doctrine of double effect cannot do anywhere near all the casuistical work that it has often been recruited for by Thomists – though not, so far as I can see, by Thomas.

The reason is that intentions are, like meanings and emotions and suspicions and thoughts and sensations and other mental phenomena, basically and

---

<sup>5</sup> Robert Bolt, *A Man for all Seasons*, p. 86. Alice is not very explicit about it, but this is clearly her thought.

<sup>6</sup> Joseph Boyle apparently thinks that uses of DDE by *non-absolutists* are not only gratuitous but illicit: "Outside the absolutist context of the Catholic tradition, DDE is not needed; and those who reject this context are not entitled to use it" (Boyle, *Who is entitled to double effect?*, "The Journal of Medicine and Philosophy" (16), p. 477). But so far as I can see, DDE makes a point about the nature of intention that has no intrinsic connection at all with the idea of a moral absolute (i.e. an exceptionless or virtually exceptionless moral prohibition). So anyone who is not a pure consequentialist at least *can* deploy DDE, perhaps *should*, in any case surely *may*. (Conversations with them suggest that Brad Hooker, Tim Mulgan, John Skorupski, and Michael Ridge all accept that there is some kind of moral distinction between the intended and the merely foreseen: indirect consequentialists to a man.)

essentially public. Anything about them that was *not* so public would not and could not be what we were talking about when we talked about intentions, meanings, sensations, etc. Like the unseen beetle in the box (*PI I*, 293), the hidden part of the mental phenomenon would simply drop out of consideration. (The point of that famous analogy is not, as some interpreters seem to think, that the box is empty, i.e. that we have no inner processes. The point is rather that we *do* have inner processes, but they are not hidden. So to speak, we see each others' beetles all the time.)<sup>7</sup>

Here is one simple example of the publicity of the mental. When my daughter Róisín was four, I once saw her leave her finger – unintentionally – in a slamming door. I did not *infer* her agony. I saw it (and heard it). Her pain was as directly perceptible to me as the slamming door was. And how directly perceptible is that? Completely directly, I would say. It is certainly a whole lot more natural to call *this* directly perceptible than any “sense-data”, “qualia”, or “phenomenal seemings” that we might like to talk about.<sup>8</sup> Thanks to recent work on mirror-neurons in experimental psychology, we may even be beginning to get a handle on *how* we perceive each other's sensations.<sup>9</sup>

A second example of the publicity of the mental. Suppose Anna, who is married to K, is watching Alexei compete in a horse-race. Her fervour in urging Alexei on has something excessive about it, something that tells everyone around her, including K, that she is in love with Alexei. Anna then realises what those around her have realised. Yet it is only now, and by “reflection” from their reactions, that she realises it herself; as if she were to find out that she is blushing only by looking in a mirror. What is manifest in Anna's behaviour is *more* manifest to others than it is to her.<sup>10</sup>

Because intentions too are normally public in this way, the right answer to the question “How do we know what someone's intention is?” is normally “Nothing simpler”. A cat stalks a bird. A deer runs toward a fence and then, seeing it, shies away. A man, as in all the old jokes, walks into a bar. Nothing is more natural for us, and few things are easier, than seeing simply from its public behaviour

---

<sup>7</sup> Cp. Wittgenstein, op. cit., I, 313: “Ich kann Schmerzen vorführen, wie ich Rot vorführe, und wie ich Gerade und Krumm und Baum und Stein vorführe. Das nennen wir eben ‘vorführen’.”

<sup>8</sup> On this see further my *Moral Perception*, “Philosophy” 2008.

<sup>9</sup> See e.g. G. Rizzolatti, C. Sinigaglia, *Mirrors in the Brain: How We Share our Actions and Emotions*, Oxford University Press 2008.

<sup>10</sup> I got this example long ago, from a class that David Pugmire gave in Southampton in 1989. The story is almost but not quite in *Anna Karenina*, Part II, Ch.28 – hence my names. (There Karenin and other spectators at a horse-race do realise Anna's passion for Vronsky from her behaviour. But Anna herself knows already.)

that some creature *means to get something*: the cat means to get the bird, the deer means to get out (and then discovers it can't), the man means to get a beer. In cases like these, involving humans,<sup>11</sup> talk about intentions is at home, and (philosophy notwithstanding) uncontroversially available to us.

I said above that my daughter's pain was directly perceptible. When the aforementioned man walks into a bar, normally his intention to get a beer is directly perceptible too. (Imagine him wiping the back of his hand across his mouth as he crosses the street, licking his lips, feeling in his pocket for his wallet as he pushes at the bar-room door.) It is not, as the behaviourists used to claim, that his intention *is* his walking-into-a-bar behaviour, any more than Róisín's screaming and clutching of her finger in her other hand *is* her pain. The behaviour *expresses* the intention, as perfectly as a clock-face shows the time. It's not that the time is *reducible to or identical with* the state of the clock-face; that would be a category mistake. Yet we can normally look at the clock-face and, from it, simply recognise the time. Just likewise, we can normally simply recognise what other agents intend, what sensations they have, and the rest of it, just by observing them. "Wenn man das Benehmen des Lebewesens sieht, sieht man seine Seele."<sup>12</sup>

"But Róisín might not have been in pain at all – she might have been tricking you with a grisly rubber finger she'd just bought in a Joke Shop. And the man walking into the bar might not intend to get a beer at all – he might be an actor, or an undercover detective. What are you directly seeing *then?*" Apparent pain, of course, or an apparent intention. The fact that *these* can be seen is no evidence that real pain and real intentions cannot *also* be seen, any more than the fact that there are forged fivers is evidence that there are no real fivers, or the fact that we seem to be able to see the right time from wrongly-set clocks is evidence that we cannot actually see the right time from rightly-set clocks. Normally with a clock, we just look and see the time. The possibility of a mistake arises afterwards. Likewise with sensations and intentions, the cases where we simply and directly observe them are the primary ones. Just as the very idea of a forged fiver depends upon the prior idea of a genuine fiver, so the idea and the possibility of a mistake about a sensation or an intention happening *sometimes* is built upon the prior fact that normally there *isn't* a mistake.

---

<sup>11</sup> Wittgenstein, op. cit., I, 647 is not, of course, about whether non-human animals have intentions too, though it is about animal-kingdom-wide continuities. From evidence elsewhere it is clear that Wittgenstein thought not, as did Aquinas (op. cit., a2ae12.5), but that is not the issue here.

<sup>12</sup> Ibidem, I, 357.

“But surely the man himself knows better than anyone else what he really intends!” If you genuinely think that, think again about my example of Anna. Or try saying it to the man’s wife; she’ll soon put you right. Of course there is a sense in which each of us is, when things go well, peculiarly intimately related to his own intentions: as Anscombe argues in *Intention*, the normal way for us to know our own intentions is directly and non-observationally, just like the normal way in which we know our other mental states, or know what we are doing with our own limbs and muscles. The directness and non-observationality of such self-knowledge doesn’t imply its incorrigibility from other, more indirect sources.<sup>13</sup>

Again, the point is not only about intentions; it generalises across the realm of the mental. In my quasi-Karenina example, it would be futile for Anna to deny what everybody else can see about her emotional state – even if it is only because they see it that she comes to see it for herself. We can even imagine people being corrected about their own pain-sensations: the St John’s Ambulance men might rush a fired-up rugby-player off the pitch, saying to her “Come on, you’re in a lot of pain” – and they might be right even though she sincerely denies it. So with intentions, it is similarly futile for the man walking into the bar to deny his beer-purchasing intention, even to himself. He might say to himself “I’ll just pop my head round the door and ask ‘Jeff been in?’”. That is what he *says*, but then that is what he has said every evening for the last 25 years – and on every such occasion he has ended up staying in said bar till chucking-out time. “Thus even the most explicit expression of an intention, on its own, is not sufficient evidence of an intention.”<sup>14</sup>

In a host of ordinary-life cases the question arises “What have I done?”; and as the posing of the question often contextually implies, the person who asks it is not always the one best-placed to answer it. One person says to another: “You intended X (adultery, to set the building on fire, to plagiarise another student’s work, to get drunk...) You *obviously* intended X, so don’t bother denying it” – and is absolutely right to say so. Absolutely right, no matter what the accused claims to the contrary. Absolutely right, too, even if the accused quite sincerely *tells himself* that adultery or arson or plagiarism etc. was not his intention. He can make little speeches in his head if he likes: “I am comforting Jean after John’s demise”,

---

<sup>13</sup> On non-observational knowledge of one’s own actions, and how it might be a species of Anscombean practical knowledge, see Adrian Haddock, “The knowledge that a man has of his intentional actions”, forthcoming in Anton Ford, Jen Hornsby, Fred Stoutland, ed., *Essays on Anscombe’s Intention* (Harvard UP). I am grateful to Adrian for showing me his paper.

<sup>14</sup> Wittgenstein, *op. cit.*, I, 641: “So ist also der am meisten explizite Ausdruck der Absicht allein keine genügende Evidenz der Absicht.”

“I am maximising our insurance return”, or whatever. These little speeches, if others get to hear them, are not irrelevant to determining the agent’s exact set of mind; but as is obvious from the practice of law-courts the world over, they are not the sole or even the main evidence of anyone’s intentions.

“But there are cases and cases. Of course sometimes we can tell someone what his intention is, even if he self-deceivingly denies it. That’s quite different from the case where something is a necessary concomitant of what he actually intends. He might simply fail to see it. Or like Anscombe’s man pumping poisoned water, he might see it and not care about it.” I quite agree: of course there are cases and cases. My point is absolutely not to deny this, but rather to build on it. What I am showing here is just that the in-advance question from others “Don’t you see what you’re doing?”, and/ or the ex-post-facto question to myself “What have I done?”, give us the means of unpicking and unpacking the detail of such cases, and the different ways in which our actions may relate to or embody our intentions.

“But sometimes the only way you can tell what someone’s intention is is by asking him – if even that is possible. Imagine someone, John, whose intention is to rescue a woman, Emma, from the rapist, George, whom he sees charging towards her from her left. John charges towards Emma himself, from her right, to pull her physically out of harm’s way; she looks right and so sees John coming, but not George; Emma takes *John* for a rapist, and shoots him dead. Meanwhile George sees this happening and makes himself scarce. Such cases show that two people can make exactly the same external movements – rushing towards Emma from one side or the other – with completely different intentions. They therefore show that intentions are internal, not external.” Again, I quite agree that *sometimes* quite different intentions dictate identical external movements, and that there could in principle be cases where no one can tell what someone’s intention was because many different intentions were compatible with his external movements, and he is no longer around to ask. (Think for instance of the controversies about Mallory and Irvine’s last known movements on Everest.) To think that such cases show that intentions are *never* discernible from external movements is simply a crude over-generalisation. Usually, in fact, they are so discernible. Think here about the other evidence that might be relevant to settling what were the intentions in the John and George case. George’s relatives might protest “You’ve got the wrong man – George was no rapist”, and cite all sorts of evidence from his former life to prove that this was not the sort of thing that George would ever have intended. Or suppose John was apprehended after the fact, and *he* protested his innocent intentions; it would be quite possible to defeat his protests by showing e.g. how habitu-

al rape was for him, or that he was both violent and obsessed by Emma. A court would call such evidence presumptive, because John does not actually perform his rape and George does not actually perform his rescue. But it is merely *secondary*? No; it is just the kind of evidence that we are always amassing about intentions. And after all, as our disbelief of John's protestation shows, a report of an intention is only presumptive evidence too.

Suppose then that I point a gun at your head and pull the trigger. Unless I don't understand guns, or have bizarre causal beliefs, or think this is only a replica, or (...), this means that I intend to kill you.<sup>15</sup> And no speech, inner or outer, on my part professing that something else is my action or my intention can make any difference to that.

Or suppose that I crush and cut off an unborn baby's head. Unless I am bizarrely ignorant of human physiology, or derangedly think it is a plastic model not a real baby, or think that God will intervene miraculously to keep the baby alive, (...), this means that I intend to kill the baby, and that what I have done is killed it. Here too I may profess to do and/ or intend not killing but something else, but it makes no difference. Any competent observer can see that my action and intention are killing ones – even if I cannot see it myself.<sup>16</sup>

In an insufficiently discussed paper on double effect, Elizabeth Anscombe says this:

At this point the Doctrine of Double Effect helps itself to an absurd device, of choosing a description under which the action is intentional, and giving the action under that intention as *the* intentional act. "I am moving what blocks that egress" ... the suggestion is that that is *all* I am doing as a means to my end. This is

---

<sup>15</sup> It is true that if I don't aim at your head, but (say) at your knee, then my intention is to *shoot* you but – probably: unless I e.g. want you to die slowly – not to *kill* you. (Sometimes in films, you see a gunman first level a gun at someone's head, and then pause and move its muzzle downwards to a different target: a nice example of the publicity not only of intentions, but of *changes* in intention.) None of that even begins to justify Grisez in writing this: "The rapist's death is not what is chosen as a means or sought as an end when the woman shoots him in the head to stop his attack (the shooting is not direct killing). Her end is to avoid being raped" (LCL p. 473). If shooting a man in the head isn't direct killing, what on earth is? How could it be more obvious that she intends to kill the rapist, and that if she does kill him, she does it intentionally?

Of course it may also be true (in real-life cases it very often will be) that she acts in the heat of the moment, that she would not shoot him if she could see any alternative, that she will help get him to hospital if he is still breathing after she has shot him, that she will feel terrible about it afterwards, that if guns were not so readily available then neither the rapist nor she would have been in their awful predicament in the first place, and so on. All of this is relevant to the moral character of what she does. None of it affects the description of her intention as to kill, and her action as intentional killing – and, I should say, entirely-justified intentional killing too.

<sup>16</sup> Perhaps this is the thought that lies behind the Vatican's condemnation of craniotomy.

as if one could say: “I am merely moving a knife through such-and-such a region of space”, regardless of the fact that that space is manifestly occupied by a human neck... “Nonsense”, we want to say, “doing that is doing this, and so closely that you can’t pretend only the first gives you a description under which the act is intentional.” For an act does not merely have many descriptions, under some of which it is indeed not intentional: it has several under which it is intentional<sup>17</sup>... Nor can you simply bring it about that you intend *this* and not *that* by an inner act of “directing your attention”. Circumstances, and the immediate facts about the means you are choosing to your ends, dictate what descriptions of your intention you must admit. (HLAE, p. 223)

Circumstances *dictate* what descriptions of your intention you *must* admit: notice the cogency and externality of these correct descriptions. Whatever else the doctrine of double effect may involve, it cannot include a permission or an exhortation to redescribe our intentions “from the inside” in whatever way we like. For the correct descriptions of our “inner processes” quite generally, including our intentions, are made externally, not “from the inside”, and objectively, not at our whim. As John Finnis, to cite one distinguished authority on these matters, apparently agrees:

[...] however “certainly foreseeable” [the very good consequences] may be, they cannot be used to characterise the act itself as, *in and of itself*, anything other than an intentional act of, say, man-killing. This is especially obvious when a blackmailer’s price for sparing his hostages is “killing that man”; the person who complies with the demand, in order to save the lives of the many, cannot deny that he is choosing an act which of itself does nothing but kill. (NLNR, p. 123)

Bad philosophical psychology leads to bad ethics; bad ethics has further, and very familiar, bad consequences. Internalism about intention, the thesis that our intentions are whatever we choose to say they are, or set ourselves to make them by way of little speeches inside our own heads, is false. This false idea in philosophical psychology leads us straight into a false idea in ethics: that we can excuse ourselves in, say, the craniotomy case by adopting an intention only to rearrange some physical stuff and not to cut off a baby’s head, even though the rearranging *is* the decapitating (as we perfectly well know). Or that we can excuse ourselves in a self-defence case by adopting an intention only to shoot a bullet

---

<sup>17</sup> MMP: “An act will fall under several intentional descriptions if it falls under any.” Cp. Thomas Aquinas, *op. cit.*, 1a2ae.12, 3.

through an attacker's brain-stem and not to kill him, even though the shooting *is* the killing (as we also know).

To take an older case, you can reason like this to prove that duelling is licit, as Pascal shows in the *Provincial Letters*:

O mon Père! lui dis-je, voilà un beau fruit de la direction d'intention! Je vois bien qu'elle est de grande étendue; mais néanmoins il y a de certains cas dont la résolution serait encore difficile, quoique fort nécessaire pour les gentilshommes. Proposez-les pour voir, dit le Père. Montrez-moi, lui dis-le, avec toute cette direction d'intention, qu'il soit permis de se battre en duel. Notre grand Hurtado de Mendoza, dit le Père, vous y satisfera sur l'heure, dans ce passage que Diana rapporte p. 5 tr. 14, r. 99. Si un gentilhomme qui est appelé en duel est connu pour n'être pas dévot, et que les péchés qu'on lui voit commettre à toute heure sans scrupule fassent aisément juger que, s'il refuse le duel, ce n'est pas par la crainte de Dieu, mais par timidité; et qu'ainsi on dise de lui que c'est une poule et non pas un homme, *gallina et non vir*, il peut, pour conserver son honneur, se trouver au lieu assigné, non pas véritablement avec l'intention expresse de se battre en duel, mais seulement avec celle de se défendre, si celui qui l'a appelé l'y vient attaquer injustement. Et son action sera tout indifférente d'elle-même. Car quel mal y a-t-il d'aller dans un champ, de s'y promener en attendant un homme, et de se défendre si on l'y vient attaquer? Et ainsi il ne pêche en aucune manière, puisque ce n'est point du tout accepter un duel, ayant l'intention dirigée à d'autres circonstances. Car l'acceptation du duel consiste en l'intention expresse de se battre, laquelle celui-ci n'a pas.<sup>18</sup>

The irony is that those who most frequently appeal to internalism about intention have a decisive tactical reason not to. They usually do so while defending absolutist views about killing against one or other sort of consequentialist attack. But the move leads them straight into check-mate, because of course the consequentialist can help himself to exactly the same sort of double-effect reasoning – and generalise it in a way that is fatal to absolutism. If we may say, with Germain Grisez<sup>19</sup>, that the craniotomy is not an act of killing but only an act of making certain physical alterations in certain material, and stress the good intentions with

---

<sup>18</sup> Pascal, *Provincial Letters*, available at [http://fr.wikisource.org/wiki/Les\\_Provinciales](http://fr.wikisource.org/wiki/Les_Provinciales) [26.1.2011].

<sup>19</sup> “[...] the baby’s death need not be included in the proposal adopted in choosing to do a craniotomy. The proposal can be simply to alter the child’s physical dimensions...” (Grisez, *Living a Christian Life*, Franciscan Press, Quincy, Illinois 1993, p. 502. Cp. Boyle o.c. 480: “the death of the one who is killed by the craniotomy in Hart's famous example is not intended in the [present] sense of intention”).

which (or the good plan as part of which) this neutral physical procedure is performed, then *pari passu* the abortionist may say that she is clearing an obstacle from the mother's womb, not killing the foetus, and stress *her* good intentions. Also *pari passu*, the nuclear bomber may say that he intends to end the war by pressing a button, not by annihilating Nagasaki. Even the concentration camp guard may say that his intention is to keep his job by mixing one chemical with another, not by releasing poison into a gas chamber. (*A fortiori*, then, there need be no moral difficulty at all about unplugging oneself from Judith Jarvis Thomson's violinist, and the reason why, as John Finnis immediately suspected, is indeed to do with DDE.)<sup>20</sup>

In each case there is no doubt *some* possible description of what the agent does under which – in supposed compliance with DDE – what is done is not *per se malum* but either good or indifferent, and under which no bad effect is allowed to be a means rather than a mere side-effect. In fact the only part of the DDE, so understood, that puts any substantive limits at all on what is permitted is the proportionality clause. Which of course is grist to the consequentialist's mill, since his starting-point was the claim that the absolutist went wrong precisely in trying to attach moral significance to actions just as such, and not to the proportion between good and evil in their consequences. To put it another way – Anscombe's<sup>21</sup> – the doctrine of double effect, *so* understood, does not stretch a point on the circumference of ethics; it destroys the centre.

"But if our intentions are not determined by what we set ourselves to make them inside our own heads, then how *are* they determined?" I speculate that at least part of the reason why the literature on double effect keeps coming back to the internalist picture is because its defenders have not seen clearly enough how to give any other answer to this question. They think internalism about intention is the only way to avoid saying what they know what they *don't* want to say – namely, what Sidgwick says ([1874] p. 202):

[I]t is best to include under the term 'intention' all the consequences of an act that are foreseen as certain or probable; since it will be admitted that we cannot evade

---

<sup>20</sup> J.J. Thomson, *A defence of abortion*; John Finnis, *The rights and wrongs of abortion*. Both in Dworkin, ed., *The Philosophy of Law*.

<sup>21</sup> "[...] there are always borderline cases in ethics. Now if you are either an Aristotelian, or a believer in divine law, you will deal with a borderline case by considering whether doing such-and-such in such-and-such circumstances is, say, murder, or is an act of injustice; and according as you decide it is or it isn't, you judge it to be a thing to do or not. This would be the method of casuistry; and while it may lead you to stretch a point on the circumference, it will not permit you to destroy the centre." (Anscombe, MMP, 185 in HLAE.)

responsibility for any foreseen consequence of our acts by the plea that we felt no desire for them, either for their own sake or as means to ulterior ends: such undesired accompaniments of the desired results of our volitions are clearly chosen or willed by us.

If we don't say that the intention of the action is the inner speech I make to myself before doing it, then isn't Sidgwick's position the only coherent alternative – that what I intend in doing it is *every* consequence of doing it that I "foresee as certain or probable"? Far from it. There are usually intermediate points available between any two extremes! And that common sense takes an intermediate point between *these* extremes is something that homely examples will quickly teach us.

Here is one: earlier this afternoon I went for a bike-ride. When I did this it was, for a start, primarily the bike-ride itself that I intended, not its consequences. The consequentialist's picture of action is always an instrumental, button-pushing one, in which what we *do* is no more than a means of securing a state of the world.<sup>22</sup> There are of course *some* kinds of action that fit the consequentialist's button-pushing picture; button-pushing, for instance. But not all do. My bike-ride today didn't. I didn't do it mainly because of how things would be once it was over. There is exercise like that, of course, as when what you want is not to *do* 100 press-ups, but to *have done* 100 press-ups. But not this bike-ride of mine. I did it for its own sake, as an activity, not as a means to any state of affairs.<sup>23</sup> (Not even the state of affairs that I should be engaged in the activity.)

Secondly, when I went for my bike-ride I foresaw, indeed foresaw as certain, all sorts of consequences which were no part of my intention. For example, I foresaw (let us say) that my going for a bike-ride would create a slight alteration in Tayside NHS's health and exercise statistics. Nonetheless, *pace* Sidgwick, that certain and foreseen consequence of my choice was not something I *intended* by

---

<sup>22</sup> Some non-consequentialists seem to me to face the same objection. For example Joseph Shaw, when he defines intention like this: "Of the upshots which agent A believes he makes more likely by acting or omitting to act, A *intends* those the anticipation of which provide motivating reason(s) for his action or omission" ("Intention in ethics", CJP YEAR). This seems to be, at best, a definition of what it is to intend an upshot, not of what it is to intend. It faces other difficulties too: why, for example, can I only intend upshots that I think what I do will make more likely? When I ask my counter-suggestible pupils to stop teasing their classmate, I may well think that asking them will be likelier to make the problem worse than better (since they are counter-suggestible). Nonetheless, my intention in asking them to stop is indeed that they should stop. What I do may be objectionable on the grounds that it is not very *phronimos* of me, but it surely isn't objectionable on the grounds that by definition, trying to get them to stop can't be what I'm doing.

<sup>23</sup> More on this in "Intuition, system, and the 'paradox' of deontology", [in:] *Perfecting Virtue*.

going for a bike ride. So far from intending this certain consequence, I wasn't remotely interested in it. In possible worlds where no such consequence held, that wouldn't have made the slightest difference to any of my choices. Anyone who says to me "Since going for a bike ride *is* affecting regional health statistics, you *must* have intended such an effect" is simply wrong. And anyone who says to me "Since you knew that this consequence had a probability of (more or less) 1, you *must* have intended it" is wrong as well.

Similarly and less fantastically, I also foresaw that going for a bike ride would certainly make it true that if anyone rang me while I was out, I wouldn't get the call. Nonetheless, I wasn't intending to avoid phone calls when I went out for my ride. You *can* intend that when you go for a bike ride. On occasion I *have* intended it. But not today. So this is a certain foreseen consequence of my action of going for a ride which, again *pace* Sidgwick, is *only sometimes* part of what I intend by actions of that type. Anyone who says to me "Since going for a bike ride *is* avoiding phone calls, you *must* have been avoiding calls today" is wrong today – even though another day he would have been right.

I foresaw these consequences as certain, yet I did not intend them. Conversely, I foresaw some other consequences as far less than certain, and yet I did intend them. For instance, I was trying to go as fast as I could. Now I foresaw that it was actually pretty unlikely that I would break my personal best for this cycling route today, since today was windy and I was tired and unfit, and I set my personal best on a still day when I was feeling really strong. Nonetheless, anyone who says to me – should I need cajoling on this point – something like "Come on, riding as fast as you can *is* trying to break your personal best, so if you broke it, you broke it intentionally" is quite correct. (Compare entering lotteries: I'd say that if you win a lottery, you win it intentionally iff you entered it intentionally, even though the chances of your winning were miniscule. More evidence here that the line between the intentional and the unintentional<sup>24</sup> is not a line between any two levels of probability.)

---

<sup>24</sup> 'Don't you mean "the line between the intended and the unintended", not "the line between the intentional and the unintentional"?' My point is, precisely, that – at least where the attempted action is completed – these are *the same line*. One way of putting my thesis in this paper is that I think "What was x's intention?" can be a misleading question; to redress its misleadingness, we are often better off with the question "What did x do intentionally?". Compare a pair of questions that are conceptually very close to these questions about intention, namely "What was x's trying?" and "What did x do by trying?". The idea of a conceptual divorce between a trying and an action which is done by that trying is not a happy one. Yet some such divorce is precisely the upshot of trying to keep our talk about intentions strictly separate from our talk about our intentional actions. Cp. footnote 29.

Notice here that my act fell under a number of intentional descriptions, not just one. (As Sidgwick perhaps accepts, and as Anscombe certainly predicts above: an act “has several [descriptions] under which it is intentional”.) If you ask me “What was your intention in going for a bike ride?” I might just respond “To go for a bike ride. Nothing else” – especially if you are a police officer. But further questioning will swiftly dispel any impression that there was really just one intention on which it is uniquely correct to see me as acting. If you ask me “Did you intend to get some fresh air?” I will say Yes. If you ask me “Were you intending to test out your new bike?” I will also say Yes. “Were you out for the exercise?” will get a Yes too, as will “Were you trying to set a new personal best?”. So too will “Did you intend to have another passing glimpse of that fine church hall in Fowlis village?” (Not all cyclists are afflicted just by Boardmania or just by Betjemia; some of us have both conditions.) Indeed, if you ask me “Did you leave the house intending to think through double effect from a quite different perspective, viz. that of the saddle?”, I’ll say Yes to that too. Or if you question me in a Kammian manner,<sup>25</sup> to see if there is a third sort of effects alongside the intended ones and the side-effects, namely effects of my choice without which I would not intend it, even though they are neither (directly) intended nor side-effects, I’m sure there will be. I would not, for instance, form the intention to go for a bike-ride unless I knew that, if I do, motorists will their best not to kill me. But I do not go for a bike-ride *in order that* motorists may do their best not to kill me.

Here, as surprisingly often elsewhere in philosophy, we suffer from the curse of the definite article. The question “What is *the* intention of your action?” seems so natural; but it imports into our thinking, and right under our noses too, the contraband assumption that there must be just one intention.<sup>26</sup> (Compare “What is *the* explanation why killing is wrong?” – a question I have complained about before.<sup>27</sup>) Quite generally, the contrary is true; to take Anscombe’s own example (*Intention* p.40), if someone is intentionally moving his arm up and down with his hand round a pump handle, it is bound to be the case that he is *also* intentionally doing something else (operating the pump, replenishing the house water supply, poisoning the household; or the intentions in some other list). That is why

---

<sup>25</sup> See Frances Kamm, *The doctrine of triple effect*, PAS 2000, p. 21-39.

<sup>26</sup> We should be similarly suspicious of talk of “*the* action”, which crops up in all sorts of unexpected places, e.g. in Gerard Hughes, who even while being much more negative about the doctrine of double effect than I would want to be, is to be found talking of “*the* (perhaps complex) action-type which that individual piece of behaviour instantiates”, and of “which action [singular] it is that the agent can properly be said to have performed” (*Human Values*, p.220, 232). As if there must (really, after due analysis) always be just one.

<sup>27</sup> Cp. my *Ethics Beyond Moral Theory*, “Philosophical Investigations” 2009.

it should cut no ice with us for a theorist of double effect to say, as too many of them do,<sup>28</sup> things like “His intention is to injure the miner whose body he uses to block the trolley, and *therefore* not to kill him”. For this simply doesn’t follow. Of course there are *some* intentions which, barring gross confusion, exclude each other: if I intend to go sailing on Loch Katrine for the next hour, then I cannot intend to go sailing on Coniston Water for the next hour. But in general, to show that someone has one intention is by no means to show that he is innocent of another.

The story so far: I deny (on Wittgensteinian grounds) the internalised view of intention, as a speech I make to myself inside my own head before acting. One way to put this is to deny that there is a significant distinction between two questions about x’s completed action: “What was x’s intention?” and “What did x do intentionally?”. And I deny (by appeal to examples) the claim that often goes with the internalised view, explicitly or implicitly, that there must be some one intention in any action which is uniquely *the* intention of that action. I deny too (also by

---

<sup>28</sup> One recent example is Lawrence Masek, “Intentions, motives, and the doctrine of double effect”, PQ 2010, p. 570: “Removing Otto’s organs saves five people by providing organs, not by killing him. According to the strict definition of intention, therefore, Lily does not intend Otto’s death. I would find the strict definition mad or absurd if it permitted Lily’s action, but it does no such thing. She does not act immorally by intending Otto’s death, but she does act immorally... by stealing. I would also reject the strict definition if it entailed that Lily could be legally punished only for stealing and not for murder. Fortunately, nothing about the strict definition denies the possibility, or the prudence, of classifying actions somewhat differently in legal and moral contexts.”

Masek does not explain what is supposed to justify these different classifications; on the face of it, jurists have typically gone in for rather a lot of very moral-looking reasoning in reaching their legal conclusions. Again, why can’t Lily avoid moral blame for stealing too? If her intentions are suitably focused, surely she won’t intend theft either. The passage just quoted suggests that Masek would answer that stealing is not an intentional category of wrong. But first, that is wildly counter-intuitive: in ordinary moral (and legal!) thought, stealing *is* an intentional concept. And secondly, if we could get away with saying that stealing wasn’t an intentional category of wrong, presumably we could make the same move about killing too.

Another recent example is Joseph Shaw, *Intentions and trolleys*, PQ 2006, p. 67 (italics mine), commenting on Glanville Williams’ case of a surgeon who “takes a fancy to a patient’s heart”, and “removes it, not with the intention of killing the patient, but just to be able to... examine it... Williams then points out that a constraint against killing formulated in terms of intention would not forbid this action. *This, of course, is perfectly true.*” No it isn’t. As any court would find, the surgeon who does what Williams describes is either of unsound mind, or *does* intend to kill his patient.

It is central to Shaw’s strategy in that article to argue that many descriptions of agents in Trolley or Tramp cases as intending to kill should be replaced by descriptions of them as intending assaults. This just poses the same question again: why should they even intend *assaults*? Why can’t they just intend what Grisez might call physical rearrangements? At this point, it seems to me, Shaw too is bound to say something like “Come on, you have to recognise that this is not just physical rearrangement but also assault”. I agree. But I ask: having taken this step, why not take the next, and admit that “you have to recognise” that it’s killing too? The answer to this is *not* “Because if you take that step you will have to go all the way to consequentialism”; for as I go on to argue here, you don’t have to go all the way. What stops you is simple facts about the conventional and natural characterisation of our actions.

appeal to examples) the Sidgwickian claim that what I intend is all the consequences of my intentional action that I foresee as certain or probable: as my examples have illustrated, one can both fail to intend what one takes to be a certain consequence of one's intentional action, and also intend what one takes to be a far from certain, even a far from probable, consequence. So if I am right, intention is not a matter of speeches in the head; and not a matter of foreseen consequences, certain, or probable above some given level, or otherwise. But then, what *is* intention a matter of?

One familiar suggestion in the literature at this point is *closeness*: what I intend is the action under the intensional description under which I actually do it, and also that action under whatever other intensional descriptions are close enough to that first description. But what is "close" here, and what is "enough"? A cottage industry in metaphysics has grown up to answer these questions. I myself have contributed to that cottage industry the suggestion that nothing is close enough, except intensional identity: the intensional description under which I do the action is, strictly speaking, the *only* description under which it is intentional.<sup>29</sup> I recant. For one thing: as argued earlier, in normal circumstances it is entirely wrong to say that if I knowingly and intentionally pull the trigger on a loaded gun pointed at your head, then I don't intentionally kill you. For another: as argued just now, there isn't in general any such thing as *the* unique "intensional description under which I actually do it".

A further problem for closeness is that an action can be intentional under both of two descriptions, one very "close" and one very "distant" indeed – while not intentional under some third description which (so to speak) lies in between as to "closeness". Example: we plant a time capsule in the ground. It's packed with interesting goodies from our time, and on the outside it says "Not to be opened till 2350 AD". Our action is intentional under the known description "Digging a hole in the ground and sticking a tin in it", which presumably is a "close" description on most if not all intuitive accounts of closeness. And it is intentional under the known description "Telling the people of 2350 about life in 2011", which presumably is a very "distant" description. But it is *not* intentional under the known description "Providing twenty-second-century looters with some possible spoil" – even though we are clearly doing that as well, and even though that description may well be intermediate in closeness between the first and the second. (This remains possible pretty much whatever we take closeness to be – temporal closeness, spatial co-location, coincidence across possible worlds, the entailment of one

---

<sup>29</sup> In *Two distinctions that do make a difference*, "Philosophy" (77) 2002, p. 211–234.

description by another, the probability of one description being true if the other is, normal causal separability<sup>30</sup>, some mix of these factors, or whatever.) If you think this is an isolated example, consider what happens every time you write a letter or an email, or make a phone call. Or consider a tyrant who signs a death-warrant and hands it to a runner, who takes it to the hangman, who waits till the joiner has finished building the gallows, and then hangs the tyrant's victim. On most accounts of closeness that I can think of, both the runner and joiner perform actions that are closer to the death of the victim than anything the tyrant does. Nonetheless, intuitively, the hangman and the tyrant both kill the victim (in different senses), while the runner and the joiner do not. (They might of course do things that make them culpably compliant in killing the victim, but that is another matter.)

To understand what counts as an intention or<sup>31</sup> an intentional action and what does not, we do not need the bad philosophical psychology of internalism about intention; nor Sidgwick's clearly mistaken claims about the intentionality of *whatever* is foreseen as certain or probable; nor do we need the fanciful pseudo-metaphysics of closeness. The contours of our category of intentional action are not the same as the contours of any of these other categories. What then *do* we need?

In one word, convention. (Recall, again, Anscombe as quoted above: "Circumstances, and the immediate facts about the means you are choosing to your ends, dictate what descriptions of your intention you must admit.") Or to use Wittgenstein's phrase, what grounds our judgements about agents' intentions is "the common behaviour of mankind":

Denke, du kämst also Forscher in ein unbekanntes Land mit einer dir gänzlich fremden Sprache. Unter welchen Umständen würdest du sagen, daß die Leute dort Befehle geben, Befehle verstehen, befolgen, sich gegen Befehle auflehnen, u.s.w.?

---

<sup>30</sup> I tried out this suggestion in *The Polymorphy of Practical Reason*, in my collection *Values and Virtues*, OUP 2007. I now think it's hopeless.

<sup>31</sup> Some remarks made by some writers on DDE, for example Joseph Shaw, suggest that they think that intentions and intentional actions are separate categories which should not be confused. In some senses no doubt this is true. In particular, and obviously enough, because something can always stop you acting, there can be an intention without an intentional action, but there can't be an intentional action without an intention. Still, the intention's *telos* is always the intentional action; an intention is *in potentia* what an intentional action is *in actu*: the characterisation of the intention is, necessarily, a characterisation of the intentional action that it will become if it is enacted: "intentio, sicut ipsum nomen sonat, significat *in aliquid tendere*" (Thomas Aquinas, op. cit., 1a2ae.12, 2c). In this sense, the two categories not only can but must be taken together. Cp. footnote 22.

Die gemeinsame menschliche Handlungsweise ist das Bezugssystem, mittels welches wir uns eine fremde Sprache deuten.<sup>32</sup>

There are ways in which it is fruitful to think of actions as analogous to linguistic utterances. (There are also problems with this analogy, most obviously (a) that utterances *are* actions of one sort, and (b) that utterances typically have just one meaning (at any rate one *semantic* meaning), whereas actions, as I've already stressed, typically fit more than one intentional description. These problems will do no harm provided we keep them in mind.) To understand another person's actions—indeed to understand them as *actions*, and her as a *person*—is to categorise them within a repertoire of possibilities which—like the vocabulary of a language—is both set and familiar, but also extensible and flexible in indefinitely many new ways. He is arguing, sewing, voting; she is listening, ploughing, deciding; they are computing the *n*th place of  $\pi$ , hauling on the main brace, cycling against the clock, dancing *Swan Lake*... and so on indefinitely. The *Bezugssystem* or "frame of reference" for our understanding of each other as agents—if you like, the *lexicon of action*—is this repertoire of possible things that other people could be doing. The same repertoire is also the constitutive backdrop to all our own choices; just as the same vocabulary guides both my interpretation of others' utterances, and my composition of my own utterances. When I act, as much as when I speak, I almost always choose to come out with something intelligible. And what it is for an action to be intelligible is for it to be relatable in some way or other, as novel in some respects and formulaic in others, to the complex network of circumstance and presupposition that constitutes the background repertoire, the lexicon of action—just as utterances are intelligible in virtue of their partly-novel, partly-formulaic relation to a literal lexicon, namely the language's vocabulary. (Ibidem, 337: "Die Absicht ist eingebettet in der Situation, den menschlichen Gepflogenheiten und Institutionen.")

Now most people's word-association sets, when they hear the word *convention*, tend to include *subjectivism*, *relativism*, *arbitrary*, and the like. I hope I have made it clear that in the present case (at least), the conventional view is not the subjective or the relativist one; in this case, conventionalism is *objectivism*, because the conventions we are dealing with are objective.

---

<sup>32</sup> Wittgenstein, op. cit., I, 206. Ibidem, I, 415: "Was wir liefern, sind eigentlich Bemerkungen zur Naturgeschichte des Menschen; aber nicht curiöse Beiträge, sondern Feststellungen, an denen niemand gezweifelt hat, und die dem Bemerkwerden nur entgehen, weil sie ständig vor unsern Augen sind."

To see this point about the objectivity of conventions compare this, from *The Hitchhiker's Guide to the Galaxy*:

It is of course well known that careless talk costs lives, but the full scale of the problem is not always appreciated.

For instance, at the very moment that Arthur said, "I seem to be having tremendous difficulty with my lifestyle", a freak wormhole opened up in the fabric of the space-time continuum and carried his words far far back in time across almost infinite reaches of space to a distant galaxy where strange and warlike beings were poised on the brink of a frightful interstellar battle.

... at that very moment the words, "I seem to be having tremendous difficulty with my lifestyle" drifted across the conference table.

Unfortunately, in the Vl'hurg tongue this was the most dreadful insult imaginable, and there was nothing for it but to wage terrible war for centuries.<sup>33</sup>

It is of course open to the hapless Arthur Dent what sounds he chooses to make with his mouth-parts. But given the conventions of the English language, it is a matter of objective fact that *if* he makes certain sounds, then whether he likes it or not, he will have said "I seem to be having tremendous difficulty with my lifestyle". Moreover, given the conventions of the Vl'hurg language, it is a matter of objective fact that *if* he makes certain sounds, then whether he likes it or not, he will have uttered the most dreadful insult imaginable in Vl'hurg – unfortunately, of course, the very same sounds. What settles who has said what is linguistic convention. So it is not settled arbitrarily or voluntaristically, but as a matter of objective fact. Just likewise, what settles who has done what is (so to speak) pragmatic convention; and how our pragmatic conventions stand is also a matter of objective fact.

Quite generally, what counts as a given action – dancing *Swan Lake*, hauling in the main brace, ploughing or cycling, or whatever – is settled by convention: by how that performance fits into the known pragmatic lexicon. Since to intend is simply to set oneself to do an action, the fact that the nature of actions is conventionally determined entails that the nature of intentions is conventionally determined too. Why, when I go for a bike ride, do I intend the bike ride, and the exercise, and the breath of fresh air, and the attempt on my personal best, and the sightseeing round Angus villages, but not (ever) the change in the NHS Tayside exercise statistics, and not (on this occasion) the avoiding of lunchtime phone

---

<sup>33</sup> Douglas Adams, *The Hitch-Hiker's Guide to the Galaxy*, Ch. 31.

calls? The answer is: because that is how our common frame of reference categorises and interprets my behaviour. This is how any competent user of that framework will interpret me. It is also, insofar as I am a competent user of the framework, how I myself will understand what I intend and what I do.

Just the same applies when we say “Come on – if you intend to fire a bullet through someone’s head then you intend to kill him”, or “Come on – if you intend craniotomy then you intend the baby’s death”, or make the various other *come on* remarks that I (at any rate) find myself longing to make so much of the time in so many of the most typical debates about double-effect reasoning. The force of these remarks too lies in their appeal to our common *Bezugssystem*, our shared framework for understanding actions. It is part of that frame of reference that e.g. cutting off a baby’s head is inseparable from killing it. And that – not anything about foreseen consequences, or “closeness”, or little speeches inside the head – is why the two intentions are inseparable.

The same convention-based approach does good explanatory work elsewhere. For example, I think it gives the right answer to the old Tramp vs. Trolley chestnut. In this, we are asked (as entrance interviewees for Merton College, Oxford, used to be asked<sup>34</sup>) what the moral difference is between diverting the famous trolley so that 5 lives are saved and one isn’t, and dissecting the familiar tramp for transplant organs that will save 5 lives at the cost of his life. The difference can’t be centrally to do with action and omission; if the doing/ allowing distinction is morally relevant to Tramp, then it is hard to see why it isn’t equally relevant to Trolley. Nor can the difference be centrally about where you direct your “inner acts of intending”; if you can give your inner acts some suitably harmless target in Trolley, it is hard to see, for reasons explained above, why you can’t do the same in Tramp. Nor again, to knock away another of my own earlier attempts,<sup>35</sup> is the key difference really a point about whether we are confronted by an already-existing problem, or as it were create the problem for ourselves – though that is the closest of these failed attempts. The real moral difference between Tramp and Trolley is, I suggest, simply that in Tramp I kill, in Trolley I don’t. More fully, and if we think it helps to add this (sometimes it may, sometimes not, depending on our explanatory context): I don’t have a killing intention in Trolley, because in that case no interpreter of my action who had a decent com-

---

<sup>34</sup> Not by me, I hasten to add.

<sup>35</sup> I did try this idea out, in various drafts. I am not sure any of them ever made it into print. I got the idea partly from Williams’ talk of “confrontation” in *Ethics and the Limits of Philosophy*, Chapter 3, and partly from Foot’s talk of the difference between “initiating” and “deflecting” (*The doctrine of double effect and the problem of abortion*, [in:] *Virtues and Vices*).

petence with the relevant conventions would interpret my action as having a killing intention. But I do have a killing intention in Tramp, because no competent interpreter of my action there would or indeed could interpret my action as having anything *but* a killing intention. That my action in Trolley is a reaction to something confronting me, and a deflection of causal processes already under way, whereas my action in Tramp is an initiation of a causal process, is also relevant: it helps us see *why* any competent interpreter of my actions will so interpret them. But the central point is the point about how we determine, in line with our shared human conventions, what count as an act of intentional killing, and what doesn't.

"But conventions are society-relative (they vary from place to place) and history-relative (they vary from time to time). So conventions can't *objectively* determine the facts about what you intend!" There is an interesting and important mistake in this inference, one which is entirely characteristic of contemporary analytic ethics. The mistake is the background assumption that "what things really are" – actions, for example – can't be historically conditioned. That is, it is the mistake that feminists, though analytic philosophers, often call "essentialism". Nothing is a bigger obstacle to real progress in the theory of action than this ahistorical assumption about the nature of action. Rejecting essentialism, we should take seriously the possibility that there are action-kinds which are not accessible to us *purely for historical reasons*. No one now, for example, can sacrifice his daughter to Hera to placate her and get a following wind to sail to Troy; and no one in Europe has ever been able to commit hara-kiri. (Perhaps even the Japanese cannot commit hara-kiri any more.) If the categories which demarcate our ethical concepts are, as the jargon has it, "thick", it follows logically that the categories that demarcate our actions must be thick as well.

So if we are asked "Why couldn't it be that intentions which were inseparable within one frame of reference were separable within another?" the answer is: It not only *could* be, it *is*. Deliberately belching at the end of a meal is, inescapably, an insult to your host in Reigate, and a complimentary act in Riyadh – or so I'm told.

None of this implies that the action-lexicon of one culture cannot be translated, at least roughly, into the action-lexicon of another; no more than we should think that, because different cultures have different languages, there can be no shared reference-point by which to translate one language into another. This is the "gemeinsame menschliche Handlungsweise" of which Wittgenstein speaks as a necessary condition of understanding humans from other cultures. (It is also

why, *pace* C.S. Lewis, we could not understand talking lions<sup>36</sup> – Wittgenstein, *op. cit.*, I, 190.)

“But conventions are partial and incompletely determinate. So they can’t fix what it is that we intend in every case.” I have two replies to this: (a) maybe it is *also* incompletely determinate what we do and do not intend, so that our standard accurately matches the vaguenesses that are actually there in what it is measuring. And (b) to echo what Austin said about ordinary language – my claim is not that the pragmatic conventions that together constitute the lexicon of action are the last word. It is only that they are the first word.

“But if we can give reasons for why the conventions hold, as you suggest above, then it is not the conventions that matter; it is these reasons.” This is not so, unless to give a reason for a convention is always to reduce the convention to some underlying level of non-conventional discourse. But that is not the only way to explain a convention; explanations can be illustrative, narrative, comparative, and many other things, as well as subsumptive. (Compare the things we might say about why murder is wrong, e.g. the things that Orwell says in his famous description “A Hanging”; Orwell points to the fact that the condemned man steps round a puddle, not to claim that human individuals are the more valuable *because* they step round puddles, such circumnavigation being a value-adding feature; but rather as one small illustration of what it is that we are valuing when we value human life. Think too of what we might say about why the past tense of “sing” is “sang”, not “singed”; here to explain is partly a matter of pointing out the history of the word, and partly a matter of citing analogues like “swim” and “swam” – and neither sort of explanation grounds the conventions of grammar in something non-conventional.)

“But if conventions change, we must be able to *control* their change, at least to some extent. And if we can do that, then can’t we *make it the case* that, e.g., an intention to perform craniotomy *becomes* separable from an intention to kill?” We certainly can control the conventions whereby we live; we are not necessarily uncritical about them. But if we criticise some one of our own (or some other society’s) conventions, we can only do this by reference to our other conventions; here it is a philosophical illusion to think that we have anything else which goes “deeper” than the conventions to criticise them with. Attempted general reforms of this sort, such as Sidgwick’s above, quickly lead us, as they led Sidgwick, into saying things that ordinary competent users of the conventions are simply going to find hopelessly implausible. As for the specific proposal to reform our conventions so

---

<sup>36</sup> Not at any rate if they were really just *lions*, which Aslan isn’t.

that craniotomy and killing come apart: any specific proposal faces the challenge to show us why, and how, the proposed reform of our conventions improves the intelligibility of action. In the case of this proposal, a response to this challenge has not even been formulated, and I have no idea how it might be met.

Besides its use in analysing the Tramp vs. Trolley chestnut, the convention-based approach helps us with some other chestnuts too; Loop and Fat Man, for instance. In Fat Man, it is not a question of diverting the trolley away from the five miners by pressing a button; it is a question of stopping the trolley by throwing a fat man onto the track in front of it. No sane or normal person, following the ordinary conventions about how we use action-descriptions, would unqualifiedly describe what I do in the straight Trolley case as an action of *killing* (and so of intentional killing), even though the one miner's death is a certain and foreseen consequence of my action. By contrast, every such person would call throwing the fat man onto the track an act of killing him. (And as above, if it is objected that the intention is only to *injure* him, the best response to that seems to be "Yes, you intend to injure him *too*.") Likewise in Loop, diverting the trolley away from the five miners and towards the one only saves the five miners *provided that* the trolley hits the one; if it doesn't hit him, the trolley will carry on round the Loop and mow down the other five miners from behind. So if you intentionally divert the trolley in Loop, you must be intending it to hit the one miner. In the original Trolley case, if you found out that the trolley had not in fact hit the one miner, you would be delighted and relieved; in the Loop variant, you would think "My plan has misfired". That shows that hitting the miner – which, at least as the story is usually told, means killing him – is part of your plan in Loop; killing him is a means to the end of saving the others, not a side-effect of pursuing that end. So here it looks to me, on the whole, as if you do have a killing intention. Loop is not straightforward, because it is an even more unusual case than most of the others in the modern casuist's panoply; but under the conventional approach, it comes out, on reflection, as pretty clearly a case of intending to kill.

Does that mean that diverting the trolley in Loop is *wrong*? It does if you think that the content of the relevant absolute prohibition is that "It is always wrong to intend to kill an innocent". But that, it seems to me, is the wrong way for an absolutist like Aquinas, or Anscombe, or Boyle, or Grisez, or Finnis, or Shaw, or myself to understand the prohibition on murder – a prohibition which, it is agreed on all sides, is exceptionless only once we have made certain important excep-

tions, perhaps just war or capital punishment for instance.<sup>37</sup> This is partly because so understanding it can very quickly lead us to want a sense in which we don't intend to kill an innocent in Loop, and this is almost certain to be the internalised-intention sense once more. There are possible cases where *whatever* I intentionally do—in the conventional sense of “intentionally do” that I have been developing—I will intentionally kill an innocent. The Crashing-Aeroplane case, where my only options are to crash the aeroplane of which I am the pilot into more innocent civilians or fewer innocent civilians, is like that. Craniotomy is like that too—or at least those particular types of craniotomy case where the child's death is certain in any case, and the mother's death is certain without craniotomy, and I am the attendant doctor (so that, given my special responsibility to my patients, intentional abstention on my part here also counts as intentional action). So are most of the cases involving uterine cancer in a pregnant woman that get described in the liter-

---

<sup>37</sup> It seems far preferable to follow Anscombe and Aquinas in saying that in a just war, or in gross civil disorder, the state and its officers have the right to intend to kill, than to follow Grisez in supposing that what happens in a just war is that the virtuous soldiers foresee but do not intend the deaths of their opponents, or similarly that what happens in self-defence is that the defender foresees but does not intend the death of his assailant. Something like this might be true in some cases of self-defence, such as the one discussed by Aquinas in my epigraph, where perhaps only blows are involved, not even swords (never mind guns, obviously). In some cases what I do is punch or perhaps just push away my assailant, and here it is perfectly feasible to say that his consequent death is *praeter intentionem*; in other cases perhaps I stab him in the guts, and here it is *not* feasible—not on an externalist reading, and not given the limitations of thirteenth-century medicine—to say that his death is *praeter intentionem*. That need not show that the externalist reading is wrong. It might show, rather, that Aquinas' rule about killing in self-defence is more stringent than internalists about intention have realised.

Another possibility about the intention to kill that seems quite patent in some cases of self-defence: perhaps some at least of these cases can be treated as cases in which the self-defendant virtually has the authority of the state and the common good to intend killing, as Aquinas would (in my view rightly) require if such intentions are to be licit. Evidently sometimes such actions are legitimated retrospectively by a not-guilty verdict, or indeed prospectively by a common-law tradition including such verdicts. This gets close to saying that such cases of self-defence are done *ministro iudicis*, “carrying out the order of the judge”. And so perhaps it suggests that such acts of intentional killing in self-defence can meet Aquinas' tough conditions for licit killing.

The *Catholic Catechism*, 1994 edition, seems to my inexpert eye to be in something of a muddle over killing and double effect: it says that “the law forbidding [the deliberate murder of an innocent person] is universally valid: it obliges each and everyone, always and everywhere” (2261) and that self-defence is only licit when one can apply double effect in the way Aquinas describes at *Summa... 2a2ae.64, 7* (2264); but also (2266–2267) that it is an open question whether capital punishment is permissible. How can this possibly be an open question, given what has already been said? The Catechism allows the state the right to punish for five reasons: redress, expiation, reformation, correction, and to pre-empt aggression (2266). But the first four of these cannot happen except through what is licit anyway; so appeal to them cannot settle, it can only beg, the question whether capital punishment is permissible. As for pre-empting aggression, even if the old fiction is maintained that a convict on death row typically remains somehow a threat to the common good because somehow an aggressor, it can hardly be supposed that *capital punishment* does not involve an intention to kill, but only the foresight of death as a side-effect.

ature. In such cases, where it simply isn't open to me to kill *no* innocents, the reasonable thing to do is not determined by the Doctrine of Double Effect at all. The reasonable thing there is to kill as few innocents as possible; which suggests in turn that the wording that we should have in the relevant absolute prohibition is something like "It is always wrong to intend to kill an innocent *where it is open to you not to*".

In any case, where I do have to kill, and in so killing am not flouting the absolute against murder, however exactly that should be defined – in such cases it is also reasonable, of course, for me to do this killing with deep repugnance and horror, and utterly unreasonable for me to do it with delight, or as part of a plan to get rid of the baby, or "to kill someone to see what it feels like". So such cases do show that there can be *some* place in ethics for making speeches to myself inside my head about what I am doing, and the aspects of it that I find morally repulsive; after all, from *Nicomachean Ethics* onwards it is a familiar enough thought that character is revealed by our attitudes, ranging from regret to delight, to the things we in one sense or another "have to" do. Even then, however, such speeches are not intentions; nor, on their own, are the plans that such speeches might express. They may capture my motives<sup>38</sup>, or morally important aspects of the intentional choices, aspects that can make the difference between permissibility and impermissibility: as Aquinas says in my second epigraph, "Quamvis et isti etiam peccent si privata libidine moveantur". But as I have argued, the test of "the common way of behaviour of humanity" shows that these speeches do not typically capture intentions.

What comes out in that last paragraph, I think, is how overworked DDE has been. Cases like Craniotomy and the Crashing Aeroplane – and, I would say, Loop – have been analysed using DDE; as if the fact that DDE is *sometimes* the appropriate casuistical tool implied that it is *always* appropriate. But it is the usual philosopher's disease, the disease of over-generalisation, to think that a principle that works perfectly well in *some* cases must be stretched and contorted so that it covers *all*, or as many as possible. What I am proposing here – following Anscombe and, as Anscombe argues, Aquinas himself<sup>39</sup> – is simply that we not

---

<sup>38</sup> For a robust recent defence of DDE that makes much of the distinction between intention and motive see Ralph Wedgwood, *Defending Double Effect*, available at <http://users.ox.ac.uk/~mert1230/DDE.a4.pdf>.

<sup>39</sup> In this paper I have not been able to cite Aquinas as much as I would have liked to, as a source for the externalism about intention that I develop. I think he *is* such a source, and that this can be shown by a careful reading of (in particular) Questions 12–20 of the *Prima Secundae*. (It is striking how often modern debates about double effect cite 2a2ae.64,7c and not 1a2ae.12. It is the latter which is the set-piece disquisition on intention; the former is barely more than an *obiter dictum*.)

overreach ourselves. Like any other moral principle, the doctrine of double effect cannot do everything. Trying to make it do too much will only lead to trouble. Indeed it already has.<sup>40</sup>

## References

- Adams Douglas, *The Hitchhiker's Guide to the Galaxy*, Orion Publishing, London 2012.
- Bolt Robert, *A Man for all Seasons*, Macmillan, Basingstoke 1985.
- Boyle Joseph, "Toward Understanding the Principle of Double Effect," *Ethics* (90) 1980.
- Boyle Joseph, "Who is Entitled to Double Effect?," *The Journal of Medicine and Philosophy* (16) 1991.
- Chappell Timothy, "Two Distinctions that Do Make a Difference. The Action/Omission Distinction and the Principle of Double Effect," *Philosophy* (77) 2002.
- Chappell Timothy, *The Polymorphy of Practical Reason*, [in:] idem, *Values and Virtues*, Oxford University Press, Oxford 2007.
- Chappell Timothy, "Moral Perception," *Philosophy* (83/4) 2008.
- Chappell Timothy, "Ethics Beyond Moral Theory," *Philosophical Investigations* (32/3) 2009.
- Finnis John, "The Rights and Wrongs of Abortion: A Reply to Judith Thomson," *Philosophy & Public Affairs* (2/2) 1973.
- Grisez Germain, *Living a Christian Life*, Franciscan Press, Quincy, Illinois 1993.
- Kamm Frances, "The Doctrine of Triple Effect and why a Rational Agent Need not Intend the Means to His End," *Proceedings of the Aristotelian Society. Supplementary Volume* (74) 2000.
- Thomas Aquinas, *Summa theologiae*, available at <http://www.corpusthomicum.org/sth0000.html>.
- Pascal Blaise, *Provincial Letters*, available at [http://fr.wikisource.org/wiki/Les\\_Provinciales](http://fr.wikisource.org/wiki/Les_Provinciales) [26.1.2011].
- Rizzolatti G., Sinigaglia C., *Mirrors in the Brain: How We Share our Actions and Emotions*, Oxford University Press, Oxford 2008.

---

However, to develop such a reading properly would take a long time, especially because a lot of what Aquinas says at least *looks* patient of either an externalist or an internalist reading. In the present context I remark only that the starting-points for the externalist reading are a close look at Aquinas' notion of the *obiectum* of an action, which so far as I can see is best translated as the *objective description* of that action, its genus and its species, and the intrinsic and essential relationship that Aquinas sees between *obiectum* and *voluntas*; a relationship which the intention mediates.

<sup>40</sup> This paper had its origin in a contribution to an Anscombe Centre for Bioethics day conference in Blackfriars in Oxford in 2011. My thanks for discussion, commentary, and help to Vivienne Brown, Adrian Haddock, John Haldane, Sir Anthony Kenny (my commentator at the Blackfriars conference), Erasmus Mayr, Eleonore Stump, Tom Pink, Anthony Price, Roger Teichmann, Ralph Wedgwood, and especially Joseph Shaw, who has generously shown me extensive drafts of his book-in-progress on DDE. None of these discussants is in any way culpable for the present paper, unless it is culpable of them to allow me to write it while foreseeing, but not intending, what would be in it.

Shaw Joseph, "Intentions and Trolleys," *Philosophical Quarterly* (56/222) 2006.

Thomson J.J., "A Defence of Abortion," *Philosophy & Public Affairs* (1/1) 1971.

Wedgwood Ralph, "Defending Double Effect," *Ratio* (24/4) 2011.

Wittgenstein Ludwig, *Philosophical Investigations*, Blackwell, Oxford 1967.