

## Agency and Two-way Powers

### Abstract

In this paper I propose a way of characterising human agency in terms of the concept of a two-way power. I outline this conception of agency, defend it against some objections and briefly indicate how it relates to free agency and to moral praise- and blameworthiness.

### Agency and Two-way Powers

In the theory of action, agency is often characterised in terms of intentionality. Accordingly, there is agency when, and only when, something is done intentionally. And doing something intentionally is construed as doing it for a reason. In this paper, I propose an alternative way of characterising human agency in terms of the concept of a particular kind of causal power: a ‘two-way’ causal power. This approach is compatible with the intentional approach but I think it is preferable to it. For one thing, and although I won’t discuss this in detail, it facilitates seeing human agency as a special case of the agency that is found throughout the natural world, both in animate and inanimate things, since that notion of agency also involves the idea of causal powers.

I start with a couple of preliminary points. First, my aim is to understand individual human agency.<sup>1</sup> So I shall not be concerned with non-human agents and shall only be concerned with the agency of individual human beings and not that of groups. In what follows, then, I shall use ‘agency’ to refer to individual human agency, unless otherwise indicated.

Second, a point that has often been noted but bears repeating. Although agency generally involves *doing* something, the reverse does not hold. In general the verb ‘do’ is not a reliable indicator of human agency, even when its grammatical subject refers to a human being because, when it is not just an auxiliary verb, ‘do’ is often used to talk about what happens to one or to ‘sub-personal’ agency. For example, fainting, falling asleep, having one’s hair cut digesting food, turning sugar into fat are all things we do, but they are not actions, or not *our* actions: they are instances of things that happen to us or are done to us; or actions or doings that are properly attributed to sub-personal systems, such as the digestive system.

So we can ask the question: Which of the things we do are instances of our agency? I shall argue that there is human agency whenever there is the exercise a distinctive kind of causal power, namely, a *two-way* causal power. The term ‘two-way power’ is medieval but the concept is Aristotelian. Aristotle talks about two-way powers or capacities, by contrast with ‘one-way’ powers. The former are characterised by the fact that when the conditions for the exercise of the power obtain, the power need not be manifested - as we shall see later, whether the power is manifested depends on the thing whose power it is.

---

<sup>1</sup> And I leave it open whether other animals ever manifest such agency.

So the idea that the power of agency is a two-way power is familiar and goes back to Aristotle – but it is also controversial.<sup>2</sup> In this paper I offer a sketch and defence of this way of characterising human agency. The paper is divided into two parts. In the first part, I explore a range of doings that can be instances of agency and suggest that they are such instances when they involve the exercise of a two-way power. In the second part I elaborate and defend the idea that the human capacity for agency can be understood in terms of the concept of a two-way power. The paper can only be a sketch, rather than a full and detailed defence of this position: my aim is to show that this is a plausible conception of human agency.

## PART 1. Agency

I. There is a pre-theoretical notion of agency according to which an agent is something or someone that makes things happen, someone with the power to cause things. This pre-theoretical notion of agency extends to animals and plants, and also to inanimate things.<sup>3</sup> Within this conception, human agency can be viewed as the exercise of our causal powers. As we shall see, not all instances of human agency consist in exercising a causal power – but many do, and most are related in one way or another to such powers. Thus the idea of a causal power, a power to cause things, is central to understanding human agency.

What sorts of things do we cause when we exercise our causal powers? We cause events to happen, as well as processes to unfold and to change, and states to obtain; and we cause objects to begin or cease to exist. We do that when, for example, we press a button, or spin a top, or spin it faster, or dry the dishes; or when we paint a picture, crack a nut, or burn down a house. But although we can cause things in all these categories, causing any of them involves causing events, for the initiation of a process, or the causing of a change within it, the obtaining of a state, and an object's beginning or ceasing to exist are events – events by causing which, we cause those other things (processes, states, the coming into existence of objects, etc.).<sup>4</sup>

In claiming that causing these changes involves causing events, I follow von Wright, who characterizes an event as a change from one state to another, or from a state to a process, or from a process to a state; or as a transformation of processes.<sup>5</sup> For instance, a change from a jumper being in the drawer to its being on the bed, or from a person's being alive to his being dead. Examples of other types of event are: a change from a ball's being stationary to its rolling along the ground and vice-versa, or a change from its rolling on the ground to its flying through the air, or from its rolling slowly to its rolling quickly, and so on.

---

<sup>2</sup> The idea is found in Aristotle's *NE* and in *EE*. It is also found in Aquinas *ST* and Reid (1788), among others. In contemporary philosophy it has been endorsed, though with slight variations by Taylor (1966), Hart (1968), Galen Strawson (1986), Geach (2000), Steward (2012), among others. I endorse this view in Alvarez (2009).

<sup>3</sup> For a defence of this view see Alvarez & Hyman (1998).

<sup>4</sup> This has been denied on the grounds that when we act we cause *states*. Coope (2007), for example, attributes this view to Aristotle. I cannot address this issue but it is enough to note here that causing a state involves causing an event, namely, the change into that state.

<sup>5</sup> Von Wright, 1956, ch. 2 §6. Von Wright does not distinguish between states and states of affairs. Without taking a position on that issue, I construe the idea that an event is a transition from one state to another as meaning that it is a transition from something's being in one state to its being in another state.

Now I turn to the question how we cause events when we exercise our causal powers. The most common and uncontroversial way in which we do so is by moving our bodies. This is true both because moving one's body is itself causing a change in or to one's body, and because most of the other changes we cause, we cause *by* moving our bodies. For example, when we cause a change in the position of a door, or in the shape of a lump of dough we do so, typically, by moving our arms and hands. But it seems that we can also cause changes without moving our bodies. For example, we can cause offence by not greeting someone, cause a death by not feeding someone, and so on.

One might reject this suggestion of the grounds that these are not genuine causal claims because causal claims relate events, or at any rate, particulars, and my not doing something is not a particular. But that seems unduly restrictive. Some causal claims are causal-explanatory claims. They don't state relations between particulars but they still identify factors that contribute to the causal explanation of an event. And sometimes, an agent's not doing something is such a causal factor on account of which the person can be said to have caused the relevant event. The conditions under which someone can be said to have caused an event by not doing something are complex, involving as they do issues about explanatory salience, duties, roles, and 'normative expectations' generally, as well as counterfactuals concerning what would have happened if the agent had done what she failed to do, among other things.<sup>6</sup> But, when the conditions are met, these are cases where a change is caused (say, a hamster's death) by an agent's not doing something (e.g. by not feeding it). Note that it is not necessary that either the not-doing or the causing be intentional.<sup>7</sup> I may forget to feed the hamster and thus cause its death unintentionally.

So it is possible to cause an event by not doing something. And, since causing an event is an instance of agency, the idea that it is possible to cause an event by not doing something appears to suggest that not doing something *itself* can be an instance of agency. Is that right?

II. Some philosophers think that just as some doings are actions, some not-doings are also actions – albeit 'negative' actions (see Vermazen, 1985). So on this view, not doing something can exemplify agency. Before we can assess those claims, it is necessary to characterise not-doings. We can think of a not-doing as, simply, the negation of a doing: for every kind of doing there is a corresponding kind of not-doing, the negation of the action-kind, which implies the absence of any particular action of the corresponding kind. Examples of actions and their corresponding negations are: uttering a word and not uttering a word, giving a present and not giving a present, going for a walk and not going for a walk, voting and not voting, etc. The term used to describe the not-doing need not be negative: 'keeping silent' is a not-doing because it consists in not doing something, namely not uttering any sound, even though the term 'keeping silent' does not involve a negation.<sup>8</sup>

---

<sup>6</sup> The term 'normative expectation' is Williams's. I discuss these issues in Alvarez 2001.

<sup>7</sup> Not even under some description. This claim depends on rejecting the idea that if one causes an event by not doing something, then the event is caused by whatever one does instead of not doing that thing. I think there is reason to reject the idea. See Anscombe (1965/6) and Alvarez (2001).

<sup>8</sup> That is why not-doings can often be referred to as 'doings': keeping silent, keeping still, remaining on one's place, refraining from kicking, etc.

Two related but distinct questions arise in connection to not-doings and agency. One is whether a not-doing is an action. The other is whether not-doings can be manifestations of agency *even if* they are not actions.

If an action is anything that can be said to be intentional or voluntary, then some instances of not-doings are actions, since they are intentional or voluntary. But note that, as well as voluntary and intentional inactivity (not-doing), there is also voluntary and intentional passivity (having something done to one or undergoing something), as when I allow myself to be anesthetized.<sup>9</sup> And yet, to have something done to one is not to perform an action. This speaks in favour of another characterisation of the term ‘action’, where it is reserved for something that involves causing change. On this characterisation, instances of not-doings are not actions. To some extent, this is a terminological choice and I shall use the term ‘action’ in the second sense.

But whether we think of them as actions or not, it seems that some cases of not doing something are *manifestations* of agency. This is so, for example, when not doing something is the manifestation of a decision or a choice, as when I decide not to greet someone *in order* to cause them offence, or to communicate my disapproval of them. I may fail to cause offence or to communicate disapproval. Nonetheless, my not greeting them is a manifestation of agency: of my decision not to greet them. My decision not to greet them needn’t be some antecedent mental event that causes me to withhold my greeting: the decision may just consist in my not greeting them in those circumstances. However, as I noted above, I may not do something, not because I decide or choose not to do it, but because, say, I forget. Whether such a not-doing counts as an instance of agency is, I think, unclear.<sup>10</sup> Be that as it may, we have seen that not doing something can be a manifestation of agency, even when it doesn’t involve causing something else, when it is a manifestation of a choice or decision about whether to cause something. In that way, such instances of agency are related to the concept of a causal power.

Something similar is true of other examples of agency. Thus, preventing, allowing, enabling and sustaining events, processes and states can also be instances of agency. They do not consist in causing an event but they are related, in different ways, to the occurrence or non-occurrence of an event of some kind that depends (in a causal sense) on my doing or not doing something. For instance, I can prevent a paper from flying away by holding it down; doing that is *not* causing but is rather preventing a change, by doing something: holding the paper down. And I can also prevent a change by not doing something, e.g. if I stand motionless in front of a laser-beam mechanism that controls a door, and thus I keep it open. Similarly, I can keep the water flowing by holding a lid up (sustaining by doing); allow a bird to fly out of a cage by not closing its door (allowing by not doing); or an ambassador may prevent a diplomatic incident by keeping quiet when provoked (preventing by not doing; though here it might be said that the not doing requires a decision to keep quiet, in which case this would be a preventing by doing). And so on.

But there seems to be a type of agency that does not involve the notion of causing events even in the indirect way just discussed. I am thinking of so-called ‘mental acts’, such as making a calculation in one’s head, trying to recall the time of a meeting, etc.<sup>11</sup> These seem

---

<sup>9</sup> See e.g. Anscombe, 1957, and Hyman, 2007.

<sup>10</sup> Perhaps they are instances of agency only when, in the relevant sense, I could have done what I forget to do – but I cannot go into that issue here.

<sup>11</sup> I have left out things like ‘deciding’ or ‘choosing’ because these have their own complexity. For a decision or a choice need not be ‘purely mental’, in the way in which trying

to be instances of agency but, although they may cause other things (my decision may cause me to be ashamed), they do not seem to *consist in* causing an event, so they do not seem to be exercises of a causal power.<sup>12</sup>

It might be thought that, since when we perform mental acts we cause events in our brains, mental acts are themselves also exercises of a causal power. But this does not seem right. For, even if we cause those events, the mental act does not *consist in* the causing of those events in the brain. And this is so even if one thinks that a mental act is identical to a brain event. For say, raising one's arm is causing one's arm to rise but making a calculation is not, *in the same sense*, causing some neuron's to fire. A different and perhaps more promising suggestion is that when, for example, I try to remember someone's name, I cause certain thoughts to occur in my mind – here it might be said that I cause the thoughts to occur. If that is right, then mental acts would also be instances of causal agency but the issue seems (to me) unclear.

So we have seen that, although the paradigmatic case of agency is the exercise of our power to move our body and thus cause change, there are other instances of agency, not-doings, preventings, allowings, mental acts, etc., which differ from the paradigm, though they also involve our having causal control over the occurrence of events. And the question arises: what do they have in common with causing change by moving one's body, so that it seems right to say that these are all instances of agency?

My suggestion, as I've already anticipated, is that these are instances of agency when they are exercises of a two-way causal power *or* are something done by exercising such a power.<sup>13</sup> So, my opening a door is an instance of agency when it is or involves the exercise of a causal two-way power, e.g. to move my arm. And similarly for not doing, allowing, preventing, for mental acts, etc. And this brings me to the second part of the paper, where I shall say more about the concept of a two-way power and defend the claim that it provides a plausible and illuminating way of characterising human agency.

## PART II.

### III. Two-way powers

---

to remember a name might be. For instance, choosing a main course in a restaurant may simply consist in ordering it and thus it would involve causing a change: a sound or a motion of a hand.

<sup>12</sup> So I agree with Steward that, to use her example, 2012, nt.17, if I add 24 and 38 mentally, one should not consider '24 and 38 being added', or the fact that they are added, as events that I cause. As she says, these are a passive way of describing the action, and a fact, respectively, and thus not an event I cause. But I disagree with her that mental acts 'too *are* exercises of an agent's power to make one's body move' (32, my italics) for the reasons given in the following paragraph above.

<sup>13</sup> 'Done by doing something else', should here be interpreted to include *in* doing something else, as when one breaks a valuable artwork in breaking a garishly decorated vase.

I have claimed that human agency requires the exercise of a two-way power and noted that the idea goes back to Aristotle who, in the *Nicomachean Ethics*, says that ‘where it is in our power to act it is also in our power not to act’ (*NE*, 1113b6).<sup>14</sup>

To understand the idea of its being in one’s power to act and not to act, something needs to be said about the notion of possibility involved. I follow Kenny (1975, 151) in holding that the relevant possibility is that of *ability* and *opportunity*. Thus, if it is in A’s power to act and also in her power not to act at  $t$ , then at  $t$  A has the ability to  $\phi$  and the ability not to  $\phi$ , and the opportunity to  $\phi$  and the opportunity not to  $\phi$ . Kenny explains the difference between ability and opportunity as follows:

An ability is something internal to the agent, and an opportunity is something external. It is difficult to make this intuitive truth precise. The boundary between internal and external here is not to be drawn simply by reference to the agent’s body: illness, no less than imprisonment, may take away the possibility of exercising my abilities without necessarily taking away the abilities themselves (1975, 133).

The notion of ability relevant to two-way powers exercised in agency involves *knowing how* to do something. As von Wright says, ability in this sense goes beyond what he calls ‘the *can do of success*’ which, he says, ‘is always relative to an occasion for acting’, whereas the ‘can do’ of knowing how ‘is *independent* of occasions for acting’ (von Wright, 1963, ch. 3 §9) and the criterion for possession of the relevant ability is, very roughly, repeated success.<sup>15</sup> This doesn’t imply that whenever I try to exercise such an ability, given the opportunity, I shall succeed because, as Austin argued (1956) with his famous golf example, I might try but fail despite having the ability and opportunity. But, as Kenny notes (1975, 142), it does mean that, in general, if I have the relevant ability and the opportunity, and I try to exercise it, I shall normally succeed.<sup>16</sup>

The ability to do something, say cook omelettes, is something that, if I have it, I have it also when I am not exercising it, and even if I rarely exercise it – though I can exercise my ability only when the circumstances permit me to do so, e.g., when I have eggs, a frying pan, I’m not tied up, etc. That much is true of all our powers and capacities. However, for my ability to cook omelettes to be a *two*-way power it must be up to me whether I exercise the ability when I have the opportunity to do so – that is, it must be up to me whether or not I cook omelettes then. Consider, for contrast, the capacity to hear certain sounds. This is a capacity I can exercise only under the right conditions (absence of interference, well-functioning organs, etc.). But it is not a capacity whose exercise is up to me: if the conditions are right, I *will* hear the sound. (Though, of course, I may be able to alter the conditions, e.g. stop my ears, so I won’t hear it). And if I don’t hear the sound, then either I’ve lost the ability (perhaps temporarily) or the conditions were not right. Thus, the capacity to hear sounds is what Aristotle calls a ‘one-way power’.

It may be tempting to think that one can understand a two-way power as the conjunction of two one-way powers. But this is not so. For one-way powers are characterised by the fact that when the conditions for their manifestation obtain, the power will be necessarily manifested. But if an agent had the ability and opportunity to  $\phi$  and also the ability

---

<sup>14</sup> See also *EE* 1223a 4-7. Aquinas follows him in saying that ‘some capacities can be realized in more than one way (*ad multa*)’ others ‘can be realized in only one way (*ad unum*)’ (*ST* 1a 2ae, 49, 4).

<sup>15</sup> The two are connected, as the ‘can do’ of know how must build on the ‘can do’ of success – and both depend ultimately on our capacity to control how we move our bodies.

<sup>16</sup> But it is important to note that this does not mean that trying is a condition for the ability.

and opportunity not  $\phi$  at  $t$ , and this were the conjunction of two one-way powers, then the agent would both  $\phi$  and not  $\phi$  at  $t$  – but that is impossible.<sup>17</sup>

Thus, if an ability I have is genuinely two-way, and if the opportunity to exercise that ability is present, then it will be up to me whether I exercise the ability. And if, as Aristotle says, whenever it is in our power to act it is also in our power not to act, that is, if acting is a two-way power, whether we act when the opportunity is present is up to us. And to say that it is up to us is to say that we determine, through our choice, whether we act or not.

I have claimed that human agency requires the exercise of a two-way power but have not said that every instance of agency *is* itself the exercise of a two-way power: sometimes, what we do is an instance of agency if, as I said above, it is something done *by* exercising a two-way power but is not itself the exercise of such a power. For instance, suppose that I unwittingly start a war by deliberately pressing a button. Surely both pressing the button and starting a war are instances of my agency – the difference is that one is intentional while the other is unintentional. However, since I wasn't aware of the connection between pressing the button and starting the war, I did not have the *ability* and hence didn't have the two-way power to start the war (at any rate, not by pressing the button). But I did have the two-way power to press the button and it was by doing so that I started the war. So this is an instance of agency that is not itself the exercise of a two-way power but is something I did by exercising my two-way power to press the button.

Thus the proposed view makes room for the possibility of unintentional action but it is neutral on the vexed question of action individuation: on whether when I do one thing by doing another, my doing the first is the same action as my doing the second. Moreover, it accommodates the qualified sense in which, as in the example above, it might be said that it was up to me to start the war – in as much as it was up to me to press the button and it was by doing that that I started the war.

IV. Although the view that acting is exercising a two-way power is old and familiar, it is thought to be problematic by some, and untenable by others. For example, it has seemed to many that 'the scientific picture' of the world presents a challenge to this conception of agency. And others have claimed that there are cases of human agency that do not involve the exercise of a two-way power and that, therefore, *that* cannot be what is distinctive about human agency – so Aristotle's claim that 'where it is in our power to act it is also in our power not to act' must be false. I shall examine some of these objections here, starting with an objection that concerns determinism.

The objection can be put as follows. The idea that human agency is the exercise of a two-way power implies the falsity of determinism, because if determinism is true, there are no two-way powers. But, although we don't know whether determinism is true, we know that

---

<sup>17</sup> I take this to be the gist of Aristotle's argument in *Metaphysics*:

It follows that as for potencies of the latter kind, when the agent and the patient meet in accordance with the potency in question, the one must act and the other be acted upon; but in the former kind of potency this is not necessary, for whereas each single potency of the latter kind is productive of a single effect, those of the former kind are productive of contrary effects, so that if the rational powers were under the same necessity, one potency would produce at the same time contrary effects. But this is impossible. Therefore there must be some other deciding factor, by which I mean desire or deliberate choice (1048a6ff).

there is human agency. So our idea of human agency cannot require the concept of a two-way power.

There are two responses one can make to this objection. One is to turn the objection on its head and say that, if the idea that agency is the exercise of a two-way power implies the falsity of determinism, then, since we know that there is such agency, determinism must be false.<sup>18</sup>

The second response is to deny the first premise which says that determinism is incompatible with the existence of two-way powers. Some compatibilist positions (e.g. those that involve a conditional or dispositional understanding of two-way powers) do not seem helpful for my purposes because they don't retain a genuine notion of a two-way power. This is because, on those positions, it is not true that whenever it is in an agent's power to act, it is also in her power not to act. Rather, according to such views, an agent's ability to act other than she does is *conditional* on factors whose presence or absence are not within the agent's control. Effectively, on these views, an agent's power to act is a one-way power. (And *mutatis mutandis* for dispositional conceptions).

However, there are other compatibilist positions that retain a genuine, non-conditional, notion of a two-way power. On these positions, if an agent has a two-way power to act, whether she exercises when the opportunity obtains is up to her. And these forms of compatibilism deny the idea that a scientific-determinist conception of the physical world is incompatible with the idea that agents have two-way powers. And they maintain that arguments to the contrary depend on equivocating between different kinds of possibility (e.g. physical and agential).<sup>19</sup>

It is not necessary for my purposes to choose between these two responses, nor indeed between the particular versions of compatibilism so long as there is a version that retains a non-conditional notion of a two-way power, because it is sufficient for me to note that jointly they show that the claim that (the possibility of) determinism undermines the idea that agency could be the exercise of a two-way power is, at the very least, inconclusive.

V. But it may be thought that there are grounds other than the possibility of determinism for rejecting this conception of agency. For there seem to be instances of human agency that do not involve the exercise of a two-way power. Among others, the following have been advanced as such cases: reflex actions and reactions; actions done under hypnosis, somnambulist actions, and more generally adjustments made during sleep; spontaneous expressions of emotion; things done under duress (e.g., threats, violence); actions that are the result of 'psychological' necessity, whether this is pathological (such as addiction and

---

<sup>18</sup> Steward (2012) seems to defend a version of this response, although she does not claim it to be a conclusive argument, and in any case her view might be more accurately rendered by saying that if determinism implies that we cannot settle anything, then, since we can settle some things, determinism must be false. As she recognizes, this response implies that the truth of determinism can be settled *a priori* – which may be thought to be untenable.

<sup>19</sup> A proper discussion of this issue is not possible here. Campbell (2005) offers an overview of some compatibilist positions, as well as his own, as does Kapitan (2011). Berofsky (2011) surveys compatibilist positions that depend on a dispositional analysis of 'can do otherwise'. Kenny (1975) defends a non-conditional compatibilist position, in chapters VII and VIII. And, more recently, List [ms] has outlined and defended in detail a version of this kind of compatibilism.



obsessive compulsions) or character/volitional necessities (e.g., Luther, or actions that result from strong emotions). And, in addition, the range of so-called ‘Frankfurt-style cases’ (and the related ‘Fischer-style’ cases), which involve brain manipulation. And there may be others.

I have discussed many of these cases elsewhere (Alvarez 2009). Here I shall summarise my response to Frankfurt-style cases and focus on other types of cases.

Frankfurt-style cases are intended to show that an agent can be morally responsible even though he could not have ‘acted otherwise’. Here, my focus is rather whether such cases are examples where an agent acts even though it was not in his power not to act. The suggestion that they are such examples depends on the claim that, in those cases, if the counterfactual intervener intervened, he could cause the agent *to act* in such a way that the agent could not refrain from acting. My contention is that most examples depend on highly questionable, if not question-begging, notions of what it is for someone (i) to choose or decide to act, or (ii) to act. Thus, some examples assume that causing a brain event can *amount to* causing a choice or decision. Here is an case from Pereboom:

But to ensure that he chooses to evade taxes, a neuroscientist implants now a device in his brain which, were it to sense a moral reason occurring with the specified force [to make it likely that he will not choose to evade taxes], would electronically stimulate his brain so that *he would choose* to evade taxes (2001,18. My emphasis).

Other examples turn on the implicit idea that manipulating an agent’s brain so that his body moves in certain ways – for instance, his finger moves so that a gun fires – amounts to causing that agent to act, that is, causing *him* to move his finger and thus fire the gun. But that is false.<sup>20</sup> Of course there’s much more to say about these cases and much has been said by myself and others elsewhere so I move on to other difficulties.<sup>21</sup>

There are many things we do that, it seems, we cannot generally avoid doing. This includes reflex actions and reactions (kicking when tapped on the knee, blinking, ducking to avoid a flying object, etc.), spontaneous expressions of emotions (smiling, cowering). And others. However, it seems plausible to argue that these doings are most often properly attributable to the agency of certain sub-personal systems. Some (kicking when tapped on the knee) are always so. For others, that seems to depend on the level of awareness operates at the sub-personal level.<sup>22</sup> It is true that some of these doings can rise or be brought to the personal level of awareness, and then it is generally possible to suppress or control them (to some extent). That is why these doings fall on the continuum between our agency and that of those sub-personal systems on which we rely for much of our activity, and these doings will seem closer to one or the other precisely to the extent to which we determine when they happen, suppress them if we choose, and generally bring them under our conscious control – that is, to the extent to which doing them involves exercising a two-way causal power to move.

A slightly different but related kind of example is adjustments made during sleep and somnambulistic actions – and the latter can be striking in the purposiveness and sometimes complexity of the behaviour. Again, what is crucial about these actions is awareness: when one is asleep, one is not aware of what one is doing (though there is a twilight zone of being half-sleep and half-aware). As above, in these examples, awareness and control operate at a

---

<sup>20</sup> Fischer-style cases (blockage cases) require different treatment. See my 2009.

<sup>21</sup> For a summary of recent work on this debate, see Fischer (2011).

<sup>22</sup> Many of these things can be done as full exercises of a two-way power, even of deliberate and explicit choice, as when we decide to smile to someone in order to reassure them.

sub-personal level, one below the level of full consciousness. And again, to that extent, these actions are more properly attributed to those sub-personal systems than to the agent.

VI. By contrast, cases of coercion (threats, bribes, etc.), forms of addiction and obsessive compulsion, and other forms of so-called volitional necessity are, though different from each other in important respects, cases where agents retains the ability and opportunity to act other than they do and where it is still, in the required sense, up to them how they act.

In all these cases, I shall claim, refraining from taking the course of action the agent takes is, for a variety of reasons, extremely unappealing or unacceptable to the agent. Sometimes, as in cases of coercion, this is because someone else makes them so. In others, it is the agent's habits, peculiarities, distorted conceptions of reality, moral values, emotional attachments, etc. that make the alternative unacceptable or wholly unappealing.

Cases of coercing work by someone else making a course of action that the agent would initially have rejected preferable to not doing it, because of the concomitant consequences of the latter. In these cases, the agent still chooses what to do, albeit under duress, perhaps very intense duress. But it might be objected that there are cases of extreme pain and terror where the agent cannot choose to refrain. Do agents in such conditions retain the ability to refrain from doing what they do (sign a confession, reveal a name)? The objector might concede that, in a sense, those agents retain the general ability to refrain but deny that it is really up to them whether they exercise their ability to, say, sign a document or not. One thing that is problematic about the question whether it was up to someone whether to act in the face of the pain, terror, exhaustion, etc. is that we have no independent way of measuring an agent's capacity to withstand such things, other than whether they did. In such circumstances, we can say that one agent resists, perhaps even dies resisting, while another gives in: does that show that it was not up to the second how he acted? I don't think it does – on the contrary (though he may act against his strongest considered preference and intentions). But it might be said that there are levels of pain that no man can endure so that, when faced with them, everyone will take whatever course of action will end the pain. But if so, what this shows is that humans will always choose to do what will end such pain – but not that in such cases they lack the ability to choose to do otherwise, or that their choice is not up to them. To be sure, in these cases such choices are coerced and, most likely, not blameworthy but they are still choices that agents make in the face of alternatives. So again, I do not think these cases are counterexamples to the view that agency is the exercise of a two-way power because it is not clear that, in the sense at issue that concerns control over exercise of ability, it was not up to the agents how they acted.

Examples of obsessive-compulsive disorders and addictive behaviour (drug addicts, alcoholics, kleptomaniacs, compulsive hand-washers, etc.) are often put forward as cases where agents cannot refrain from doing what they do – but this, in the relevant sense that concerns us here (ability and opportunity) is not clearly true. The claim that these agents lack the ability to refrain, that is not up to them how they act, depends on the idea that their desires to engage in the relevant actions are 'irresistible'. But, again, in the absence of a measure of an agent's ability to resist particular desires, other than whether they resist them, I think these claims should be construed to express not an inability to refrain but rather the perhaps immense difficulty in refraining, given the intensity of the sensations that accompany the desire (withdrawal symptoms, the memory of the pleasure or relief in engaging in the activity etc.) and, sometimes, the lack of effective contrary motivation to do so. For it is worth noting that inveterate addicts of all sorts sometimes give up on their addiction from one day to the next, perhaps as a result of a clearer realisation of the extent and implications (for themselves

and others) of the addiction. And it is also worth noting that agents who display obsessive-compulsive behaviour often show high levels of control: attention to detail, avoidance of risk, careful planning, desisting from the action if the circumstances require it, etc. which suggests that they have the ability to refrain from doing what they do and that they will do so if they see reason (within their pathology) to do so – that is, it is up to them how they act. Moreover, they are often treated with behavioural therapies that depend on the gradual habituation to *refraining* from the behaviour – but if they lacked the ability to refrain, if it really was not up to them whether to act or not, how could the therapy begin?

VII. The final kinds of examples I shall discuss along these lines are those of so-called ‘volitional necessities’. One such case is what Williams calls ‘moral incapacities’, where an agent is said to be morally incapable of doing certain things: lying, betraying a partner, etc.. Do these cases present a difficulty to my account? If an agent is indeed (morally or emotionally) unable to do something, then it is not in her power to do it or not to do it, that is, it is not up to her whether she does it or not. But if that is so, if there are indeed such incapacities, these are not objections to my account unless we insist that in those cases the agent’s not doing those things are instances of agency – but why should we so insist when we do not insist that, for example, someone’s not flying unaided or their not speaking Dutch when they don’t know the language is an instance of agency? But it is questionable that at least some of the examples often given are indeed cases of inability to refrain in the sense relevant here. Luther is said to have claimed that he could not recant. But first, this is not the same as saying that he could not have refrained from publishing his views – he surely could not have but chose not to refrain, perhaps because he thought it his duty of conscience. And as for whether he could recant, the right description seems also to say that he chose not to because recanting was, to him, a morally unacceptable choice.<sup>23</sup> There is no more reason to think that Luther had no ability to recant (he clearly had the opportunity) rather than that he chose not to do so because it would go against his conscience, than there is to think that a relative who says that she cannot put you up for the night has no ability to do so, rather than that she chose not to, perhaps because it would cause her a great deal of inconvenience. To be sure, there are important differences: doing the thing that is said to be impossible might destroy the moral integrity of the person in one case, while in the other it would simply disrupt her domestic arrangements. But this does not affect the point about ability and its being up to the agent how she acts.

But there may be cases where it seems the agent really cannot refrain: for example, a mother may be unable to refrain from rushing into a burning house to save her children.<sup>24</sup> But does this suggest that she had no ability and opportunity to refrain, that it was not up to her how she acted? I don’t see it does. For suppose that the mother is told that by running into the house she’d make it more likely that her children will die. Surely she would then not rush in – though she may still feel a strong inclination to do so. This, I contend, shows that she *had* the ability and opportunity to refrain in the actual case and that it was up to her how she acted – what she didn’t have before but would have after being so informed is *the motivation* to refrain from rushing into the house.

---

<sup>23</sup> It seems that the words ‘Here I stand, I can do not other’ are a later addition to his declaration before the Diet of Worms, which in fact was just: ‘I cannot and will not recant anything, since it is neither safe nor right to go against conscience. May God help me.’ The famous words are not included in records by witnesses of his appearance before the Diet of Worms and their origin is uncertain.

<sup>24</sup> The example is a variant of Steward’s, 2012, 183.

One might object to these remarks as follows. Perhaps it is true that, had the agent had different motives, she *could* have refrained. But this does not show that *given* the motives she actually had, she could have refrained. But this objection depends on treating motives (reasons, desires, inclinations) to act otherwise as *conditions* for being able to act otherwise. But that seems wrong, for it conflates conditions for the exercise of an ability with conditions for the presence of the ability. The idea that someone *would not* refrain in the absence of motives to do so is not to be equated with the claim that she *could not* refrain in the absence of such motives.

But suppose, the objector might reply, that the agent could not have had different motives. In that case, surely she could not have acted other than she did. But that still does not follow. What follows is that, in the circumstances, she would not have acted differently. But it is up to an agent to choose to act other than she does, even if she is most unlikely to do so, given her current motives. So long as she's a rational agent guided by her reasons and desires, she will choose to act according to her motives. But that is consistent with the claim that, even with such motives, she could have chosen otherwise, for she had the ability and opportunity to choose not to act as she did.<sup>25</sup>

So, to summarise my response to the general objection that some cases of human agency do not involve the exercise of a two-way power, I claim that these alleged counterexamples are cases where the doings are not *our* actions but those of sub-personal systems; or cases where it is not true that the agent lacked the ability and opportunity to act or not to act, and that it was up to her whether she acted or not – and that is what my account of agency requires.

VIII. Finally, I shall say something about certain cases of not-doing which are often thought to refute the two-way power conception of agency. Consider Locke's example of the man who is carried while asleep 'into a room where is a person he longs to see and speak with'. He awakes and stays in the room, though unknown to him the room is locked fast, so he cannot get out. Is his staying not an instance of agency even though he could not have left, that is, even though his staying is not the exercise of a two-way power?<sup>26</sup>

It is indisputable that the man stays in the room and that he could not have left, since he lacked the opportunity. But what is not indisputable is whether the agency that we should attribute to the man is merely staying in the room (his failing to leave), rather than his *choosing* to stay in the room. After all, it should be noted that, although we *could not* characterise his staying itself as, say, and act negligence or recklessness, given that he could not have left, we *could* characterise his choosing to stay (of his failure to choose to leave) in those terms. And if that is right, if his agency lies in his choosing to stay and his staying is merely a manifestation of that choice, then the case does not present a problem because, although he could not have left, our man could certainly have *chosen* to leave. His choosing to leave might just have been his trying to open the door, just as his choosing to stay needn't have been anything other than his staying while being aware that that is what he was doing.

---

<sup>25</sup> This will no doubt raise the 'luck objection' in the reader's mind. In my view the objection, at least on some versions, depends on a questionable assumption that if an action is not necessitated by an agent's motives then it is a matter of chance – but the alternative is that it is a matter of choice. But of course much more would need to be said to address the objection properly.

<sup>26</sup> Locke says his staying is voluntary but not free, since he cannot forbear (*Essay Bk II, Ch XXI, Of Power, §10*).

He could not have executed his choice, since the door was locked, but he could have made the choice. And since he could have chosen to leave (i.e. refrained from choosing to stay), his choosing to stay, which was his agency, was an exercise of a two-way power.

## Conclusion

I have proposed and tried to defend a conception of human agency as the exercise of a two-way power and have claimed that such a conception enables us to accommodate a whole range of doings and not doings that we seem inclined to characterise as instances of agency. This suggestion, as we have seen, is linked to the idea that whether we exercise our abilities, when the opportunity to do so is present, is up to us: that we can choose whether to do so or not.

Perhaps because of this, many who have defended similar views, have presented them as accounts of *free* action. But I think the conditions for agency I have given provide only a necessary condition for free agency, not a sufficient one. For, at least on some very plausible conception of what it is to act freely, many exercises of two-way powers are not free, for they are constrained, since they are done because of coercion, obligation, compulsions, psychological disorders, etc.<sup>27</sup> And whether someone's agency is so constrained is a fact that will affect the extent to which we think that agent is praise- and blameworthy for what he does and fails to do, and for their consequences. However, when, how, to what extent, and why different factors affect freedom of agency, and the corresponding praise- or blameworthiness attaching to it, are extremely complex questions that I have not tried to address in this paper.<sup>28</sup>

Maria Alvarez

King's College London

Department of Philosophy

King's College London

Strand

LONDON

WC2R 2LS

maria.alvarez@kcl.ac.uk

---

<sup>27</sup> In that respect I disagree with Galen Strawson (1986), who says that our actions are all free – though the disagreement may be terminological, since he accepts that there is a sense of 'free' in which they are not.

<sup>28</sup> I am grateful to several people for comments on previous drafts, specially John Hyman, Erasmus Mayr and participants in research seminars at Bristol, Nottingham, Oxford and Sheffield, as well as the audience of the Aristotelian Society meeting of 10<sup>th</sup> December 2012.

## BIBLIOGRAPHY

- Alvarez, Maria, 2009, 'Actions, Thought-experiments and the "Principle of Alternate Possibilities"', *Australasian Journal of Philosophy*, 87: 1, 61-81.
- , 2001, 'Letting Happen, Omissions and Causation', *Grazer Philosophische Studien*, 61: 63-81.
- Alvarez, Maria and John Hyman, 1998, 'Agents and their Actions', *Philosophy*, 73: 219-245.
- Anscombe, G.E.M., 1957, *Intention*, (Oxford: Blackwell).
- , 1965/6, 'Comment on Mr Bennett's "Whatever the Consequences"', *Analysis*, 26: 25.
- Aquinas, Thomas, 1960–73, *Summa Theologiae*, ed. Gilby, T., (Cambridge: Blackfriars).
- Aristotle, 1984, *The Complete Works of Aristotle*, ed. Barnes, J. (Princeton NJ: Princeton University Press).
- Austin J. L. (1961) 'Ifs and Cans', *Philosophical Papers*, J.O. Urmson and G. J. Warnock (eds), (Oxford: Clarendon Press), 205-232.
- Berofsky, Bernard, 2011, 'Compatibilism without Frankfurt: Dispositional Analyses of Free Will', *The Oxford Handbook of Free Will*. Second Edition, ed. Kane, R. (Oxford: Oxford University Press), 153-174.
- Campbell, J.K., 2005, 'Compatibilist Alternatives', *Canadian Journal of Philosophy*, 35, 3: 387-406.
- Clarke, Randolph, 2009, 'Dispositions, Abilities to Act, and Free Will: The New Dispositionalism', *Mind*, 118: 323-351.
- Coope, Ursula, 2007, 'Aristotle on Action' *Proceedings of the Aristotelian Society*, Supplementary Volume, 81(1): 109-138.
- Frankfurt, Harry, 1969, 'Alternate Possibilities and Moral Responsibility', *Journal of Philosophy*, 66 (4): 829-39.
- Fischer, John Martin, 2011. 'Frankfurt-Type Examples and Semi-Compatibilism: New Work' in *The Oxford Handbook of Free Will*. Second Edition, ed. Kane, R. (Oxford: Oxford University Press), 243-265.
- Geach, P.T., 2000. 'Intention, Freedom and Predictability', in *Logic, Cause and Action: Essays in Honour of Elizabeth Anscombe*, ed. Teichmann, R., (Cambridge: Cambridge University Press), 73–81.
- Hornsby, Jennifer, 1980, *Actions*, (London: Routledge and Kegan Paul).
- Hyman, John, 2007, 'Three Fallacies about Action', in *Proceedings of the 29<sup>th</sup> International Wittgenstein Symposium*, (Hölder-Pichler-Tempsky).
- Kapitan, Tomis, 2011, 'A Compatibilist Reply to the Consequence Argument', in *The Oxford Handbook of Free Will*. Second Edition, ed. Kane, R. (Oxford: Oxford University Press), 131-150.

- Kenny, A.J.P., 1975, *Will, Freedom and Power*, (Oxford: Blackwell).
- List, Christian, 'Free Will, Determinism, and the Possibility of Doing Otherwise' (ms).
- Locke, John, 1979 [1689], *An Essay Concerning Human Understanding*, ed. P.H. Nidditch (Oxford University Press).
- Pereboom, Derk. 2001, *Living Without Free Will*, (Cambridge: Cambridge University Press).
- Reid, Thomas, 1969 [1788], *Essays on the Active Powers of the Human Mind*, ed. Brody, B.A., (Cambridge Mass.: MIT Press).
- Steward, Helen, 2012, *A Metaphysics for Freedom*, (Oxford: Oxford University Press).
- Strawson, Galen, 1986, 'On the Inevitability of Freedom from the Compatibilist Point of View', *American Philosophical Quarterly*, 23 (4): 393-400.
- Taylor, Richard, 1966, *Action and Purpose*, (Englewood Cliffs, NJ: Prentice-Hall).
- Vermazen, Bruce, 1985, 'Negative Acts', in *Essays on Davidson: Actions and Events*, eds. Vermazen B. and Hintikka, M.B., (Oxford: Oxford University Press).
- Von Wright, G.H., 1963, *Norm and Action*, (London: Routledge and Kegan Paul).
- Williams, B.A.O, 1993, 'Moral Incapacity', *Proceedings of the Aristotelian Society*, New Series, 93: 59-70.